McGRAW-HILL

ENCYCLOPEDIA OF SCIENCE & TECHNOLOGY

www.MHEST.com

16 S - SON

Sable

A carnivore classified in the family Mustelidae along with martens and fishers. The sable (*Martes zibellina*) has a very slender, long, and supple body (see **illustration**). Coloration varies but is usually dark brown on the back and slightly lighter on the flanks, belly, and head. The winter pelage (the coat, which includes fur plus longer, coarser guard hairs) is long, silky, and luxurious with a particularly strong, thick undercoat; the summer pelage is shorter, coarser, duller, and darker. The legs and tail are dark. The limbs are short and terminate in five toes with semiretractile claws. The soles of the feet are covered with extremely dense, stiff hairs. The bushy tail serves as a balancing rod. The dental formula is I 3/3, C 1/1, PM 4/4, M 1/2 × 2, for a total of 38 teeth. Adult sables have a head and body length of 35–55 cm (14–22 in.), including a tail length of 12–19 cm (5–7 in.). They weigh 0.5–2 kg (1–4 lb). *See* DENTITION.

Ecology. Sables are generally solitary and inhabit both coniferous and deciduous forests, preferably near streams. Though they prefer dense, dark coniferous forests, they occur also in stonebirch or larch forests, as well as in forest tundras and swamps. They avoid cultivated areas. Sables move by leaping and are very agile and quick. They spend most of their time on the ground, although at times they climb into trees and rocks. Burrows several meters long that begin under or among rocks, logs, or roots lead to the enlarged nest chamber, which is lined with fur and dry vegetation.

The sable (*Martes zibellina*).

Sables may be active at any time during the day or night. Food consists principally of rodents, but pikas, birds, fish, honey, nuts, and berries may also be consumed. In some regions, they are known to catch salmon migrating upstream during spawning.

Adult sables have few enemies, although young animals may be taken by larger carnivores and predatory birds. The mortality rate of the young is high, with 75–85% dying in their first year of life due mainly to climate and lack of food.

Reproduction and development. Breeding occurs from June to August, with the single annual litter being born the following April or May. The total period of pregnancy lasts 245–298 days due to the process of delayed implantation, in which the fertilized egg remains inactive in the uterus for a considerable time before attaching to the uterine wall and continuing its development. Implantation occurs in February or March; thus, actual embryonic development takes approximately 25–35 days. Usually 3 or 4 (ranging from 1 to 7) young comprise a litter. Newborn sables are blind, deaf, helpless, and weigh 25–35 g (0.9–1.3 oz). The auditory canals open between 23 and 24 days; the eyes open between 30 and 36 days; and the young leave the nest for the first time at about 45 days of age. Weaning occurs at about 7 weeks. Sexual maturity is not attained until about 27 months. The maximum known longevity of sables in captivity is 15 years.

Population status. The sable originally ranged throughout the entire taiga (boreal coniferous

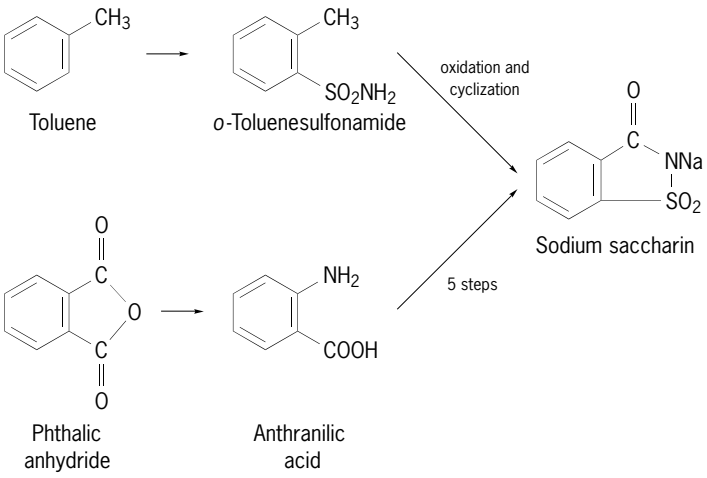
forest) zone from Scandinavia (northern Finland) to eastern Siberia and North Korea including Sakhalin and Hokkaido islands. Because of excessive trapping for its valuable fur, it disappeared from much of its range by the early 1900s. Hunting sable was completely prohibited from 1936 until 1940. However, programs of protection and reintroduction have allowed the species to increase its numbers and distribution, so that exceptionally dense populations are once again found in various areas, chiefly in the mountains of the northernmost parts of its former range. *See* CARNIVORA. Donald W. Linzey

Bibliography. D. Macdonald (ed.), *The Encyclopedia of Mammals*, Andromeda Oxford Ltd., 2001; R. M. Nowak, *Walker's Mammals of the World*, 6th ed., Johns Hopkins University Press, 1999.

Saccharin

The sodium salt of *o*-sulfobenzimide. The free imide, called insoluble saccharin because it is insoluble in water, has limited use as a flavoring agent in pharmaceuticals. The sodium and calcium salts are very soluble in water and are widely used as sweetening agents.

Saccharin was first synthesized in 1879 by I. Remsen and C. Fahlberg. Fahlberg discovered the sweet taste of the chemical by accident. Commercial production of saccharin starting from toluene began in 1900. A process for making saccharin from phthalic anhydride was developed about 1950, and saccharin is manufactured by both of these processes, which are shown in the following reaction schemes:



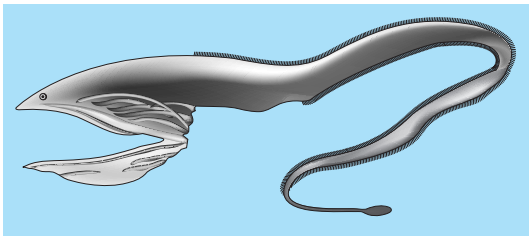
Sodium saccharin is 300–500 times sweeter than cane sugar (sucrose). The saccharin salts are used to improve the taste of pharmaceuticals and toothpaste and other toiletries, and as nonnutritive sweeteners in special dietary foods and beverages. Using noncaloric saccharin in place of sugar permits the formulation of low-calorie products for people on calorie-restricted diets and of low-sugar products for diabetics. Some people experience a bitter aftertaste from saccharin. This can be minimized by proper flavor formulation and by using saccharin as a cosweetener along with another sweetening agent such as aspartame or cyclamate. *See* ASPARTAME.

The regulatory status of saccharin varies from country to country. The United States allows its use in special dietary foods but requires a label statement: "Use of this product may be hazardous to your health. This product contains saccharin which has been determined to cause cancer in laboratory animals." Canada does not allow the use of saccharin in foods or drugs. Most countries permit some uses of saccharin but restrict applications and levels of use as a means of controlling the amount that people consume. *See* FOOD MANUFACTURING. Karl M. Beck

Bibliography. T. E. Furia (ed.), *CRC Handbook of Food Additives*, 2d ed., vol. 2, 1980; *Kirk-Othmer Encyclopedia of Chemical Technology*, 3d ed., vol. 22, 1983.

Saccopharyngiformes

Teleost fishes in the subdivision Elopomorpha whose member taxa have a leptocephalous (elongated and flatterned side to side) larval stage. They are the most aberrant and bizarre of all fishes, having lost a number of structures common to most fishes, such as the symplectic bones, opercular bones, branchiostegal bones, scales, pelvic fins, ribs, pyloric caeca, and swim bladder. The caudal fin is absent or rudimentary, the gill opening small and ventral, the dorsal and anal fins long, and the jaw and hyomandibular bones greatly elongate and attached to the neurocranium by only one condyle (see **illustration**).



Gulper eel (*Saccopharynx ampullaceus*).

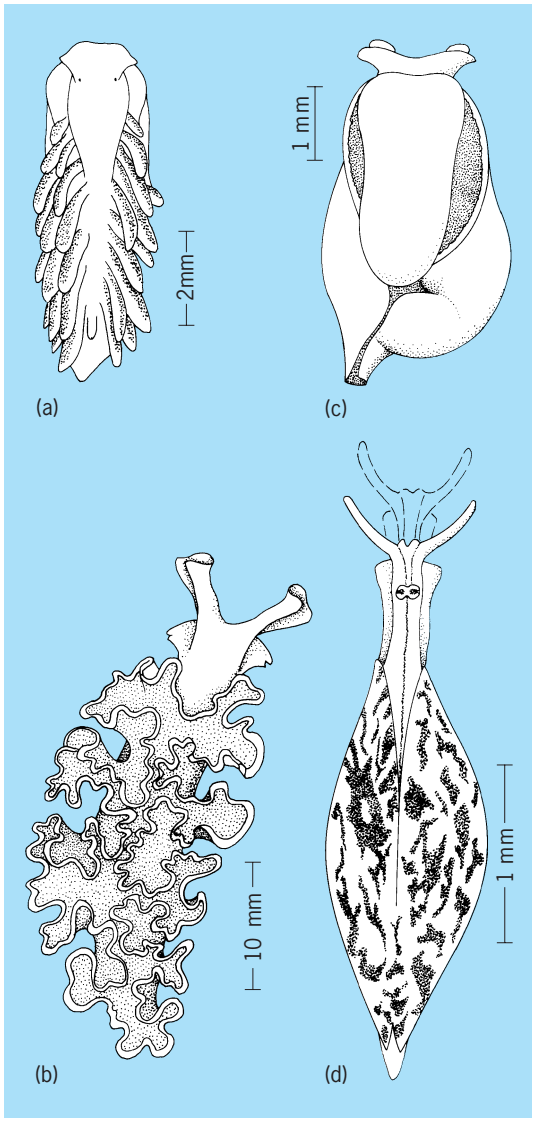
These fishes live at great depths in total darkness. They are adapted to take advantage of any prey that comes their way, large or small. The enormous jaws and distensible gut allows them to capture and swallow prey much larger than themselves.

Twenty-six species in five genera and four families are known from the bathypelagic areas of the Atlantic, Indian, and Pacific oceans. *See* OSTEICHTHYES; TELEOSTEI. Herbert Boschung

Bibliography. E. B. Bohlke (ed.), Orders Anguilliformes and Saccopharygiformes, pp. 1–655, in *Fishes of the Western North Pacific*, pt. 9, vol. 1, Sears Foundation for Marine Research Memoir (Yale University).

Sacoglossa

An order of the gastropod subclass Opisthobranchia containing a thousand living species of herbivorous sea slugs; sometimes called Ascoglossa. They occur



Representative sacoglossans: (*a*) *Stiliger*, (*b*) *Tridachia*, (*c*) *Volvatella*, and (*d*) *Berthelinia*.

in all the oceans, at shallow depths, reaching their greatest size and diversity in tropical seas.

Members of this abundant order have two common features: the herbivorous habit [except for *Olea* and a species of *Stiliger* (**illus.** *a*) which eat the eggs of other mollusks] and the possession of a uniseriate radula in which the oldest, often broken, teeth are usually retained within the body, not discarded. The radular teeth are stilettolike and are used to puncture individual plant cells so that the fluid contents may be swallowed. Dissociated chloroplasts from the prey may be retained and "farmed" in cells of the digestive gland, persisting in *Tridachia* (illus. *b*) for up to 6 weeks. This remarkable endosymbiosis occurs also in the widespread genus *Elysia*.

Sacoglossans possess varied shells; there may be a single shell, capacious and coiled (*Volvatella*, illus. *c*; *Cylindrobulla*), a flattened open dorsal shell (*Lobiger*), or two lateral shells (the bivalved gastropods, *Berthelinia*, illus. *d*). In the highest sacoglossans (*Elysia*, *Limapontia*) the true shell is completely lost after larval metamorphosis.

The association between sacoglossan species and the algae of the sea bottom may be extraordinarily precise, as evidenced in the shallow waters of Florida. *See* OPISTHOBRANCHIA. T. E. Thompson

Bibliography. S. P. Parker (ed.), *Synopsis and Classification of Living Organisms*, 2 vols., 1982; T. E. Thompson, *Biology of Opisthobranch Molluscs*, I, Ray Society, London, 1976.

Safe

An extremely strong container, usually made of metal and sometimes built into a wall or vault, designed to protect money or other valuables. A bank safe is fireproof, strongly built, and equipped with a well-locked door. A furrier's safe provides, in addition, cooling, ventilation, and protection against insects. A safe or vault is a complete structure, the walls, floor, and ceiling being as much a part of the protective features as the more conspicuous door. For utmost protection, a vault is so large and heavy that it cannot be removed. Combination locks eliminate the keyhole as a weakness, and eliminate the key, which could be stolen or duplicated. A time clock built into the locking mechanism unlocks the vault entirely from the inside. The time lock operates only after the elapsed time, set when the vault was closed. Walk-in vaults are furnished with shelves, locked drawers, and cabinets for systematic and compact storage of valuables. The vault may be air conditioned and may have a telephone extension inside. *See* STRUCTURAL MATERIALS.

Frank H. Rockett

Safety glass

A unitary structure formed of two or more sheets of glass between each of which is interposed a sheet of plastic, usually polyvinyl butyral. In usual manufacture, two clean and dry sheets of plate glass and a sheet of plastic are preliminarily assembled as a sandwich under slight pressure in order to produce a void-free bond. The laminate is then pressed at 75–225 lb/in.^{2} (0.5–1.5 megapascals) under heat at 239–302°F (115–150°C) long enough to unite. For use in surface vehicles the finished laminated glass is approximately $^1/_4$ in. (6 mm) thick; for aircraft it is thicker.

Because of the ability of the plastic interlayer to yield rather than rupture under load, and because of the adherence of the glass to the plastic, the laminated glass presents less hazard from flying or scattered glass chips in event of damage than would an equal thickness of glass. Alternatively, with chemically tempered glass, the laminate shatters into blunt granules. From this feature, such laminates are termed safety glass. However, the nonrigid plastic layer lowers the modulus of rupture of the laminate to approximately 60% of that of monolithic glass at room temperature. The thermoplastic interlayer

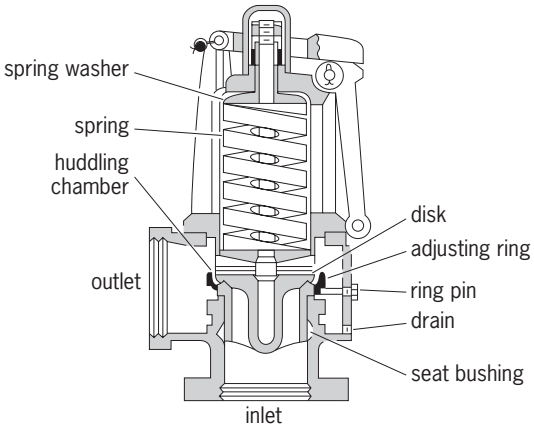
also makes other properties of safety glass appreciably temperature-dependent. Optically, safety glass has substantially the properties of the glass sheets of which it is made. However, above about 160°F (70°C) the organic plastic may deteriorate.

Most safety glass is used in automotive vehicles where the plastic interlayer is 0.030 in. (0.75 mm) thick. It is also used in ships, locomotives, railroad cars, aircraft, safety goggles, and viewing windows. Formed from tempered glass and in multiple layers built up to greater thickness, the laminate is used in bullet-resisting structures in banks, jewelry display cases, military vehicles, and test-chamber windows. The glass or plastic can be tinted to provide color filtering. Where resistance to shattering due to fire is a consideration, wire mesh is embedded in plate glass during manufacture. *See* GLASS. Frank H. Rockett

Bibliography. G. McLellan and E. B. Shand, *Glass Engineering Handbook*, 1984.

Safety valve

A relief valve set to open at a pressure safely below the bursting pressure of a container, such as a boiler or compressed air receiver. Typically, a disk is held against a seat by a spring; excessive pressure forces the disk open (see **illus.**). Construction is



Typical safety valve.

such that when the valve opens slightly, the opening force builds up to open it fully and to hold the valve open until the pressure drops a predetermined amount, such as 2–4% of the opening pressure. This differential or blow-down pressure and the initial relieving pressure are adjustable. Adjustments must be set by licensed operators, and settings must be tamperproof. The ASME Boiler Construction Code gives typical requirements for safety valves. *See* VALVE.

Theodore Baumeister

Safflower

An oilseed crop (*Carthamus tinctorius*) that is a member of the thistle (Compositae) family and produces its seed in heads (see **illus.**). Flowers vary in

Safflower (*Carthamus tinctorius*). (*USDA*)

color from white through shades of yellow and orange to red. Most cultivars of safflower are spiny. The seed is shaped like a small sunflower seed, and is covered with a hull that may be white with a smooth surface or off-white to dark gray with a ridged surface. Depending on hull thickness, the oil content varies from 25 to 45%. *See* ASTERALES.

Origin. Safflower was domesticated several thousand years ago in the Middle East, and spread at an early date in Indian and also along the Nile valley to modern-day Sudan. At first it was used as a source of dye, obtained from the flowers, and later as a source of oil. Until the discovery of aniline dyes, safflower and indigo served as the main source of dye for textiles. Over much of the Middle East, safflower is grown on a small scale for its dried flowers, which are used to color foods. *See* DYE.

Adaptation. Safflower is adapted to an environment that provides abundant supplies of moisture to the roots from seeding through to the end of flowering, and little or no moisture to the aboveground parts of the plant during and after flowering. Such an environment is provided in south-central India, where the summer monsoon rains fill the soil with moisture, which carries the crop from planting in October–November through to harvest in February–March, when the weather is very dry. In areas with a Mediterranean climate (California, Australia, and Mexico) safflower develops vegetatively during or just after the wet winters and matures in the dry summers. Deep soils with high-moisture-holding capacity have been best for safflower. At least 120 days are required from seeding to maturity.

Culture. Where safflower has been grown as a dryland crop, it has been sown in a flat seedbed with rows 12–30 in. (30–75 cm) apart. If the rows are 24–30 in. (60–75 cm) apart, there is room for mechanical cultivation. Under irrigation, safflower has been sown on raised beds, with one row on each bed if the centers are 24–30 in. (60–75 cm) apart, and two rows on beds if the centers are more than 30 in. (75 cm) apart. Frequently the herbicide trifluralin is incorporated into the seedbed before planting to prevent the development of several species of weeds. Nitrogen and phosphorus fertilizers should be applied at about the same rates used for cereal crops.

Plants may be harvested with a grain combine adjusted to provide no damage to the seed. Moisture

content of the seed should be 8% or less for safe storage.

Utilization. There are two types of safflower oil, both with 6–8% palmitic acid and 1–2% stearic acid. One type, the standard or polyunsaturated type, has 76–79% linoleic acid and 11–17% oleic acid. This high-linoleic type has been used in soft margarines marketed in small tubs, in salad oils, and in the manufacture of paints and varnishes. It has had limited use for frying foods, because heat causes it to polymerize and form a tough film on the cooking vessel. The second type of oil, called the high–oleic-acid or monounsaturated type, has 76–79% oleic acid and 11–17% linoleic acid. Its fatty-acid composition is similar to that of olive oil, but the flavor is bland. High-oleic safflower oil is a premium frying oil, and has been reported to be used in some cosmetics. *See* FAT AND OIL (FOOD).

The meal left after the extraction of the oil may contain 20–45% protein, depending on the amount of hull removed from the seed before processing. The meal is used as a poultry and livestock feed. *See* ANIMAL FEEDS.

There is a small but continuing market for dried safflower to be used for ornamental purposes. Types with no spines, long erect branches, and red flowers are preferred. Plants are harvested when in full bloom, and dried in a shaded location.

Production. In the United States, fluctuations in production have been mostly a consequence of the relative prices of safflower and wheat. Production data in other countries are not known definitely, but those in Mexico are estimated to be in the same range as those in the United States, and those in Australia about 20% of those in the United States. California has been the leading state in the production of safflower.

Pests. Insect pests have been a serious limiting factor to safflower production in the Old World, the larvae of the safflower fly (*Acanthiophilus helianthi*) being the most serious. In areas where safflower has been recently introduced, only a few insects such as western flower thrips (*Frankliniella occidentallis*) and lygus bugs (*Lygus hesperus*) have caused moderate damage.

Diseases which cause serious damage but which are controlled by resistant cultivars are rust due to *Puccinia carthami* and wilts due to *Fusarium oxysporum carthami* and *Verticillium dahliae*. Only moderate resistance has been found to a root rot caused by species of *Phytophthora*, so that control has been achieved by careful management of irrigations. Where humidities are high during flowering, gray mold caused by *Botrytis cinerea* and leaf spot caused by *Alternaria carthami* have been serious diseases. *See* PLANT PATHOLOGY. P. F. Knowles

Saffron

The plant *Crocus sativus*, a member of the iris family (Iridaceae). A native of Greece and Asia Minor, it is now cultivated in various parts of Europe, India, and China. This crocus (see **illus.**) is the source of a potent yellow dye used for coloring foods and medicine. The dye is extracted from the styles and stigmas of the flowers, which appear in autumn. It takes 4000 flowers to produce 1 oz (28 g) of the dye. *See* DYE; LILIALES. Perry D. Strausbaugh; Earl L. Core

Saffron (*Crocus sativus*), of the family Iridaceae.

Sage

A shrubby perennial plant in the genus *Salvia* of the mint family (Limiaceae). There are several species, including garden, or true, sage (*S. officinalis*), the sage most commonly used in foods. Many varieties of garden sage are known, but the Dalmatian type possesses the finest aroma. Mediterranean, or Greek, sage (*S. fructicosa*) is also used in foods, but is of inferior quality. Clary sage (*S. sclarea*) and Spanish sage (*S. lavandulifolia*) are used in perfumes and cosmetics. *See* LAMIALES.

Garden sage is native to southern and eastern Europe, and is still cultivated extensively there and in the United States and Russia. It is a plant of low stature (2 ft or 60 cm), with hairy, oblong grayish-green leaves about $1^1/_2$–2 in. (4–5 cm) long. Sage does best in warm, dry regions, with full sun. Sage plants are established by seed or, for better uniformity, by vegetative propagation (stem cuttings). Two or three harvests of sage, with yields of 1200–2000 lb/acre (1050–1800 kg/hectare) of dried leaves, are possible by the second year of cultivation. Plantings last 2–6 years.

The quantity of essential oil is higher if the sage plant is harvested before flowering. To preserve the essential oil content and leaf color, sage is dried, as are most other herbs. Once dried, sage is separated

from the stems and made available to consumers as whole, rubbed (crushed), or ground leaves.

The dried leaves are among the most popular spices in western foods, and are used as an essential ingredient in sausages, pork dishes, and poultry stuffings. Sage is highly aromatic and fragrant, with a pungent, slightly bitter and astringent taste. Both the dried leaves and essential oil of sage are used in flavoring and for antioxidant properties in cheeses, pickles, processed foods, vermouth, and bitters. *See* SPICE AND FLAVORING.

Seth Kirby

Sagittarius

The Archer, a large, zodiacal, southern summer constellation (see **illustration**). It is identified with the centaur Chiron of Greek mythology, knowledgeable in music and medicine and skilled in hunting. (Other myths identify Chiron with the constellation Centaurus.) Sagittarius contains the center of the Milky Way Galaxy, so it is especially rich in star clusters and star clouds. However, it never rises high in the sky in the Northern Hemisphere. Visitors to the Southern Hemisphere marvel at Sagittarius's overall brightness and richness. *See* CENTAURUS; MILKY WAY GALAXY; STAR CLOUDS; STAR CLUSTERS; ZODIAC.

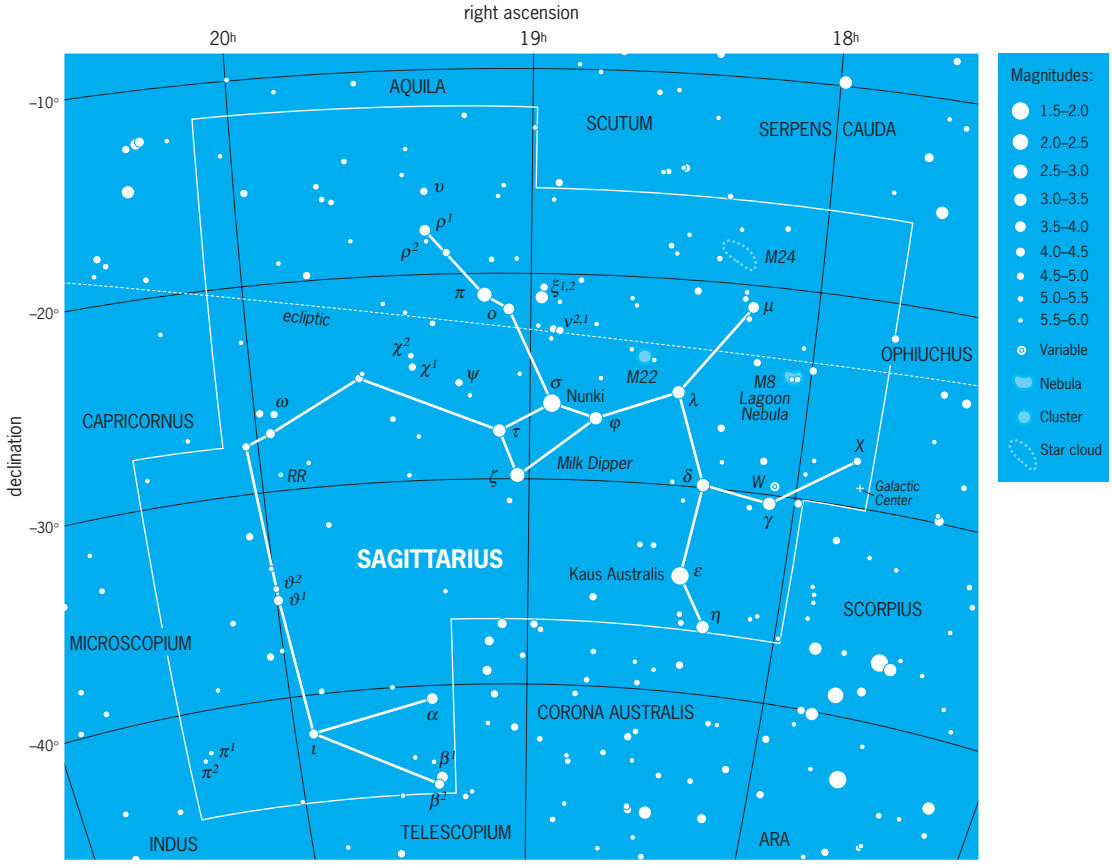
At the center of this large constellation, there is an asterism (star pattern) known as the Milk Dipper, since the Milky Way can be seen through its dipper shape. Several Messier objects (noted in Charles Messier's eighteenth-century catalog of nonstellar objects) include the Lagoon Nebula (M8), the globular star cluster M22, and the open star cluster M24. *See* MESSIER CATALOG.

The modern boundaries of the 88 constellations, including this one, were defined by the International Astronomical Union in 1928. *See* CONSTELLATION.

Jay M. Pasachoff

Sahel

A semiarid bioclimatic zone bordering the southern margins of the Sahara Desert in Africa. It extends between 13 and 18°N and stretches across the width of the African continent from the Atlantic Ocean to the Red Sea, an area of about 5 million square kilometers (1.9 million square miles). Within this area are large portions of the countries of Mauritania, Senegal, Gambia, Mali, Niger, Chad, Sudan, and Eritrea, as well as the northern parts of Burkina Faso, Nigeria, and Ethiopia (**Fig. 1**). The actual boundary of the



Modern boundaries of the constellation Sagittarius, the Archer. The celestial equator is 0° of declination, which corresponds to celestial latitude. Right ascension corresponds to celestial longitude, with each hour of right ascension representing 15° of arc. Apparent brightness of stars is shown with dot sizes to illustrate the magnitude scale, where the brightest stars in the sky are 0th magnitude or brighter and the faintest stars that can be seen with the unaided eye at a dark site are 6th magnitude. (*Wil Tirion*)

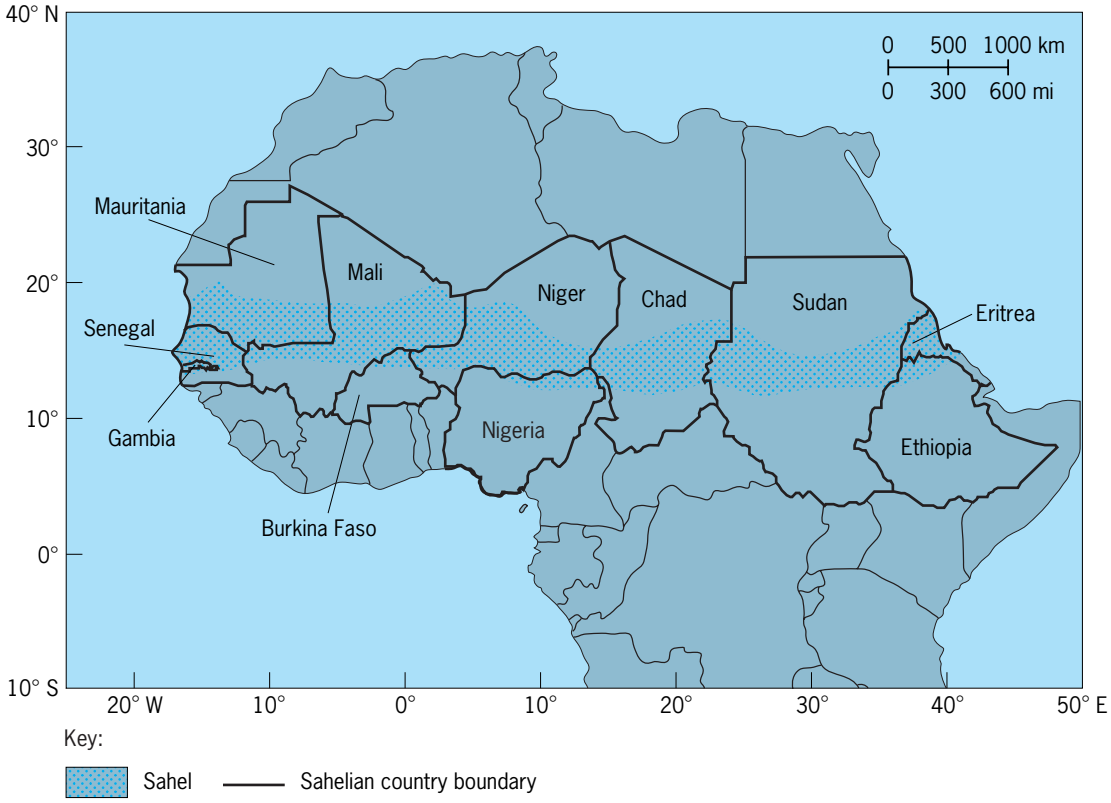


Fig. 1. Location of the Sahel. The Sahelian countries are named.

Sahel is neither as precise nor as rigid as suggested by lines of latitude. For practical purposes, the Sahel is usually defined in terms of its climate and vegetation. *See* AFRICA.

Climate. Two major but contrasting air masses dominate the climate of the Sahel: the dry, tropical continental air from the northeast and the humid, tropical maritime air from the southwest. These air masses converge in a zone of humidity and temperature contrast known as the Inter-Tropical Discontinuity (ITD; over the oceans this boundary is known as the Inter-Tropical Convergence zone, ITCZ). The ITD migrates south to north in response to the apparent position of the overhead sun and the relative intensity of the subtropical high-pressure systems located approximately over the Azores in the North Atlantic Ocean and St. Helena in the South Atlantic. In December–January, a strengthening of the Azores high and a corresponding weakening of the St. Helena causes the ITD to shift south to the coastal margins of West Africa. In the eastern Sahel, where the moderating effect of a coastline is lacking, the ITD is displaced much farther south (below the Equator). As a result, most of North Africa, including the Sahel, comes under the influence of the tropical continental air mass. The wind, known locally as the harmattan, is dry, dusty, and frequently chilly during early morning. In July–August, the dipole reverses and the ITD penetrates deep into West Africa, bringing monsoon rainfall to the Sahel. The length of the rainy season is about 2 or 3 months in the northern Sahel (July–August/September), increasing to 4 months in the southern Sahel (June–September). The mean annual rainfall also shows pronounced south–north gradients, decreasing from about 600–750 mm (24–30 in.) at the boundary of the Sahel and the Sudan savanna in the south to 100–200 mm (4–8 in.) in the north. *See* AIR MASS; CLIMATOLOGY; MONSOON METEOROLOGY; TROPICAL METEOROLOGY.

Rainfall varies significantly from one year to the next. At Niamey, Republic of Niger, for example, the total rainfall received in 1998 was 1159 mm (46 in.), more than double the average of 544 mm (21.4 in.) for the 30-year period from 1970 to 2000, and nearly four times the 324 mm (12.8 in.) received in the severe drought year of 1983. Such high variability complicates agricultural planning and is responsible for the droughts that have plagued the region. *See* DROUGHT.

Temperature is high throughout the year, but tends to be highest between March and June with mean monthly temperatures between 30 and 35°C (86 and 95°F), although daily maximum temperatures frequently exceed 40°C (104°F). The coldest months are December to February with mean monthly temperatures between 20 and 25°C (68 and 77°F). Potential evapotranspiration is also high throughout the year and exceeds rainfall amounts in all months except during the core of the rainy season (July–August) in the southern Sahel.

The Sahel is of interest climatically because it provides the most dramatic and sustained decline in rainfall recorded anywhere in the world (**Fig. 2**).

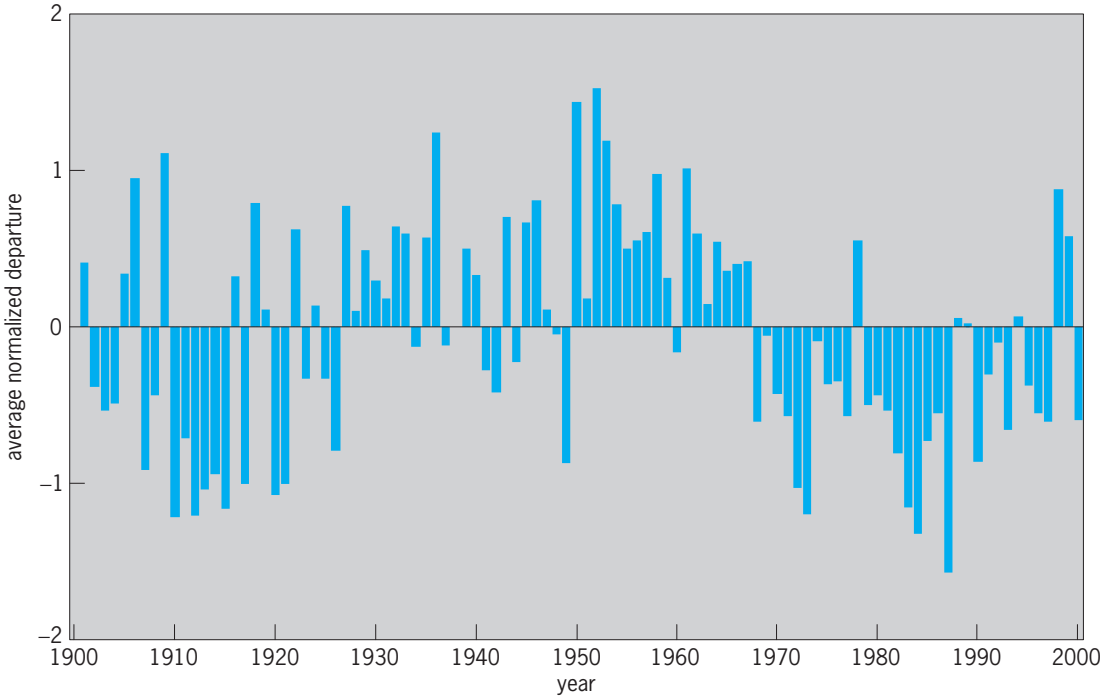


Fig. 2. Standardized rainfall variability over the Sahel during the twentieth century.

Moreover, the livelihoods of its inhabitants are closely linked to the climate, principally rainfall, and finely balanced so that fluctuations in rainfall have serious consequences on agricultural and socioeconomic systems. During the late 1960s and early 1970s, sharply decreased rainfall resulted in devastating drought and famines that caused the deaths of 100,000 to 200,000 people and approximately 12 million head of livestock. The poor rains continued throughout the 1970s and 1980s, producing another drought throughout the region in 1983–1985 that forced millions of people to abandon their homes in search of food. Rainfall conditions recovered somewhat during the mid-1990s, but they have not fully returned to predrought levels.

The causes of these droughts and climate variability in the Sahel generally are not fully understood. Major research efforts utilizing computer modeling are focusing on the remote forcing of Sahelian rainfall through ocean-atmosphere interaction or regional feedback processes involving changes in vegetation and land cover characteristics.

Vegetation. Sahel vegetation is transitional between the Sudan savanna woody grassland to the south and the open desert to the north. The vegetation consists mostly of open grassland and thorny woody species. Annual grasslands dominate the northern Sahel, while wooded grasslands occur on sandy soils in the southern part. Among the most important woody plant species are various species of acacia, combretum, and baobab (*Adansonia digitata*). In some locations, the vegetation is concentrated in strips separated by patches of bare soil. This formation produces a unique pattern called tiger brush because viewed from above it resembles the strips on a tiger's back. The reasons for this peculiar arrangement are not known. *See* DESERT; SAVANNA.

In recent decades, the vegetation in the Sahel has come under intense pressure mainly from increased demand for farmland and fuel wood to satisfy rapidly increasing urban populations. Around major cities, such as Niamey, the landscape has been largely cleared of woody species to a radius of more than 30 km (19 mi). Such changes have raised concerns about desertification—permanent land degradation from the effects of climatic variability and human activities. Such rapid land use changes may be reinforcing the effects of natural climatic variability and helping to perpetuate droughts and increasing aridity. *See* DESERTIFICATION.

Landscape features. Sahel landscape lacks dramatic topographic expression. Apart from the Red Sea hills (1270 m; 4200 ft) on the Eritrean border, sand dunes, and isolated erosion relics, most of the region consists of flat monotonous plains or gently undulating landscape generally below 600 m (2000 ft). The highest points are formed by two late Tertiary volcanic mountains in the Western Sudan: the Jebel Mara (3071 m; 10,075 ft) and the Jebel Gurgeil (2400 m; 7900 ft). A little farther west, the Devonian sandstone-capped Ennedi Plateau in eastern Chad stands 1450 m (4760 ft) high. In east-central Mali, the Dogon sandstone plateau descends gently westward to the River Niger floodplains but terminates abruptly in the southeast, forming cliffs 300 m (1000 ft) high. Known as the Bandiagara escarpment, these cliffs have been inhabited for the past millennium by the Dogon people, famous for their traditional beliefs, art, and cosmology. Farther north,

the Adrar des Iforas, a southern extension of the Ahaggar massif, reaches heights of 700 m (2300 ft). *See* MASSIF.

Rivers and surface water resources. The Niger is the most important river in the Sahel. Originating in the Fouta Djallon highlands in Guinea, the river flows northeast toward the fringe of the Sahara Desert before turning southeast in a great bend to empty into the Gulf of Guinea. This unusual course takes the river through semiarid West Africa for much of its 4200 km (2600 mi) length, providing a source of water and livelihood that helps to mitigate the effects of droughts. The original Tuareg name *egerou n-igereou*, meaning river of rivers, bears eloquent testimony to the significance that the inhabitants have attached to the river historically.

The Bani River joins the River Niger in central Mali, forming a vast inland delta. During wet periods, the delta may attain a surface area of about 80,000 km^2 (30,900 mi^2). The inland delta is important hydrologically because it helps to dissipate a significant portion of the river flow through evaporation and infiltration. *See* DELTA.

Other major rivers in the Sahel include the Senegal in the west, and in the east the Nile, which traverses the zone on its way to the Mediterranean Sea. East of Niamey are the Dallols, a system of fossil river channels that evidently supported a sizable river during the recent geologic past.

Lake Chad is the only natural lake in the Sahel. Once the sixth largest lake in Africa, this inland drainage lake has shrunk from a surface area of about 25,000 km^2 (9650 mi^2) in 1963 to less than 1400 km^2 (540 mi^2) in 2001; it is $^1/_{20}$ of the size that it was 40 years ago. As a result, Lake Chad, which supports nearly 20 million people, is increasingly referred to as the disappearing lake. The primary cause is reduced rainfall throughout the Sahel that began during the late 1960s. However, the construction of dams and increased abstraction of irrigation water on the major river systems that discharge into the lake have worsened the problem. These river systems include the Chari and Logone in Cameroon and the Komadougou-Yobe in Nigeria.

People and economy. The Sahel has traditionally been a cross roads for people and culture from North and West Africa. Indeed, some of the largest and best-organized indigenous African empires and kingdoms, including Nubia, Ghana, Mali, Songhay, and Kanem-Bornu, were located either wholly or substantially within the Sahel. These kingdoms engaged in the trans-Saharan trade with North Africa, which helped the spread of Islam into Sub-Saharan Africa, created a cultural mosaic, and gave rise to some of the legendary cities of Africa, including Tombouctou.

Rain-fed farming, nomadic livestock herding, and agropastoralism (a mixture of sedentary agriculture and livestock keeping) are the major economic activities. Unreliable rainfall makes agriculture, whether cultivated or livestock herding, very risky. Additionally, limited use of technology and modern agricultural aides such as fertilizer or pesticides results in very poor yields, making the zone susceptible to drought-induced famines. Small-scale irrigated agriculture on river floodplains is becoming widespread, especially near large urban centers.

Economically, the countries in the Sahel are poor. Of the 173 countries ranked on the 2002 United Nations Human Development Index, Sudan is the highest-rated Sahelian country at 139. The remaining eight countries rank among the 25 poorest nations in the world. These countries have primarily agrarian underdeveloped economies, and the strain of dealing with persistent drought, famines, and desertification has made many of them dependent on external aid.

Aondover Tarhule

Bibliography. M. Glantz, The West African Sahel, in M. Glantz (ed.), *Drought Follows the Plow*, pp. 33–43, Cambridge University Press, 1994; M. Hulme, Climatic perspectives on Sahelian desiccation: 1973–1998, *Global Environ. Change*, 11:19–29, 2001; M. Tiffen and M. Mortimore, Questioning desertification in dryland Sub-Saharan Africa, *Nat. Resources Forum*, 26:218–233, 2002; United Nations Development Program, *Human Development Report 2002: Deepening Democracy in a Fragmented World*, Oxford University Press, 2002; F. White, *The Vegetation of Africa*, Natural Resources Research, UNESCO, 1983.

Saint Elmo's fire

A type of corona discharge observed on ships under conditions approaching those of an electrical storm. The charge in the atmosphere induces a charge on the masts and other elevated structures. The result of this is a corona discharge which causes a spectacular glow around these points. This effect was accentuated by the contours of early sailing vessels, which usually had several masts and much rigging. The display thus created under darkened conditions was quite striking. The effect was not understood through much of that era and led to much superstition on the part of sailors. *See* CORONA DISCHARGE.

Glenn H. Miller

Salenioida

An order of Echinacea in which the apical system includes one or several large angular plates covering the periproct, with the other characters similar to those of the hemicidaroid urchins. There are two families. The Acrosaleniidae, an extinct group confined to the Jurassic and Cretaceous, had the anus displaced to one side of the periproct as a result of the unequal growth of the suranal plates covering the periproct; the tubercles were perforate and crenulate. The Saleniidae, in which the tubercles are imperforate, ranged from the Jurassic onward, with two surviving deep-sea genera. *See* ECHINACEA; ECHINODERMATA; ECHINOIDEA; HEMICIDAROIDA.

Howard B. Fell

Salicales

An order of flowering plants, division Magnoliophyta (Angiospermae), in the subclass Dilleniidae of the class Magnoliopsida (dicotyledons). The order consists of the single family Salicaceae, with about 350 pieces. There are only two genera, *Salix* (willow) and *Populus* (poplar and cottonwood), the former by far the larger. The Salicales are dioecious, woody plants, with alternate, simple, stipulate leaves (see **illus.**) and much reduced flowers that are

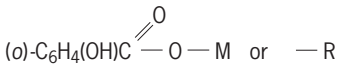
Eastern American cottonwood (*Populus deltoides*), showing female catkins with ripening capsules and cottony seeds. (*John H. Gerard*, *National Audubon Society*)

aggregated into catkins. They have numerous ovules on parietal placentae in a compound, unilocular pistil, which ripens into a usually two-valved capsule. The mature seeds are plumose-hairy and are distributed by the wind. *See* DILLENIIDAE; MAGNOLIOPHYTA; MAGNOLIOPSIDA; PLANT KINGDOM; POPLAR; WILLOW.

Arthur Cronquist; T. M. Barkley

Salicylate

A salt or ester of salicylic acid having the general formula shown below and formed by replacing the carboxylic hydrogen of the acid by a metal (M) to give a salt or by an organic radical (R) to give an ester. Alkali-metal salts are water-soluble; the others, insoluble. Sodium salicylate is used in medicines as an antirheumatic and antiseptic, in the manufacture of dyes, and as a preservative (illegal in foods). Salicylic acid is used in the preparation of aspirin. The methyl ester, the chief component of oil of wintergreen, occurs free and as the glycoside in many plants. This ester is used in pharmaceuticals as a component of rubbing liniment for its counterirritant effect on sore muscles. It is also used as a flavoring agent and an odorant. The phenyl ester (salol) and others are used medicinally. *See* ASPIRIN.



Elbert H. Hadley

Saline evaporites

Deposits of bedded sedimentary rocks composed of salts precipitated during solar evaporation of surface or near-surface brines derived from seawater or continental waters. Dominant minerals in ancient evaporite beds are anhydrite (along with varying amounts of gypsum) and halite, which make up more than 85% of the total sedimentary evaporite salts. Many other salts make up the remaining 15% (**Table 1**); their varying proportions in particular beds can be diagnostic of the original source of the mother brine. *See* ANHYDRITE; EVAPORATION; GYPSUM; HALITE; SALT (CHEMISTRY); SEAWATER; SEDIMENTARY ROCKS.

TABLE 1. Major evaporite minerals*

Mineral	Formula
Anhydrite	$CaSO_4$
Antarcticite	$CaCl_2 \cdot 6H_2O$
Aphthitalite (glaserite)	$K_2SO_4 \cdot (Na,K)SO_4$
Aragonite[LS]	$CaCO_3$
Bassanite	$CaSO_4 \cdot 1/2\ H_2O$
Bischofite	$MgCl_2 \cdot 6H_2O$
Bloedite (astrakanite)	$Na_2SO_4 \cdot MgSO_4 \cdot 4H_2O$
Burkeite	$Na_2CO_3 \cdot 2Na_2SO_4$
Calcite[LS]	$CaCO_3$
Carnallite	$MgCl_2 \cdot KCl \cdot 6H_2O$
Dolomite[LS]	$Ca_{(1+x)}Mg_{(1-x)}(CO_3)_2$
Epsomite	$MgSO_4 \cdot 7H_2O$
Gaylussite	$CaCO_3 \cdot Na_2CO_3 \cdot 5H_2O$
Glauberite	$CaSO_4 \cdot Na_2SO_4$
Gypsum	$CaSO_4 \cdot 2H_2O$
Halite	NaCl
Hanksite	$9Na_2SO_4 \cdot 2Na_2CO_3 \cdot KCl$
Hexahydrite	$MgSO_4 \cdot 6H_2O$
Ikaite[LS]	$CaCO_3 \cdot 6H_2O$
Kainite	$4MgSO_4 \cdot 4KCl \cdot 11H_2O$
Kieserite	$MgSO_4 \cdot H_2O$
Langbeinite	$2MgSO_4 \cdot K_2SO_4$
Leonhardtite	$MgSO_4 \cdot 4H_2O$
Leonite	$MgSO_4 \cdot K_2SO_4 \cdot 4H_2O$
Loewite	$2MgSO_4 \cdot 2Na_2SO_4 \cdot 5H_2O$
Mg-calcite[LS]	$(Mg_xCa_{1-x})CO_3$
Magnesite[LS]	$MgCO_3$
Mirabilite	$Na_2SO_4 \cdot 10H_2O$
Nahcolite	$NaHCO_3$
Natron	$Na_2CO_3 \cdot 10H_2O$
Pentahydrite	$MgSO_4 \cdot 5H_2O$
Pirssonite	$CaCO_3 \cdot Na_2CO_3 \cdot 2H_2O$
Polyhalite	$2CaSO_4 \cdot MgSO_4 \cdot K_2SO_4 \cdot 2H_2O$
Rinneite	$FeCl_2 \cdot NaCl \cdot 3KCl$
Sanderite	$MgSO_4 \cdot 2H_2O$
Schoenite (picromerite)	$MgSO_4 \cdot K_2SO_4 \cdot 6H_2O$
Shortite	$2CaCO_3 \cdot Na_2CO_3$
Sylvite	KCl
Syngenite	$CaSO_4 \cdot K_2SO_4 \cdot H_2O$
Tachyhydrite	$CaCl_2 \cdot 2MgCl_2 \cdot 12H_2O$
Thermonatrite	$NaCO_3 \cdot H_2O$
Thernadite	Na_2SO_4
Trona	$NaHCO_3 \cdot Na_2CO_3$
Van'thoffite	$MgSO_4 \cdot 3Na_2SO_4$

*After J. K. Warren, *Evaporite: Their Evolution and Economics*, Blackwell Scientific, Oxford, 1999.

†Less saline alkaline earth carbonates or evaporitic carbonates are indicated by LS; the remainder are evaporite salts. Less common minerals, such as monohydrocalcite, hydromagnesite, the borates, nitrates, and strontium salts, are not listed.

Today, brines deposit their salts within continental playas or coastal salt lakes and lay down beds a few meters thick and tens of kilometers across. In contrast, ancient, now-buried evaporite beds are often much thicker and wider; they can be up to hundreds of meters thick and hundreds of kilometers wide. Most ancient evaporites were formed by the evaporation of saline waters within hyperarid areas of huge seaways typically located within arid continental interiors. The inflow brines in such seaways were combinations of varying proportions of marine and continental ground waters and surface waters. There are few modern depositional counterparts to these ancient evaporites, and none to those beds laid down when whole oceanic basins dried up, for example, the Mediterranean some 5.5 million years ago. *See* BASIN; DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; MEDITERRANEAN SEA; PLAYA.

Some of the thicker ancient salt beds, especially those dominated by the salt mineral halite (common table salt), have flowed and folded like toothpaste and have been squeezed through the adjacent sediments to form large slugs of salt unconnected to the original salt bed. In the Gulf of Mexico, such slugs of Jurassic age salt have climbed upward through more than 10 km (6 mi) of Tertiary-age sediment and moved laterally more than 60–70 km (37–44 mi) into the gulf. *See* GULF OF MEXICO; JURASSIC; TERTIARY.

Evaporite minerals and brine evolution. Evaporite salts precipitate by the solar concentration of seawater, continental water, or hybrids of the two. The chemical makeup, salinity (35‰), and the proportions of the major ions in modern seawater are near-constant in all the world's oceans, with sodium (Na) and chloride (Cl) as the dominant ions and calcium (Ca) and sulfate (SO_4) ions present in smaller quantities [a Na-(Ca)-SO_4-Cl brine]. It is a matter of ongoing scientific debate whether the same ionic proportions have characterized seawater in times past. Halite and gypsum anhydrites have been the major products of seawater evaporation for at least the past 2 billion years, but the proportions of the more saline minerals, such as sylvite/ magnesium sulfate ($MgSO_4$) salts, appear to have been more variable. *See* CALCIUM; CHLORINE; ION; MAGNESIUM; SODIUM; SULFUR.

Modern marine evaporites. When modern seawater evaporates, a predictable sequence of evaporite salts precipitates from increasingly concentrated hypersaline waters (**Table 2**). The first mineral to precipitate is calcium carbonate ($CaCO_3$), usually as aragonite; this begins in mesohaline waters when the brine reaches twice the concentration of seawater (40–60‰) and achieves a density of 1.10 g/cm^3. As the brine continues to concentrate and approaches about five times the concentration of seawater (that is, 130–160‰), gypsum (calcium sulfate dihydride; $CaSO_4 \cdot 2H_2O$) precipitates from penesaline waters with brine densities of 1.13 g/cm^3. At surface temperatures higher than 45°C (113°F) but in the same salinity range, anhydrite ($CaSO_4$), the anhydrous form of calcium sulfate, tends to precipitate in preference to gypsum. At 10–12 times the original seawater concentration (340–360‰) and at densities of 1.22 g/cm^3, halite (NaCl) drops out of the supersaline waters. After halite, the bittern salts (potassium or magnesium sulfates/chlorides) precipitate from supersaline waters at concentrations more than 70–90 times that of the original seawater. If calcium sulfate is still precipitating at these higher salinities, it tends to be as anhydrite rather than gypsum. By this bittern stage of seawater concentration, the brine density is in excess of 1.30 g/cm^3 and the brines have a viscosity similar to that of liquid honey. *See* ARAGONITE; WATER DESALINATION.

The increasing density of the concentrating brine increases the likelihood that it will slowly seep into underlying sediments where it alters its surroundings via a process known as brine reflux. A common outcome of brine reflux in many ancient seawater-fed evaporite basins is the conversion of underlying limestone into dolomite via the process of reflux dolomitization. *See* DOLOMITE; LIMESTONE.

Evaporation of continental waters (from both river and ground-water inflow) precipitates a less predictable suite of evaporite salts. Unlike marine-derived brines, evaporating continental waters draw on a more diverse reservoir, and so ionic proportions in the various brines are far more variable (**Table 3**). The varying ionic proportions in the final brine depend on the types of rock that have been leached by the inflowing waters. Flow through limestone aquifers produces inflow waters rich in calcium and bicarbonate (HCO_3); dolomite weathering and dissolution generate magnesium (Mg); igneous and metamorphic matrices yield silica-rich Ca-Na-HCO_3 waters. Pyritic shales and other sulfide-rich sediments will contribute sulfate ions, whereas basic and ultrabasic rocks tend to produce alkaline Mg-HCO_3 waters. *See* IGNEOUS ROCKS; METAMORPHIC ROCKS; PYRITE.

In all, H. P. Eugster and L. A. Hardie (1978) distinguished five major water types that precipitate

TABLE 2. Mineral paragenesis and classification of concentrating seawater*

Brine stage (hypersaline)	Mineral precipitate	Salinity, ‰	Degree of evaporation	Water loss, %	Density, g/cm^3
Normal marine	Seawater	35	1x	0	1.04
Mesohaline or vitahaline	Alkaline-earth carbonates	35–140	1–4x	0–75	1.04–1.10
Penesaline	$CaSO_4$ (gypsum/anhydrite)	140–250	4–7x	75–85	1.10–1.126
Penesaline	$CaSO_4 \pm$ halite	250–350	7–11x	85–90	1.126–1.214
Supersaline	Halite (NaCl)	>350	>11x	>90	>1.214
Supersaline	Bittern salts, K-Mg Salts	Variable	>60x	>99	>1.29

*After J. K. Warren, *Evaporites: Their Evolution and Economics*, Blackwell Scientific, Oxford, 1999.

TABLE 3. Typical brine compositions of seawater and various salt lake waters

	Na	Ca	Mg	K	Cl	SO_4	HCO_3
Seawater	10,762	410	1,293	399	19,353	2,709	142
Bristol Dry Lake, U.S.A.	57,400	43,330	1,070	3,330	173,000	210	—
Coastal salina, South Australia	99,700	600	10,700	2,700	171,100	7,200	200
Dallol Salt Pan, Danakil Africa	46,020	23,290	16,570	8,860	168,560	0	—
Dead Sea, Middle East (lower brine)	39,700	17,180	42,430	7,590	219,250	420	220
Freefight Lake, Canada (lower brine)	48,531	395	15,192	3,734	10,316	118,318	15,360
Great Salt Lake, U.S.A.	105,120	420	10,420	8,480	181,730	21,730	—
Laguna de Tirez, Southeast Spain	68,000	1,100	45,000	2,600	27,500	116,000	440
Lake Assal, northeast Africa	50,300	8,310	6,760	3,100	113,600	860	52
Lake Magadi, Africa	119,500	—	—	1,470	71,750	2,100	3,250
Lake Van, Turkey	7,740	10	90	510	5,450	2,340	2,190

evaporite beds in closed continental evaporite basins: (1) Ca-Mg-Na-(K)-Cl, (2) Na-(Ca)-SO_4-Cl, (3) Mg-Na-(Ca)-SO_4-Cl, (4) Na-CO_3-Cl, (5) Na-CO_3-SO_4-Cl. As any one of these waters concentrates within a particular evaporite basin, it deposits a characteristic suite of evaporite minerals (**Table 4**). First precipitates are the alkaline-earth carbonates: low-magnesian calcite, high-magnesian calcite, aragonite, and dolomite. The mineralogy of this initial precipitate depends on the Mg/Ca ratio of the parent brine. The proportions of calcium, magnesium, and bicarbonate ions in the brackish inflow waters determine the subsequent evaporation pathway of the brine. *See* CALCITE; CARBONATE MINERALS.

Very ancient marine evaporites. It is chemically impossible for sodium carbonate minerals to form by evaporation of modern seawater. Assuming that the proportions of calcium to sulfate and bicarbonate in seawater have not changed much in the last 600–800 million years, the presence of sodium carbonate minerals such as trona or its pseudomorphs (minerals of the same outward crystal form) in ancient evaporite beds would indicate nonmarine settings. In the oceans of the very early Earth (Archean), however, the proportions of dissolved ionic components were different than today as levels of methane and carbon dioxide in the atmosphere were much higher and the atmosphere was highly reducing (little or no free oxygen). Back then, solar evaporation of the earliest oceans precipitated trona along with halite; gypsum did not form. The various types of evaporite pseudomorphs retained in Archean sediments support this rather different seawater brine evolution. Halite pseudomorphs occur in many evaporitic early-Precambrian rocks 2–4 billion years ago. Widespread gypsum/anhydrite pseudomorphs become commonplace only around 1700 million years ago as rising levels of atmospheric oxygen meant that the oxidized form of sulfur, the sulfate ion, became a stable ionic phase in the world's oceans. Only pseudomorphs, rather than beds of the actual salts, are preserved in rocks this old. The oldest actual evaporite minerals are found in Neoproterozoic beds deposited some 600–700 million years ago in Oman and Australia. *See* ARCHEAN; ATMOSPHERE; CARBON DIOXIDE; EARTH; LAVA; METHANE; PRECAMBRIAN; PROTEROZOIC.

TABLE 4. Major evaporite minerals associated with the five brine types

Brine type	Saline mineral*
Ca-Mg-Na-(K)-Cl	Antarcticite
	Bischofite
	Carnallite
	Halite
	Sylvite
Na-(Ca)-SO_4-Cl	Techyhydrite
	Glauberite
	Gypsum
	Halite
	Mirabilite
	Thernadite
Mg-Na-(Ca)-SO_4-Cl	Bischofite
	Bloedite
	Epsomite
	Glauberite
	Gypsum
	Halite
	Hexahydrite
	Kieserite
	Mirabilite
	Thernadite
Na-CO_3-Cl	Halite
	Nahcolite
	Natron
	Thermonatrite
	Trona
Na-CO_3-SO_4-Cl	Burkeite
	Halite
	Mirabilite
	Nahcolite
	Natron
	Thernadite
	Thermonatrite

*For formulas, see Table 1.

Modern distribution. Modern bedded evaporite deposits typically accumulate in saline lakes and playa mudflats within ground-water discharge regions in the arid and semiarid deserts of the world. Coastal deposits of evaporites also occur in the same desert belts in areas fed by seepage of seawater into isolated coastal depressions or mudflats (coastal sabkhas). Continental playa deposits, such as Lake Eyre and Salar de Uyuni, typically contain much larger areas and volumes of salts than modern coastal deposits, but still do not approach the aerial extent or thicknesses of their ancient counterparts. Evaporite salts also form as lake precipitates and efflorescences in the cold polar deserts of Antarctica, but the amount of salt in these regions pales to insignificance when compared with the volumes forming in settings closer to the Equator. Brine mixing and freezing, rather than direct solar concentration, play

a much more important role during crystallization of the polar deposits.

Salt-accumulating playas define discharge areas within internal drainage basins (endoheic basins); they are areas where more water is leaving the basin via evaporation than is entering it as rainfall, surface, or subsurface inflow. Modern evaporite-filled playas and lakes mostly form in deserts, in two belts located between 15 and 45° north and south of the Equator. Due to changing atmospheric circulation and climate distribution, almost all modern continental playas, in both the Northern and Southern hemispheres, have experienced numerous water-full (humid) versus dry (arid playa) stages in the last few hundred thousand years of glacial-interglacial interactions. *See* DESERT; GLACIAL EPOCH.

Ancient distribution. Due to the high solubility of evaporite minerals, almost all sizable ancient evaporite deposits are less than 600–800 million years old. This is not a result of different intensities of dissolution; otherwise, evaporites would be more common today than, say, in the Permian. Rather, there have been times and tectonic configurations in the past when evaporite deposition was much more commonplace and much more widespread than it is today. At distinct times in the past—early Cambrian of Siberia, Permian of the southern United States and northwest Europe, Jurassic of the Gulf of Mexico and the Arabian Gulf, the Late Miocene in the Mediterranean—whole seaways were characterized by the deposition of widespread evaporites. The scale and depositional diversity of these ancient evaporites far exceed that seen in modern deposits; at times evaporites were precipitated across whole sedimentary basins as units hundreds of meters thick and thousands of kilometers wide. *See* CAMBRIAN; EUROPE; MIOCENE; PERMIAN.

For example, the Late Miocene some 5.5 million years ago was the last time that a large evaporite-covered basin formed on the Earth's surface; it occupied the position of the present Mediterranean Sea. At that time, the Mediterranean Sea had dried up into four major subbasins where thick sequences (0.3–2.0 km; 0.2–1.2 mi) of subaqueous gypsum, halite, and bittern salts were precipitating from seawater-derived brines. Almost all these evaporite beds are dominated by sedimentary textures whose present-day counterparts can only be found precipitating in shallow waters (<10 m or 33 ft deep) of much less extensive modern saline lakes.

The documentation in the 1970s of an ocean-wide shoalwater depositional setting for most Mediterranean evaporites revolutionized the understanding of how widespread ancient saline evaporites were deposited. Geologists realized that there is no modern counterpart of a desiccated oceanic basin; yet in the Late Miocene (Messinian) salinity crisis, the brine-filled depressions on the Mediterranean floor were filled by 2-km-thick (3-mi) units of halite and gypsum in less than 300,000 years. These widespread evaporites were laid down on the floors of a chain of shallow saline seas/lakes, with some of the water surfaces as much as 2000 m (6600 ft) below ambient sea level. The lakes were fed by a combination of seawater seepage and river inflow, but there was no surface connection with the adjacent Atlantic and Tethyan oceans. At that time a gigantic desert extended across the old sea floor from Spain to Israel and from northern Italy to Libya. The isolation that allowed the Mediterranean Ocean to dry up was created by the collision of Africa with Eurasia via plate tectonics. *See* PLATE TECTONICS.

Compared with evaporite deposits of today, the wider extent, greater depositional and chemical diversity, and greater thickness of evaporites deposited on ancient platforms and across whole basins were due to three main factors:

1. Times of warmer worldwide climate created wider latitudinal belts for evaporite deposition and preservation. For example, there is a clear diminution in evaporite-depositing latitudes from the Jurassic to the Cretaceous, coinciding with a change from an arid to a more humid worldwide climate. Yet, even in the Cretaceous, with its transition from evaporites to coal (indicating humid climate) at around 50° paleolatitude, the extent of evaporite deposition was still wider than today.

2. Huge shallow epeiric seas often formed large shelf and shallow inland seaways in arid areas, creating areas of extensive platform and intracratonic evaporites. This style of evaporite deposit was commonplace during times of long-term tectonic quiescence and during low-amplitude sea-level fluctuations of greenhouse Earth. It was often associated with deposition of extensive carbonate platforms. Shoal-water evaporite beds in the backreef of the Permian Basin of west Texas and New Mexico were deposited in this fashion, as was the Cretaceous Ferry Lake Anhydrite of east Texas; the Red River Formation of the Williston Basin, Wyoming; and the Red Heart Dolomite of the Georgina Basin, Australia.

3. The correct tectonic and climatic conditions existed for the formation of thick (>100 m or 330 ft) evaporite deposits extending across whole depositional basins (for example, the thick salt beds of the Castile/Salado formations of west Texas, and the Zechstein strata of northwest Europe). These basinwide domains formed during tectonically active times, when a combination of climate and tectonism restricted seawater inflow to vast ocean-margin and lacustrine basins that had formed by incipient rifting, continental collision, or transtensional faulting. The last such saline giant formed 5.5 million years ago, when the collision of Africa with Eurasia converted the Mediterranean-Tethyan Sea into a chain of shoal-water evaporite lakes and the Zagros Basin into an evaporite-filled depression.

John Warren

Bibliography. G. I. Alsop, D. J. Blundell, and I. Davison, *Salt Tectonics*, Geol. Soc. Spec. Publ., vol. 100, London, 1996; N. P. James and A. C. Kendall, Introduction to carbonate and evaporite facies models, in R. G. Walker and N. P. James (eds.), *Facies Models: Responses to Sea Level Change*, pp. 265–275, Geological Association of Canada, 1992; J. R. Li et al., Responses of evaporite mineralogy to inflow

water sources and climate during the past 100 Ky in Death Valley, California, *Geol. Soc. Amer. Bull.*, 109(10):1361–1371, 1997; J. L. Melvin (ed.), *Evaporites, Petroleum and Mineral Resources*, Developments in Sedimentology Series, vol. 50, Elsevier, Amsterdam, 1991; M. G. Rowan, Structural styles and evolution of allochthonous salt, central Louisiana outer shelf and upper slope, in M. P. A. Jackson, D. G. Roberts, and S. Snelson (eds.), *Salt Tectonics: A Global Perspective*, pp. 199–228, Amer. Ass. Petrol. Geol. Mem., no. 65, 1995; B. C. Schreiber (ed.), *Evaporites and Hydrocarbons*, Columbia University Press, New York, 1988; J. K. Warren, *Evaporites: Their Evolution and Economics*, Blackwell Scientific, Oxford, 1999.

Salmonelloses

Diseases caused by *Salmonella*. These include enteritis (90–95% of all salmonelloses in the United States) and septicemia with or without enteritis (5–10%). *Salmonella typhi*, *S. paratyphi A*, *B*, and *C*, and occasionally *S. cholerae suis* cause particular types of septicemia called typhoid and paratyphoid fever, respectively; while all other types may cause enteritis or septicemia, or both together.

Typhoid fever. This type has an incubation period of 5–14 days. Typhoid fever is typified by a slow onset with initial bronchitis, diarrhea or constipation, a characteristic fever pattern (increase for 1 week, plateau for 2 weeks, and decrease for 2–3 weeks), a slow pulse rate, development of rose spots, swelling of the spleen, and often an altered consciousness; complications include perforation of the bowel and osteomyelitis. *Salmonella typhi* can be isolated from the blood in the first 10 days and later from the feces and urine. The ileum shows characteristic ulcerations. Fecal excretion of *S. typhi* usually ends by the sixth week, but 2–5% of convalescents, mostly women, become chronic carriers. The organism may also persist in the gallbladder.

Typhoid fever leaves the individual with a high degree of immunity. Antibodies in the serum are detectable from the second week on: Anti-O antibodies signify a present infection, whereas anti-H antibodies, which appear later, are observed in convalescence, after vaccination, or after an earlier infection with an antigenically (H) related *Salmonella* species. Anti-Vi is often observed in carriers. Both anti-H and anti-Vi tend to persist. *See* ANTIBODY; IMMUNOLOGY.

Vaccination with an oral vaccine gives an individual considerable protection for about 3 years, which may be broken by large oral challenge doses. *See* VACCINATION.

The only effective antibiotic is chloramphenicol, although a few resistant strains have been reported in Mexico. Ampicillin may be helpful but is more useful in treating carriers. If gallbladder disease is present, cholecystectomy should be considered in carriers.

Preventive measures must concentrate on sanitation, since this exclusively human salmonella is transmitted most often by fecal contamination of water or foodstuffs. *See* ANTIBIOTIC; WATER-BORNE DISEASE.

Paratyphoid fever has a shorter course and is generally less severe than typhoid fever. Vaccination is an ineffective protective measure.

Enteric fevers. Enteric fevers, that is, septicemias due to types of *Salmonella* other than those previously mentioned, are more frequent in the United States than typhoid and paratyphoid fever but much less frequent than *Salmonella* enteritis. In children and in previously healthy adults, enteric fevers are most often combined with enteritis and have a favorable outlook. In certain predisposed individuals (for example, those under adrenocortical steroid treatment or those suffering from sickle cell anemia, malaria, or leukemia), septicemia with or without enteritis and metastatic infection in other organs may occur; the prognosis then depends on the underlying illness. The organisms involved are the same as those causing *Salmonella* enteritis. Chloramphenicol or ampicillin are used in treatment. However, strains resistant to both drugs have been observed. *See* DRUG RESISTANCE.

Enteritis. Inflammation of the small bowel due to *Salmonella* is one of the most important bacterial zoonoses. The most frequent agents are *S. typhimurium*, *S. enteritidis*, *S. newport*, *S. heidelberg*, *S. infantis*, and *S. derby*. The incubation period varies from 6 h to several days. Diarrhea and fever are the main symptoms; the intestinal epithelium is invaded, and early bacteremia is probable. Predisposed are persons with certain preexisting diseases (the same as for enteric fevers), very old and very young individuals, and postoperative patients. Chronic carriers exist but are rare in comparison with post-typhoid carriers. The main reservoir is animals, with transmission occurring chiefly through foodstuffs.

As would be expected, enteritides occur more frequently during the summer months. There is no immunity to this salmonellosis. Antibody determinations in the patient's serum may or may not yield positive results. In spite of laboratory effectiveness of many antibiotics, antimicrobial treatment serves only to prolong the carrier state and has no effect on the disease; neither has vaccination. Prevention must concentrate on improving sanitation in commercial food production (for example, improvements in slaughtering methods, slaughterhouse conditions, and transport facilities; and such measures as exclusion of carriers and decontamination of offal and powdered bulk food such as egg powder).

Alexander von Graevenitz

Bibliography. E. J. Bowmer, The challenge of salmonellosis, *Amer. J. Med. Sci.*, 247:467, 1964; F. Cabello (ed.), *Biology of Salmonella*, 1993; J. I. Cohen et al., Extraintestinal manifestations of *Salmonella* infections, *Medicine*, 66:349, 1987; E. van Oye (ed.), *The World Problem of Salmonellosis*, 1964.

Salmoniformes

An order of generalized fishes, including the salmons and trouts, characterized by soft or articulated fin rays; the presence (usually) of an adipose dorsal fin; cycloid scales (scales absent in few species); pelvic fins in the abdominal region and each with more than six rays and an associated axillary process (absent in Osmeridea); a pelvic girdle free from the pectoral girdle; a pectoral fin placed low on the side and more or less horizontal; an upper jaw usually bordered by premaxillae and maxillae; nonprotractile premaxillae; the presence (usually) of an orbitosphenoid bone as well as a mesocoracoid arch in the pectoral girdle; the absence of Weberian apparatus connecting the swim bladder with the inner ear; a physostomous (connected to the esophagus) air bladder; the larvae are not leptocephalous-like; and there are no luminescent organs.

In the past several groups presently considered distinct orders—namely, Argentiniformes, Osmeriformes, Salmoniformes, Esociformes, Stomiiformes, and Myctophiformes—were placed, in various combinations, in Salmoniformes. The last two named orders are Neoteleosts; the remaining four orders constitute the superorder Protacanthopterygii. Herein the osmerids and salmonids are arbitrarily united in Salmoniformes as follows:

Order Salmoniformes
- Superfamily Osmeridea
  - Family Osmeridae
    - Salangidae
    - Sundasalangidae
- Superfamily Galaxioidea
  - Family Retropinnidae
    - Lepidogalaxiidae
    - Galaxiidae
- Superfamily Salmonoidea
  - Family Salmonidae (trouts and salmons)
    - Subfamily Coregoninae
      - Thymallinae
      - Salmoninae

Osmeridae (smelts). Smelts are small, silvery, minnow-like fishes, distinguished by a large mouth, large teeth in the jaws, few pylonic caeca or none, and deciduous scales. Absence of an axillary process associated with the pelvic fins readily distinguishes smelts from the salmons and trouts. They occupy cold and temperate marine waters of the Northern Hemisphere, namely the Arctic, Atlantic, and Pacific oceans. Several species are known to complete their life cycle in freshwater. Others spawn in the lower reaches of rivers from where the larvae drift seaward. Rainbow smelt (*Osmerus mordax*), a species in North America, spawn on wave-swept ocean beaches where they create excitement among locals, who find the fish easy to harvest. There are seven genera and 13 species.

Salanagidae (icefishes). Icefishes are transparent or translucent fishes with a poorly ossified skeletons, flattened head, and adipose fin, and are virtually scaleless. They are anadromous (live in salt water but spawn in freshwater) and freshwater fishes known from Sakhalin Island, off the coast of Japan, and from Korea, China, and northern Vietnam, where they are locally important food fishes. There are four genera and 11 species.

Sundasalangidae (noodlefishes). Noodlefishes are similar to icefishes but lack an adipose fin. They occur in the fresh and brackish waters of Borneo and southern Thailand. Sexually mature at less than 15 cm (6 in.), Noodlefishes are among the smallest protacanthopterygian. There is one genus and two species.

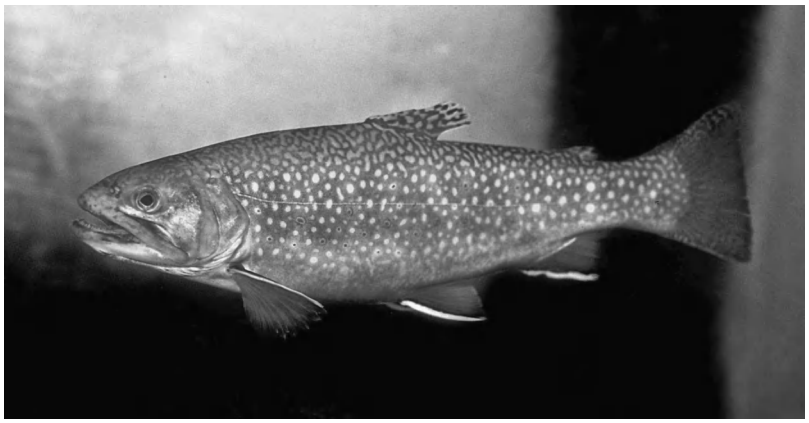
Retropinnidae (New Zealand smelts). Small plankton-feeding fishes rarely exceeding 13 cm (5 in.), New Zealand smelts differ from the above families by having after an adipose fin, a hard keel on their belly, and lacking the right gonad. They emit a cucumber-like odor when captured. Occuring in southeastern Australia, Tasmania, and New Zealand, they complete their life cycle in freshwater; however, some species spend early stages in the marine environment and return to freshwater to spawn. There are six genera and 10 species.

Lepidogalaxiidae (salamanderfishes). The single species (*Lepidogalaxis salamandroides*) in this family is an extraordinary fish. Its eyes lack muscles. Compensating for the inability to rotate the eye, it has the ability to move its head downward and sideways, which is most unusual for a fish. Endemic to a small area in southwestern Australia, salamanderfishes can survive periodic droughts by burrowing in the sand.

Galaxiidae (galaxiids). The following combination of characteristics serves to distinguish galaxiids from other salmoniforms: absence of maxillary, vomer, palatine, and basibranchial teeth; presence of paired gonads; absence of an adipose fin in most species; and lack of a cucumber-like odor. While most species spend their entire lives in freshwater, all species breed in freshwater. A few are catadromous (live in freshwater and spawn in salt water), the larvae and juveniles spending some time at sea. Some survive dry season by aestivating. *See* AESTIVATION AND HIBERNATION.

Galaxiidae is strictly a Southern Hemisphere family, with an interesting distribution. On a distribution map their range indications appear as small blobs skipping from Lord Howe Island and New Caledonia to New Zealand, across wide spans of ocean to Southern Australia, then to the tip of South Africa, and farther across the south Atlantic to the tip of South America. Galaxiids are the dominant freshwater fishes where they occur. There are eight genera and 40 species.

Salmonidae. Salmonids are distinguished from the osmerids and galaxiids by the following combination of characters: large adult size; more robust body; dorsal fin near the middle of the back, the origin of the dorsal fin is over or in advance of a vertical line drawn from the origin of the pelvic fins; presence of pelvic axillary process and adipose fin; gill membranes extending far forward and free from the



Brook trout (*Salvelinus fontinalis*). (*Photo courtesy of Johnny Jensen; © www.jjphoto.dk*)

isthmus; and parr with vertical bars in most species. The family consists of three subfamilies.

Coregoninae (whitefishes and ciscoes). Coregonids have fewer than 16 dorsal fin rays; large scales, fewer than 110 in lateral series; toothless maxilla; and usually small toothless vomer. They range from Asia and Siberia to northern North America. There are three genera and 32 species, three of which are endemic to Bear Lake in Utah-Idaho.

Thymallinae (graylings). A long, high dorsal fin with more than 17 rays distinguishes the graylings. Four species represent the single genus, *Thymallus*, in the freshwaters of Europe, Mongolia, northern Asia, and North America.

Salmoninae (salmons and trouts). The salmonids are distinguished by having fewer than 16 dorsal fin rays, small scales numbering 110 or more (usually many more), and maxillary teeth.

Generally the large seagoing anadromous species that return to their natal stream to spawn and die are called salmons, whereas the smaller species that live to spawn several seasons are called trouts (see **illustration**). All species spawn in freshwater. An ultrasensitive sense of smell and the ability to fix in their memory the odors of streams they traverse on their journey to sea serve as a "road map" for salmons to return to their place of birth.

Salmons and trouts are Holarctic in native distribution; however, species such as *Salvelinus fontinalis*, *Salmo trutta*, and *Oncorhynchus mykiss* (long known as *Salmo gairdneri*) have been introduced throughout the world. Worldwide salmons and trouts are probably more studied, more fished for recreation purposes, more exploited commercially, more genetically manipulated, and more introduced in foreign countries than any other group of fishes. These fishes are very important economically. *See* AQUACULTURE.

There are seven genera and about 30 species. Ongoing studies will probably add to the numbers.

Herbert Boschung

Bibliography. R. J. Behnke, *Native Trout of Western North America.* Amer. Fish. Soc. Monog. 6, American Fisheries Society, Bethesda, Maryland, 1992; R. J. Behnke, *Trout and Salmon of North America*, Free Press, New York, 2002; T. M. Berra, *An Atlas of Distribution of Freshwater Fish Families of the World*, University of Nebraska Press, Lincoln, 1981; R. H. Smith, *Native Trout of North America*, 2d ed., Frank Amato Publications, Portland, Oregon, 1994.

Salt (chemistry)

A compound formed when one or more of the hydrogen atoms of an acid are replaced by one or more cations of the base. The common example is sodium chloride in which the hydrogen ions of hydrochloric acid are replaced by the sodium ions (cations) of sodium hydroxide. There is a great variety of salts because of the large number of acids and bases which has become known.

Classification. Salts are classified in several ways. One method—normal, acid, and basic salts—depends upon whether all the hydrogen ions of the acid or all the hydroxide ions of the base have been replaced:

Class	*Examples*
Normal salts	$NaCl$, NH_4Cl, Na_2SO_4, Na_2CO_3, Na_3PO_4, $Ca_3(PO_4)_2$
Acids salts	$NaHCO_3$, NaH_2PO_4, Na_2HPO_4, $NaHSO_4$
Basic salts	$Pb(OH)Cl$, $Sn(OH)Cl$

The other method—simple salts, double salts (including alums), and complex salts—depends upon the character of completeness of the ionization:

Class	*Examples*
Simple salts	$NaCl$, $NaHCO_3$, $Pb(OH)Cl$
Double salts	$KCl \cdot MgCl_2$
Alums	$KAl(SO_4)_2$, $NaFe(SO_4)_2$, $NH_4Cr(SO_4)_2$
Complex salts	$K_3Fe(CN)_6$, $Cu(NH_3)_4Cl_2$, $K_2Cr_2O_7$

In general, all salts in solution will give ions of each of the metal ions; an exception is the complex type of salt such as $K_3Fe(CN)_6$ and $K_2Cr_2O_7$. In such salts the ionization is entirely as shown by reactions (1) and (2). No detectable quantities of Fe^{3+} or Cr^{3+} from

$$K_3Fe(CN)_6 \rightarrow 3K^+ + Fe(CN)_6{}^{3-} \quad (1)$$

$$K_2Cr_2O_7 \rightarrow 2K^+ + Cr_2O_7{}^{3-} \quad (2)$$

these salts exist in solution because of the strong bonding of these ions in the complex ions. However, in those complex salts where the bonding is weak, ions of the metal can be detected; for example, in Na_2CdCl_4, the cadmium complex ion ionizes appreciably as in reaction (3). The elements with unfilled

$$CdCl_4{}^{2-} \rightarrow Cd^{2+} + 4Cl^- \quad (3)$$

inner electron shells form complex salts readily. The alum type of salt is a sulfate including the univalent

cation of a relatively strong base and a trivalent metal ion such as Al^{3+}, Fe^{3+}, or Cr^{3+}.

Double salts include ions of nearly enough the same size to fit into the same crystal lattice.

Hydrolysis. The solutions of some normal salts are neutral, but those of others are acidic or basic. This results from the reactor of the ions of salt with water. This reaction is called hydrolysis. Examples are shown in reactions (4)–(6). The resulting solu-

$$CO_3^{2-} + HOH \rightleftharpoons HCO_3^- + OH^- \quad (4)$$

$$CH_3COO^- + HOH \rightleftharpoons CH_3COOH + OH^- \quad (5)$$

$$NH_4^+ + HOH \rightleftharpoons NH_4OH + H^+ \quad (6)$$

tion will be acidic or basic, depending upon whether the hydrolysis produces an excess of hydrogen or hydroxide ion. *See* HYDROLYSIS.

Theories of acids and definition of salts. The development of more general theories of acids and bases in the twentieth century has required a broadening of the concept of salts.

The Brönsted theory of acids lays emphasis on the process of the reaction between acids and bases, and not so much on the product except that the products are other acids and bases, as in reactions (7) and (8).

$$acid_1 + base_2 \rightleftharpoons acid_2 + base_1 \quad (7)$$

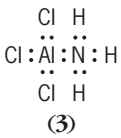
$$HCl + H_2O \rightleftharpoons H_3O^+ + Cl^- \quad (8)$$

Since the Brönsted theory extends the proton theory of acids to solvents other than water, the original definition of a salt must be expanded as follows: A salt is an electrovalent compound that contains some cation other than the solvated proton and some anion other than the anion which is the conjugate base of the solvent. In the water system the salt should not contain the H_3O^+ and OH^- ions alone; in the liquid ammonia system it should not contain the NH_4^+ and NH_2^- ions alone.

In terms of the Lewis theory of acids and bases, compound (**1**) is an acid, and compound (**2**) is a base. Hence the salt should be compound (**3**). Here the salt

(**1**)

(**2**)



(**3**)

is not limited to replacement of the H^+ with a cation of a base. Rather, it is any aggregate of molecules, atoms, or ions joined together with a coordinate covalent bond. Such compounds correctly can be called salts; however, by common parlance, the term salt usually refers to an electrovalent compound, the classical example of which is sodium chloride. *See* ACID AND BASE; CHEMICAL BONDING. Alfred B. Garrett

Bibliography. S. Radel and M. Navidi, *Chemistry*, 2d ed., 1994.

Salt (food)

The chemical compound sodium chloride. It is used extensively in the food industry as a preservative and flavoring, as well as in the chemical industry to make chlorine and sodium. Historically, salt is one of the oldest materials used in people's food. *See* CHLORINE; SALT (CHEMISTRY); SODIUM.

Manufacture. Salt was originally made by evaporating seawater (solar salt). This method is still in common usage; however, impurities in solar salt make it unsatisfactory for most commercial uses, and these impurities also lead to clumping. Salt, freshly produced from seawater evaporation ponds, may contain large numbers of halophilic (salt-loving) microorganisms. These occasionally cause spoilage of meat, fish, vegetables, and hides when salt has been used in preservation.

In the United States refined salt is obtained from underground mines located in Michigan and Louisiana. Salt is usually handled during the refining processes as brine. These processes are discussed below.

Grainer salt. This type of salt is made by evaporation of brine in long shallow pans, as large as 18 ft (5.5 m) wide, 1.5 ft (4.6 m) deep, and 150 ft (46 m) long. The daily capacity of such a grainer may be 80 tons (73 metric tons). A scraping conveyor continually removes the crystallizing salt from the bottom of the grainer. The salt is then filtered, dewatered, dried, cooled, and rolled to break clumps. Grainer salt is usually the coarsest in grain and highest in impurities.

Vacuum pan salt. Salt brine is boiled at reduced pressure. A triple-effect evaporator is used; the first stage uses relatively light vacuum, but this is increased until in the third it is quite high and the salt solution boils at about 110°F (43°C). Production is continuous and the production cycle takes 48 h. A 20% salt slurry is brought out from the bottom of the third evaporator at the same time fresh brine is admitted; thus impurities are washed from the surface of the outgoing crystals. The salt slurry is filtered, dewatered, and high-temperature-dried at 350°F (177°C) before screening and packing.

Alberger process. Salt brine is heated to high pressures in heaters and then is passed to a graveller. A graveller is a large cylindrical vessel filled with stones which serves as a deposition site for calcium sulfate. The brine proceeds to flashers, where the pressure is gradually reduced to that of the atmosphere, and salt begins to crystallize. The brine and salt mixture is discharged to a large open pan; the crystallized salt is pumped to a centrifuge for dewatering, then dried in a rotary dryer.

Use. Large users of salt in the food industry are pickle makers and meat packers. In the pickle industry salt is used as brine to which fresh cucumbers are added. A selective fermentation then proceeds

which is governed by salt concentration. Cost and time are dictating a movement toward "fresh-pack" pickles, in which salt and spices are added with the fresh cucumber, the jar is sealed, and the whole package is heated for pasteurization preservation. In the meat packing industry salt is added to fresh meat as a preservative, as in salt pork, or along with nitrates or nitrites as the first step in the production of cured meats, such as hams or bacon. *See* FOOD ENGINEERING.

Additives. Salt is liable to clumping during periods of high humidity, so preventives are added. Materials used include magnesium carbonate and certain silicates. Iodides are added in those areas where iodine deficiencies exist.

Roy E. Morse

Bibliography. T. P. Coultate (ed.), *Food: The Chemistry of Its Components*, 3d ed., 1996; B. Kraus, *Complete Guide to Sodium*, 1990; Y. Pomeranz and C. E. Meloan, *Food Analysis: Theory and Practice*, 3d ed., 1994.

Salt dome

An upwelling of crystalline rock salt and its aureole of deformed sediments. A salt pillow is an immature salt dome comprising a broad salt swell draped by concordant strata. A salt stock is a more mature, pluglike diapir of salt that has pierced, or appears to have pierced, overlying strata. Most salt stocks are 0.6–6 mi (1–10 km) wide and high. Salt domes are closely related to other salt upwellings (**Fig. 1**), some of which are much larger. Salt canopies, which form by coalescence of salt domes and tongues, can be more than 200 mi (300 km) wide. *See* DIAPIR.

Distribution. Exploration for oil and gas has revealed salt domes in more than 100 sedimentary basins that contain rock salt layers several hundred meters or more thick. The salt was precipitated from evaporating lakes in rift valleys, intermontaine basins, and especially along divergent continental margins. Salt domes are known in every ocean and continent (except little-explored Antarctica) but are scarce in the Pacific Ocean, where few salt-bearing basins have escaped destruction by plate convergence. *See* BASIN; PLATE TECTONICS.

Composition. Salt domes consist largely of halite (NaCl, common table salt). Other evaporites, such as anhydrite ($CaSO_4$), gypsum ($CaSO_4 \cdot 2H_2O$), and more rarely sylvite (KCl), carnallite ($KCl \cdot MgCl_2 \cdot 6H_2O$), and other potash and magnesian sulfates and chlorides, form thinner layers within the rock salt. Nonevaporite inclusions range from disseminated grains to massive rafts 2 mi (4 km) wide. Salt is highly soluble, so it is remarkable that salt as much as 800 million years old is still preserved in domes in Iran, Oman, and Pakistan. *See* HALITE; SALINE EVAPORITES.

Subsurface dissolution by ground water leaves a compacted residue, known as cap rock, which mantles the crest and shoulders of the salt dome. Residual anhydrite typically undergoes reactions involving water, hydrocarbons, and bacteria to yield gypsum, calcite ($CaCO_3$), sulfur, and sulfides of iron, zinc, lead, and barium. *See* ANHYDRITE; CALCITE; GYPSUM.

Formation. Crystalline salt flows by creep as a viscous or power-law fluid. Creep is promoted by fine

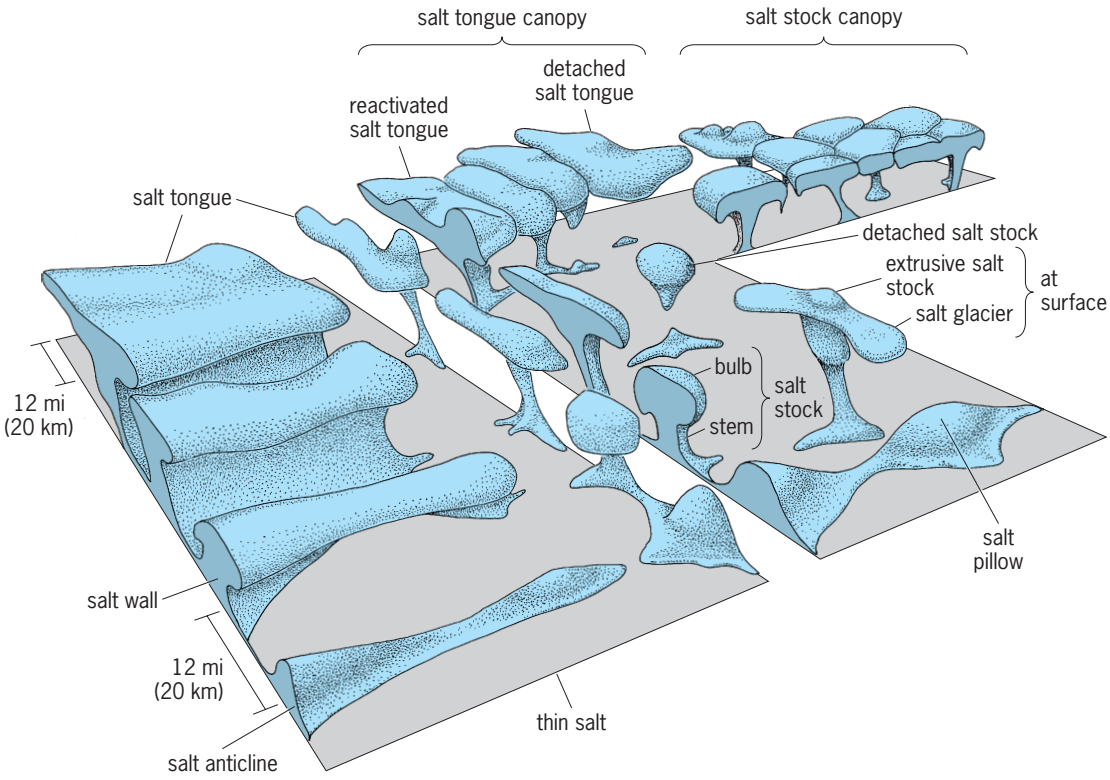


Fig. 1. Salt domes and related structures; the overlying and surrounding sediments have been removed for clarity.

grain size, heat, and traces of water. Salt acts as a weak but pressurized geologic lubricant that allows stronger strata above and below to deform partly independently of each other. Salt thus facilitates deformation, whether driven by plate tectonics or merely by gravity (the process of halokinesis). *See* CREEP (MATERIALS).

Salt diapirs are initiated and driven upward by differential loading: laterally varying sedimentary loads are created by erosion, deformation, or uneven sedimentation. As salt diapirs pierce overlying strata, they become discordantly encased in brittle rocks. Three piercement modes are recognized: reactive, active, and passive (**Fig. 2**). Reactive diapirs are emplaced while a basin is stretched (Fig. 2*a*). Salt upwells into structurally thinned zones, such as grabens. Reactive rise stops when regional stretching ceases. If this diapir becomes tall enough and its roof strata thin enough, the diapir becomes active (Fig. 2*b*). The pressure of salt is enough to break through and shoulder aside the thin roof. Salt then emerges under air or water (Fig. 2*c*). This passive diapir remains at or near the surface while sediments accumulate around it. The passive diapir grows taller because its base sinks below the surface along with the surrounding strata, in a process known as downbuilding. A passive diapir is a slow-moving fountain that can extrude, even under water, to form salt glaciers. These are flowing today in the Zagros Mountains of Iran. The many salt tongues and sheets in the Gulf of Mexico are thought to have originated as submarine salt glaciers, partly protected from dissolution by residual cap rock and veneers of marine sediment. Spreading salt diapirs can coalesce to form a variety of canopies (Fig. 1). These buried extrusions can act as source layers for younger generations of salt domes. The lower parts of salt walls can be squeezed together by compression: the salt is displaced upward, where it bulges its roof. The closing sides of the squeezed salt wall can eventually meet to form a salt weld in which only a smear of salt remains or vanishes entirely. *See* GRABEN.

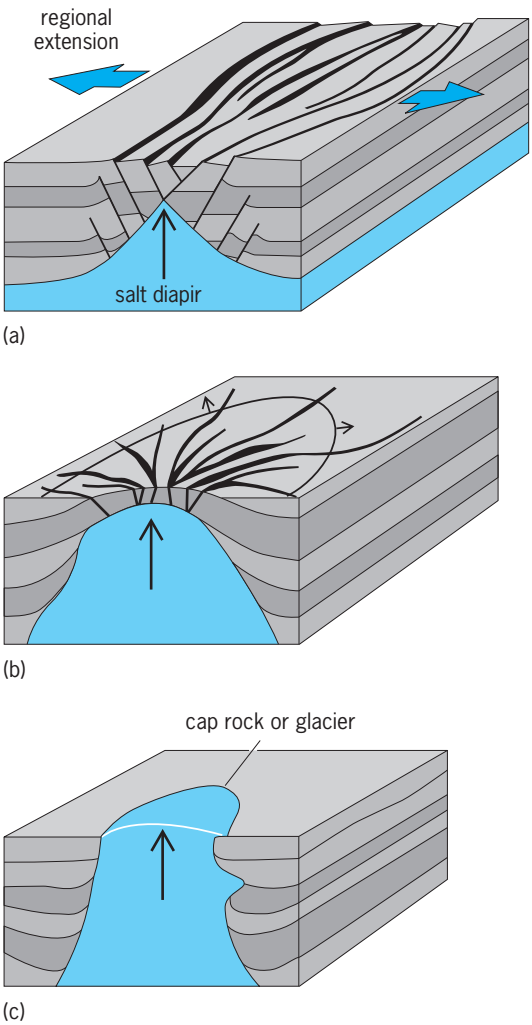


Fig. 2. Three modes of salt diapirism and their characteristic structures. (*a*) Reactive piercement. (*b*) Active piercement. (*c*) Passive piercement.

Detection. Salt domes exposed at the surface have been known for millennia as sources of salt, a historically prized commodity. In the search for salt or hydrocarbons, buried salt domes were discovered by increasingly sophisticated means: by searching for mounds, saline or sulfurous springs, and hydrocarbon seeps; by drilling; by conducting gravity surveys that respond to the low density of salt and the high density of cap rock, and magnetic surveys, where salt is less magnetic than surrounding sediments; and by seismic refraction and reflection surveys that recognize salt by its high sonic velocity and lack of internal reflections. *See* PETROLEUM GEOLOGY; PROSPECTING; SEISMIC EXPLORATION FOR OIL AND GAS.

Economic importance. Salt domes supply industrial commodities, including fuel, minerals, chemical feedstock, and storage caverns. Giant oil or gas fields are associated with salt domes in many basins around the world, especially in the Middle East, North Sea, and South Atlantic regions. The most intensively explored of these basins is the United States Gulf Coast and Gulf of Mexico, where 80% of the oil and gas fields are associated with salt structures and form the United States' largest exploration province. Salt canopies obscure vast regions of little-explored potential hydrocarbons. The 1901 oil gusher at Spindletop salt dome in Texas initiated the world's continuing dependence on—and exploration for—prolific liquid fossil fuels. Oil and gas are trapped around salt domes by arching, fracturing, or faulting of surrounding strata, or by sedimentary facies changes caused by dome rise, and are sealed in traps by the impermeable salt itself.

The cap rock of salt domes provides the main source of elemental sulfur, which is readily extracted by the Frasch process. Salt domes on the United States Gulf Coast have yielded 65% of the world's sulfur production. About 65% of the United States salt is mined as rock or as brine dissolved from salt domes. Potassium salts in German salt domes are a major source of potash for fertilizers. *See* FERTILIZER; SULFUR.

Salt domes are also used to store crude oil, natural gas (methane), liquefied petroleum gas, and

radioactive or toxic wastes. Bottle-shaped caverns up to 2000 ft (600 m) tall can be created by dissolving salt with water injected down a well into the dome and extracting the resulting brine. More than 500 million barrels of crude oil are stored in United States salt domes as an emergency supply called the Strategic Petroleum Reserve. *See* OIL AND GAS STORAGE.

M. P. A. Jackson

Bibliography. G. I. Alsop, D. J. Blundell, and I. Davison, Salt Tectonics, *Geol. Soc. London Spec. Publ.*, no. 100, 1996; M. T. Halbouty, *Salt Domes: Gulf Region, United States and Mexico*, 2d ed., 1979; M. P. A. Jackson, D. G. Roberts, and S. Snelson (eds.), Salt Tectonics: A Global Perspective, *Amer. Ass. Petrol. Geol. Mem.*, no. 65, 1995; M. P. A. Jackson et. al., Structural dynamics of salt systems, *Annu. Rev. Earth Planet. Sci.*, 1994.

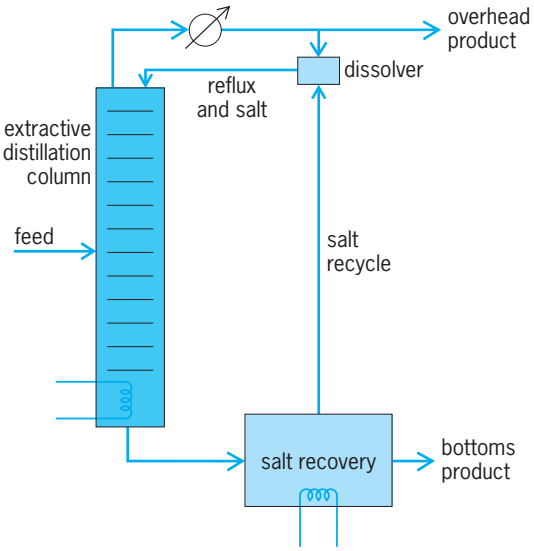
Salt-effect distillation

A process of extractive distillation in which a salt that is soluble in the liquid phase of the system being separated is used in place of the normal liquid additive introduced to the extractive distillation column in order to effect the separation.

Process. Extractive distillation is a process used to separate azeotrope-containing systems or systems in which relative volatility is excessively low. An additive, or separating agent, that is capable of raising relative volatility and eliminating azeotropes in the system being distilled is supplied to the column, where it mixes with the feed components and exerts its effect. The agent is subsequently recovered from one or both product streams by a separate process and recycled for reuse. In conventional extractive distillation, the agent is a liquid, usually one having fairly low volatility. An example is ethanol-water separation using ethylene glycol as the added agent. *See* AZEOTROPIC MIXTURE.

In salt-effect distillation, the flow sheet is essentially the same as for a liquid agent, although the subsequent process used to recover the agent for recycling is different; that is, evaporation is used rather than distillation. A typical flow sheet is shown in the **illustration**. The salt is added to the system by being dissolved in the reentering reflux stream at the top of the column. Being nonvolatile, it will reside in the liquid phase, flowing down the column and out in the bottom product stream. (An alternative method would be to use regular distillation up to the azeotrope composition, salt-effect distillation to cross it, and regular distillation to complete the separation.)

Liquid and salt agents work in a similar manner: their molecules or ions exert attractive forces, which result in the formation of association complexes in the liquid phase with the molecules of the feed components. An agent is used whose attraction is preferential toward the molecules of the less volatile component of the system being separated. Thus, the volatility of the higher-boiling component is reduced by an amount greater than that of the lower-boiling component. As a result, the relative volatility is increased, and the azeotrope is eliminated if the effect is sufficiently strong and sufficiently selective. *See* COORDINATION COMPLEXES.



Flow sheet for salt-effect distillation. (*After W. F. Furter, Salt effect in distillation: A technical review, Chem. Eng., 219:173–177, 1968*)

Advantages and disadvantages. The advantage of salts over liquids as separating agents is that the forces of attraction exerted by their ions toward molecules of the higher-boiling feed component may be much stronger than those that can be exerted by the molecules of a liquid agent. With a salt, a much smaller quantity of separating agent may have to be fed, recovered, and recirculated for a particular separation than in the case of a liquid. This results in savings in costs of energy and capital, particularly in the separating-agent recovery process. For example, 6 mol % of potassium acetate present in the liquid phase will completely eliminate the ethanol-water azeotrope, whereas conventional azeotropic distillation of this system using a liquid agent such as benzene or *n*-pentane may require an agent-to-feed ratio ranging anywhere from 1:1 to 5:1.

A second advantage is that the salt, being completely nonvolatile, does not distribute itself partially into the vapor phase in the extractive distillation column. Therefore, no provision is required to recover agent from the overhead product stream. With a liquid agent, however, a so-called knockback section of column is usually required above the agent's feed point to recover that portion of the agent that has vaporized in the column and to prevent it from contaminating the overhead product.

The main disadvantage of a salt over a liquid agent is found in solubility limitations. It is generally much more difficult to find a salt that is both effective and soluble in the system being distilled than it is to find a liquid with these properties. Yet, in that limited number of systems for which a salt exists that has a strong and selective effect and is sufficiently soluble, the advantages of a salt over a liquid can be substantial, particularly in the saving of energy.

Applications. The first large-scale use of salt-effect distillation was in producing ethanol for blending

with gasoline to produce gasohol. Most gasoline sold in Europe from the late 1920s to World War II contained 10% ethanol, which was produced from sugarbeets and refined by the HIAG process, a salt-effect distillation that uses a mixture of potassium and sodium acetates as the separating agent. Over 100 HIAG plants were built worldwide, with the last ceasing operation in 1965, in Brazil. Interest in salt-effect distillation revived in the late 1970s and throughout the 1980s among gasohol producers in the United States and elsewhere. *See* ALCOHOL FUEL.

The major commercial use of salt-effect distillation is in the concentration of aqueous nitric acid, using the salt magnesium nitrate as the separating agent instead of sulfuric acid, which was the liquid agent used in earlier extractive distillation processes. Other commercial applications exist, including acetone-methanol separation using calcium chloride and isopropanol-water separation using the same salt.

Salt-effect distillation has not achieved a wider use in the past for several reasons: the chemical theory is very complex; engineers much prefer handling of fluids over handling of solids when there is a choice; and, most importantly, when energy is cheap there is little incentive for improved energy efficiency. With its strong advantages in energy economy over conventional extractive distillation, salt-effect distillation can be expected to grow as energy becomes increasingly costly. *See* AZEOTROPIC DISTILLATION; DISTILLATION.

William F. Furter

Bibliography. W. F. Furter, Salt effect in distillation, Pt. II, *Can. J. Chem Eng.*, 55:229–239, 1977; H. Z. Kister, *Distillation Design*, 1992.

Salt gland

A specialized gland located around the eyes and nasal passages in marine turtles, snakes, and lizards, and in birds such as the petrels, gulls, and albatrosses, which spend much time at sea. Salt glands are compound tubular glands which apparently arise as invaginations of the epithelium of the nasal passage or from the developing conjunctival sac of the eye. In the marine turtle it is an accessory lacrimal gland which opens into the conjunctival sac. In seagoing birds and in marine lizards it opens into the nasal passageway. Salt glands copiously secrete a watery fluid containing a high percentage of salt, higher than the salt content of urine in these species. As a consequence, these animals are able to drink salt-laden seawater without experiencing the dehydration necessary to eliminate the excess salt via the kidney route. *See* EXCRETION; GLAND.

Olin E. Nelsen

Salt marsh

A maritime habitat characterized by grasses, sedges, and other plants that have adapted to continual, periodic flooding. Salt marshes are found primarily throughout the temperate and subarctic regions.

Fig. 1. Plant zonation in a salt marsh, Canary Creek Marsh, Lewes, Delaware: smooth cordgrass (*Spartina alterniflora*) at the left; salt-meadow hay in the middle; and marsh elder (*Iva frutescens*) and groundsel tree (*Baccharis halimifolia*) at the upland edge on the right. (*Courtesy of F. C. Daiber*)

The tide is the dominating characteristic of a salt marsh. The salinity of the tide defines the plants and animals that can survive in the marsh area. The vertical range of the tide determines flooding depths and thus the height of the vegetation, and the tidal cycle controls how often and how long vegetation is submerged. Two areas are delineated by the tide: the low marsh and the high marsh. The low marsh generally floods and drains twice daily with the rise and fall of the tide; the high marsh, which is at a slightly higher elevation, floods less frequently. *See* MANGROVE.

Marsh formation. Salt marshes usually are developed on a sinking coastline, originating as mud flats in the shallow water of sheltered bays, lagoons, and estuaries, or behind sandbars. Where tidal flow is impeded, sediments accumulate with the help of submerged vegetation. Eventually, the surface becomes exposed at low tide, stranding seeds or fragments of plant roots. The first colonizers are *Spartina*, *Salicornia*, and *Puccinellia;* and with continuing sedimentation, the soil matures and secondary invaders such as *Spartina patens* (salt-meadow hay) and *Distichlis spicata* (salt grass) arrive. The marsh surface then takes on a slope, with the landward edge higher, and finally plant zonation becomes evident (**Figs. 1** and **2**). *See* ESTUARINE OCEANOGRAPHY.

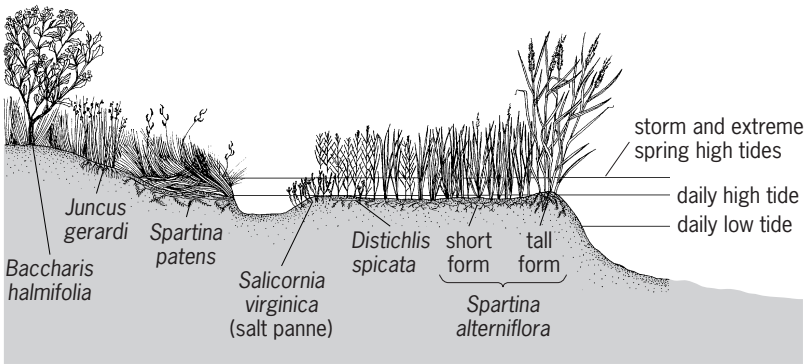


Fig. 2. Salt-marsh profile of the Delaware estuary. Such salt marshes occur primarily in the lower estuary, where they are dominated by smooth cordgrass (*Spartina alterniflora*). The typical profile from creek bank to upland features a levee next to the creek bank that is colonized by the tall form of *Spartina*. Away from the creek bank, the broad expanse of marsh is dominated by short-form *Spartina* intermixed with several subdominant species, including salt-meadow hay (*Spartina patens*) and glasswort (*Salicornia virginica*). (*Courtesy of Karen Grosz*)

Physiography. The primary physiographic features of salt marshes are creeks and pans. Typically, during the development of mud flats, minor surface irregularities cause water to be deflected into channels. Once channels are formed, scouring takes place, and the constant flooding and ebbing of the tide prevents vegetation from colonizing them. Salt-marsh creeks are an example of a stream bed produced by two conflicting processes—sedimentation and erosion.

In the early stages of marsh development, plant colonization is irregular, and parts of the bare surface, or pans, become surrounded by vegetation. As the marsh surface rises, such bare spots lose their outlets for tidal water. Filled with water, pans prevent the growth of vegetation, but if they are drained, vegetation invades. The so-called rotten spot is an irregularly shaped pan formed by mats of debris deposited by storm tides. This debris, combined with excessive salinity, kills the vegetation below.

Plants. Salt marshes are formed where salinity is high, ranging from 20 to 30 parts per thousand of sodium chloride. Proceeding up the estuary, there is a transitional zone where salinity ranges from 20 to less than 5 ppt. In the upper estuary, where river input dominates, the water has only a trace of salt. This varying salinity produces changes in the marsh—in the kinds of species and also in their number. Typically, the fewest species are found in the salt marsh and the greatest number in the fresh-water tidal marsh.

The salt marshes along the Atlantic shore of the temperate latitudes of North America are dominated by smooth cordgrass (*Spartina alterniflora*) and salt-meadow hay, and at the landward edge by black grass (*Juncus gerardi* or *J. roemarianus*), marsh elder (*Iva frutescens*), and groundsel tree (*Baccharis halimifolia*). The brackish-water marsh is dominated by big cordgrass (*Spartina cynosuroides*) and three-square sedge (*Scirpus americanus*). The fresh-water marsh is dominated by spadder dock (*Nuphar lutem*), pickerelweed (*Pontederia cordata*), broadleaf arrowhead (*Sagittaria latifolia*), and water hemp (*Amaranthus canabinus*).

Animals. Animals—from protozoa to mammals—abound in salt marshes. Just as the plants do, these creatures display a zonation determined by salinity, frequency, and duration of tidal submergence; examples are the ribbed mussel (*Geukensia demissa*), salt-marsh snail (*Melampus bidentatus*), and salt-marsh mosquito (*Aedes sollicitans*). The fauna of a tidal marsh can be placed in three major subdivisions: marine, terrestrial, and marsh species. The marine species, such as the bay anchovy (*Anchoa mitchilli*) and blue crab (*Callinectes sapidus*), have a center of distribution in the estuary. The terrestrial species, such as the meadow vole (*Microtus pennsylvanicus*) and raccoon (*Procyon lotor*), are air breathers; they live on the marsh or at the landward edge and make excursions onto the marsh. The marsh species are unique to the marsh. Most invertebrate species, such as the ribbed mussel, salt-marsh snail, and fiddler crab (*Uca* sp.), are derived from marine ancestors with aquatic larvae; as part of their life cycle, they require two or more adjacent but different ecological communities. Other marsh species include the greenhead fly (*Tabanus nigrovittatus*), mummichog (*Fundulus heteroclitus*), clapper rail (*Rallus longirostris*), seaside sparrow (*Ammospiza maritima*), and, in brackish-water marshes, the muskrat (*Ondatra zibethica*).

The large variety of plants and animals in the marsh leads to an assortment of plant–animal relationships. Direct interactions include trampling of vegetation and feeding, reproductive, and residential activities. Indirect interactions, which become evident some time after the occurrence of plant–animal interplay, include nutrient cycling and seed dispersal. The greater the intensity of interaction, the sooner and the greater the effect.

Productivity. The salt marsh is one of the most productive ecosystems in nature. This productivity is made possible by the involvement of solar energy in the photosynthetic process of higher rooted plants of the marsh and the algae growing on the surface muds. In addition to solar energy, tidal energy repeatedly spreads nutrient-enriched waters over the marsh surface, subsidizing the production process.

Some of this enormous supply of live plant material may be consumed by marsh animals, but the most significant values are realized when the vegetation dies and is decomposed by microorganisms to form detritus. Dissolved organic materials are released, providing an essential energy source for bacteria that mediate wetland biogeochemical cycles (carbon, nitrogen, and sulfur cycles). *See* BIOGEOCHEMISTRY.

Detritus helps form the base of the marsh–estuarine food web. It is fed upon directly by detritivores, small invertebrates and fishes living in the tidal creeks, which are in turn fed upon by coastal fishes such as menhaden (*Brevoortia tyrannus*), croaker (*Micropogonias undulatus*), spot (*Leiostomus xanthurus*), and summer flounder (*Paralichthys dentatus*), which spend a portion of their larval and juvenile stages in tidal creeks or in the adjoining estuary. These fishes then move out of the tidal creeks and estuary into the coastal waters, where they in turn are consumed by larger predators. This sequence represents an important transfer of marsh-derived production to the coastal ocean. *See* BIOLOGICAL PRODUCTIVITY; FOOD WEB.

Roles. The salt marsh has two general roles: one is human-based and the other is natural. It has been a source of hordes of biting flies, of pasturage for livestock, and of income from agricultural crops, shellfish, fish, animal pelts, gravel, and clay. It has been ditched, drained, impounded, and filled with trash from human activity, and it has served as a site for homes and factories.

The salt marsh also serves as a sediment sink, a nursery habitat for fishes and crustaceans, a feeding and nesting site for waterfowl and shorebirds, a habitat for numerous unique plants and animals, a nutrient source, a reservoir for storm water, an erosion control mechanism, and a site for esthetic

pleasures. An understanding of and an appreciation for the importance of salt marshes has led to federal and state legislation aimed at their protection. *See* PLANTS, SALINE ENVIRONMENTS OF; WATER CONSERVATION.

Franklin C. Daiber

Bibliography. P. Adam, *Saltmarsh Ecology*, 1990; R. H. Chabreck, *Coastal Marshes: Ecology and Wildlife Management*, 1988; V. J. Chapman, *Salt Marshes and Salt Deserts of the World*, 2d suppl. reprint ed., 1974; V. J. Chapman, *Wet Coastal Ecosystems*, 1977; F. C. Daiber, *Animals of the Tidal Marsh*, 1982; F. C. Daiber, *Conservation of Tidal Marshes*, 1986; L. R. Pomeroy and R. G. Wiegert (eds.), *The Ecology of a Salt Marsh*, 1981; D. S. Ranwell, *Ecology of Salt Marshes and Sand Dunes*, 1972.

Salviniales

A small order of heterosporous, leptosporangiate ferns (division Polypodiophyta) which float on the surface of the water. The delicate, branching stem becomes as much as 1 in. (2.5 cm) or so long and is provided with small, simple to bifid or more or less dissected leaves. The sporangia are enclosed in specialized appendages of the leaves, called sporocarps. A sporocarp contains numerous microsporangia or a single megasporangium, but not both types. The order contains only a single family, with two widely distributed genera, *Salvinia* and *Azolla*, with only about 20 species in all. *Azolla* is of some interest because of its symbiotic relationship with *Anabaena azollae*, a nitrogen-fixing blue-green alga which inhabits specialized chambers in the leaves. *See* POLYPODIOPHYTA.

Arthur Cronquist

Samarium

A chemical element, Sm, atomic number 62, belonging to the rare-earth group. Its atomic weight is 150.35, and there are seven naturally occurring isotopes; ^{147}Sm, ^{148}Sm, and ^{149}Sm are radioactive and emit α particles. *See* PERIODIC TABLE.

Samarium oxide is pale yellow, is readily soluble in most acids, and gives topaz-yellow salts in solutions. Samarium has found rather limited use in the ceramic industry, and it is used as a catalyst for certain organic reactions. One of its isotopes has a very high cross section for the capture of neutrons, and therefore there has been some interest in samarium in the atomic industry for use as control rods and nuclear poisons. *See* LANTHANUM; RARE-EARTH ELEMENTS.

Frank H. Spedding

Bibliography. F. A. Cotton et al., *Advanced Inorganic Chemistry*, 6th ed., Wiley-Interscience, 1999; N. Kaltsoyannis and P. Scott (eds.), *The F Elements (Oxford Chemistry Primers, 76)*, Oxford University Press, 1999.

1												13	14	15	16	17	18
1 H	2											5 B	6 C	7 N	8 O	9 F	2 He
3 Li	4 Be											13 Al	14 Si	15 P	16 S	17 Cl	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	31 Ga	32 Ge	33 As	34 Se	35 Br	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	49 In	50 Sn	51 Sb	52 Te	53 I	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	81 Tl	82 Pb	83 Bi	84 Po	85 At	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg						86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

Sampled-data control system

A type of digital control system in which one or more of the input or output signals is a continuous, or analog, signal that has been sampled. There are two aspects of a sampled signal: sampling in time and quantization in amplitude. Sampling refers to the process of converting an analog signal from a continuously valued range of amplitude values to one of a finite set of possible numerical values. This sampling typically occurs at a regular sampling rate, but for some applications the sampling may be aperiodic or random. The sampling process consists of an ideal switch which passes the time-varying analog signal (**Fig. 1***a*) to a hold device which stores the instantaneous analog amplitude of the signal while it is converted to a binary number. The ideal sampler allows the signal to pass to the hold device at regular intervals for a very short period of time. The effective

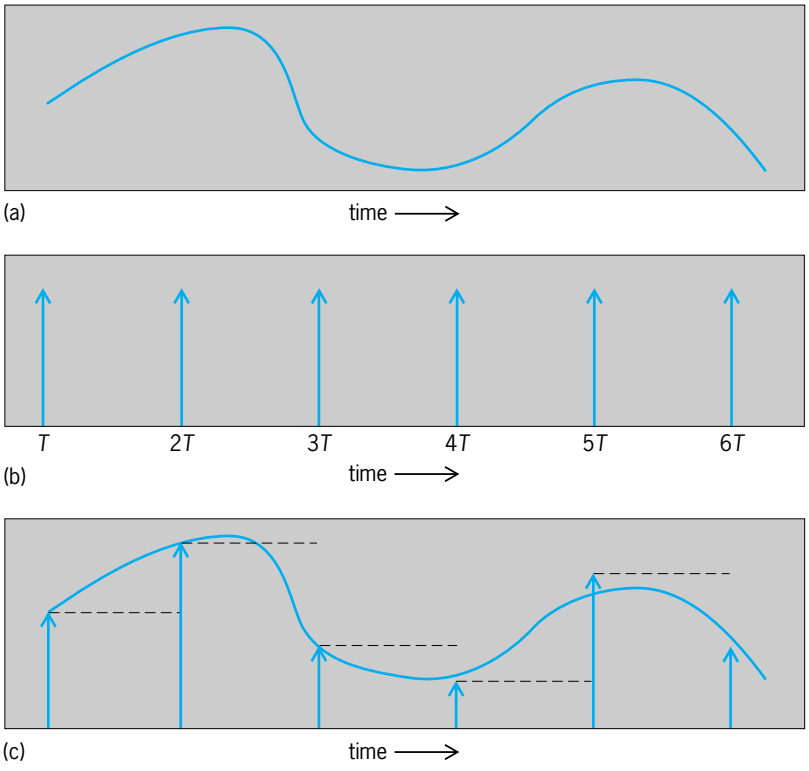


Fig. 1. Sampling process. (*a*) Continuous, analog signal to be sampled. (*b*) Sampling signals, corresponding to an ideal sampler that allows the signal to pass to the hold device at regular intervals of *T* seconds. (*c*) Resulting quantized values occurring at discrete instants of time. The horizontal lines represent the values of the analog signal as seen by the digital compensator, as well as the typical output of a digital-to-analog converter.

sampling signal is thus a series of delta functions (Fig. 1*b*). The resulting sampled signal (Fig. 1*c*) has the appearance of a pulse-amplitude-modulated signal and, for theoretical analysis, is often considered to be the product of a sequence of Kronecker delta functions with the original analog signal. *See* PULSE MODULATION.

A sampled-data control system can be decomposed into four components (**Fig. 2**): the plant which is to be controlled, the analog-to-digital converter, the digital device which implements the compensator, and the digital-to-analog converter.

Plant. While the device to be controlled is usually referred to as the plant, sampled-data control systems are also used to control processes. The term plant refers to machines or mechanical devices which can usually be mathematically modeled by an analysis of their kinematics, such as a robotic arm or an engine. A process refers to a system of operations such as a batch reactor for the production of a particular chemical, or the operation of a nation's economy. The output of the plant which is to be controlled is called the controlled variable. A regulator is one type of sampled-data control system, and its purpose is to maintain the controlled variable at a preset value (for example, the robotic arm at a particular position, or an airplane turboprop engine at a constant speed) or the process at a constant value (for example, the concentration of an acid, or the inflation rate of an economy). This input is called the reference or setpoint. The second type of sampled-data control system is a servomechanism, whose purpose is to make the controlled variable follow an input variable. Examples of servomechanisms are a robotic arm used to paint automobiles which may be required to move through a predefined path in three-dimensional space while holding the sprayer at varying angles, an automobile engine which is expected to follow the input commands of the driver, a chemical process that may require the pH of a batch process to change at a specified rate, and an economy's growth rate which is to be changed by altering the money supply. *See* PROCESS CONTROL; REGULATOR; SERVOMECHANISM.

Analog-to-digital converter. The analog-to-digital converter changes the sampled signal into a binary number so that it can be used in calculations by the digital compensator. The word length (number of bits of resolution) of the analog-to-digital converter limits the fundamental accuracy of the control system and determines its maximum speed of operation. Typical word lengths are 10–12 bits, but greater accuracies are available for some applications. For example, a 10-bit analog-to-digital converter quantizes the analog signal into 2^{10}, or 1024, discrete levels, which yields approximately 0.1% resolution. The conversion of a continuous valued signal into one of 2^n allowed values (where n is the word length) creates the equivalent of an additive noise called quantization noise. This noise is the difference in amplitude between the continuous signal and the amplitudes of the sampled signal (Fig. 1*c*). The word length of the analog-to-digital converter and subsequent computations is selected to keep this noise to an acceptable level.

The analog-to-digital converter also sets the maximum speed of operation of the sampled-data control system since it takes some time, usually microseconds, to effect the conversion. The Nyquist criterion requires that a system be sampled at greater than twice its maximum frequency component in order to properly represent a signal. Since the analog-to-digital converter is used to convert plant feedback signals, it is the dynamics of the plant which determine the minimum acceptable sampling rate, and this is reflected in the selection of the analog-to-digital converter conversion speed. Typical practical sampling rates are 8 to 10 times the maximum plant frequency. *See* ANALOG-TO-DIGITAL CONVERTER.

Digital-to-analog converter. Since a digital controller computes the control signal used to drive the plant, a digital-to-analog converter must be used to change this binary number to an analog voltage. The output of a digital-to-analog converter followed by a simple zero-order hold would be a step function (Fig. 1*c*), with the plant's dynamics being slow relative to the discontinuities in the output of the zero-order hold. *See* DIGITAL-TO-ANALOG CONVERTER.

Digital compensator. The digital compensator in the typical sampled-data control system (Fig. 2) takes

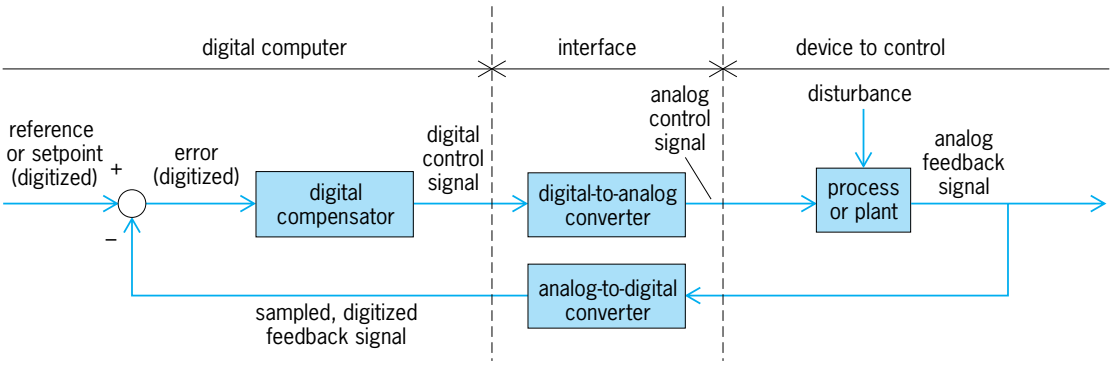


Fig. 2. Typical closed-loop, sampled data control system. The device to be controlled has continuously valued feedback signals which are sampled by an analog-to-digital converter, operated on by the digital compensator, and then converted back to analog signals for control of the plant.

the digitized values of the analog feedback signals and combines them with the setpoint or desired trajectory signals to compute a digital control signal, to actuate the plant through the digital-to-analog converter. A compensator is used to modify the feedback signals in such a way that the dynamic performance of the plant is improved relative to some performance index. In general, control systems are used to reduce output errors, maintain proper plant output in response to disturbances, compensate for plant wear or changing parameters, or make an uncontrollable plant controllable.

Sampled-data control systems are designed by first developing a mathematical model for the plant or process to control. From this model, inherent capabilities can be computed and performance deficiencies identified. A sampling rate can be selected and a compensator designed through well-established procedures in order to meet the desired performance measure. Compensators can be designed through a direct digital approach operating on sampled data signals, or they can be designed as continuous systems and then converted to sampled-data systems, with some trial and error required in both cases to meet the desired performance goals. *See* CONTROL SYSTEMS; DIGITAL COMPUTER; DIGITAL CONTROL.

Kenneth J. Hintz

Bibliography. C. H. Houpis and G. B. Lamont, *Digital Control Systems*, 2d ed., 1992; K. Ogata, *Discrete-Time Control Systems*, 2d ed., 1995; G. A. Perdikaris, *Computer Controlled Systems*, 1991; J. R. Ragazzini and G. F. Franklin, *Sampled-Data Control Systems*, 1958.

Sand

Unconsolidated granular material consisting of mineral, rock, or biological fragments between 63 micrometers and 2 mm in diameter. Finer material is referred to as silt and clay; coarser material is known as gravel. Sand is usually produced primarily by the chemical or mechanical breakdown of older source rocks, but may also be formed by the direct chemical precipitation of mineral grains or by biological processes. Accumulations of sand result from hydrodynamic sorting of sediment during transport and deposition. *See* CLAY MINERALS; GRAVEL; MINERAL; ROCK; SEDIMENTARY ROCKS.

Origin. Most sand originates from the chemical and mechanical breakdown, or weathering, of bedrock. Since chemical weathering is most efficient in soils, most sand grains originate within soils. Rocks may also be broken into sand-size fragments by mechanical processes, including diurnal temperature changes, freeze-thaw cycles, wedging by salt crystals or plant roots, and ice gouging beneath glaciers. *See* WEATHERING PROCESSES.

Composition. Because sand is largely a residual product left behind by incomplete chemical and mechanical weathering, it is usually enriched in minerals that are resistant to these processes. Quartz not only is extremely resistant to chemical and mechanical weathering but is also one of the most abundant minerals in the Earth's crust. Many sands dominantly consist of quartz. Other common constituents include feldspar, and fragments of igneous or metamorphic rock. Direct chemical precipitation or hydrodynamic processes can result in sand that consists almost entirely of calcite, glauconite, or dense dark-colored minerals such as magnetite and ilmenite. Depending on local source rocks, sand may consist of more unusual minerals, such as gypsum, halite, or phosphate. Sand that consists largely of calcite in the form of shell fragments or skeletal debris is common in environments where biological activity is great but input from other sediments is low. *See* FELDSPAR; QUARTZ.

The compositions of sand most closely resembles the composition of source rocks when chemical and mechanical weathering processes are subdued, as in arid (desert) or tundra (arctic) settings. Where high relief allows for little soil development and rapid transport of sediment, sand is exposed to the weathering environment only briefly, and sands also are little altered by chemical or mechanical weathering. Conversely, in low-relief settings in humid or tropical climates, sand composition may differ greatly from the compositions of source rocks.

Texture. Sand can be described texturally in terms of the distribution of grain sizes, the roundness and shapes of grains, and the surface textures of grains. The mean size of sand grains is largely determined by the energy of the transport medium; higher current velocities or greater wave energy can transport grains of larger size. The coarsest grains are deposited at the first point where current velocity decreases, generally near the source area, and so grain size generally decreases with increased transport distance. The range of grain sizes present is referred to as sorting; sand containing a large variation in grain sizes is poorly sorted, whereas sand of nearly a single grain size is well sorted. Sorting is strongly influenced by transport and deposition processes; river sand tends to be poorly sorted, beach sand is intermediate, and eolian sand is generally well sorted.

The shape of grains is usually inherited from the source rock. For example, sand grains derived from a schist tend to be flat and tabular, whereas sand derived from a granite may consist of grains that are nearly equidim e nsional. Grain shape is not to be confused with grain roundness, a measure of the angularity of the grain edges. Well-rounded grains are smooth, while angular grains have sharp edges. Mechanical weathering in high-energy environments (such as beaches) or in low-viscosity transport mediums (air) can round sand grains, as can chemical dissolution in soils. The surface textures of sand grains, as seen under extreme magnification, can be useful indicators of the origin and transportational history of sand. For example, sand from an intense chemical weathering environment may be covered with dissolution etch pits (**Fig. 1**), and sand from eolian environments tends to be frosted.

Transport. Sand can be transported by any medium with sufficient kinetic energy to keep the grains

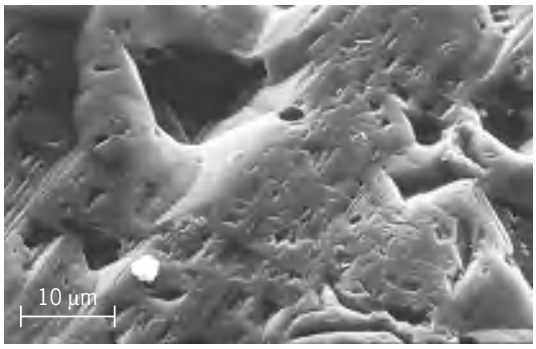


Fig. 1. Scanning electron photograph of the surface of a quartz grain from an intense chemical weathering environment. The surface texture shows numerous dissolution etch pits.

in movement. In nature, sand is commonly transported by rivers, ocean currents, wind, or ice. Rivers are responsible for transporting the greatest volume of sand the greatest distances. When rivers deposit sand at the oceans edge, it is commonly remobilized by waves or transported along shore by powerful longshore currents. Sand, like coarser and finer sediments, may also be entrained by glacial ice and transported to the sea by glacial flow. Where sand is widely exposed to wind currents, such as in river bars, on beaches, and in glacial outwash plains, it may be picked up and transported by the wind. As it moves, whether subaqueously or subaerially, sand commonly is organized into complex moving structures (bedforms) such as ripples and dunes (**Fig. 2**). *See* DUNE.

Occurrence. Accumulations of sand result from hydrodynamic sorting of sediment during transport. Most dramatic are the vast inland sand seas (ergs) in desert areas and accumulations at beaches. Probably at least as much sand is present, however, in river beds and temporarily stored in river floodplains as alluvium. Sand accumulations occur where a sudden decrease in current velocity causes the current to drop any carried sand. Thus, points where rivers or currents suddenly decelerate, such as deltas or the shallow continental shelf, are likely sites for the accumulation of sand. Sand is rare in the deep sea because currents have been robbed of their sand by nearshore deposition. Nevertheless, catastrophic subsea avalanches of mixtures of sediment and water, known as turbiditity currents, can carry appreciable amounts of sand to the deep sea. *See* DELTA; TURBIDITY CURRENT.

Fig. 2. Dunes produced by the transport of sand by wind, White Sands National Monument, NM. The sand in these dunes is made up almost entirely of gypsum ($CaSO_4$). Despite their appearance, the dunes are only a thin veneer of sand that is steadily moving across the countryside.

Geologic significance. Sand composition and texture can be used to tell much about the geologic history of the region in which it is found. Sand composition, for example, reflects the composition of source rocks, which in turn is controlled by the plate tectonic setting. With the advent of the plate tectonic model, it was noted (especially by W. R. Dickinson and colleagues) that the compositions of sands and sandstones could be related to the tectonic environments of source terranes, and sandstone composition has since been used as a tool for inferring the tectonic setting of ancient sequences. However, sand composition also is modified by chemical and mechanical processes. Thus, if source rocks remain unchanged through geologic time, changes in the composition of sands produced from them may be used as a tool for assessing paleoclimate. Finally, grain texture (roundness, surface markings, and so on) may be used to judge the nature (eolian, fluvial, glacial) and distance of transport, or the extent of reworking in the depositional environment, such as at beaches. *See* DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; PLATE TECTONICS.

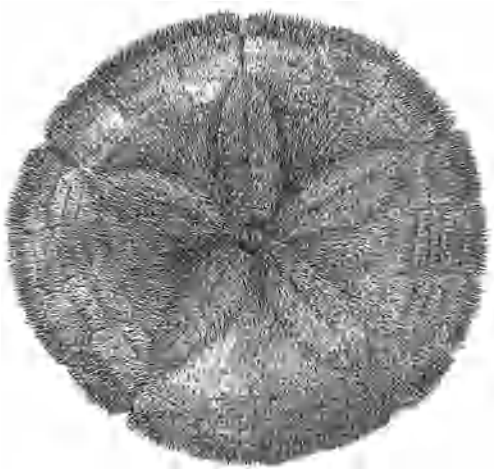
Economic importance. Sand and gravel production is second only to crushed stone among nonfuel minerals in the United States. Although sand and gravel has one of the lowest average per ton values of all mineral commodities, the vast demand makes it among the most economically important of all mineral resources. Sand and gravel is used primarily for construction purposes, mostly as concrete aggregate. Pure quartz sand is used in the production of glass, and some sand is enriched in rare commodities such as ilmenite (a source of titanium) and in gold. *See* CONCRETE.

Mark J. Johnsson

Bibliography. M. J. Johnsson and A. Basu, Processes Controlling the Composition of Clastic Sediments, *Geol. Soc. Amer. Spec. Pap.*, no. 284, 1993; A. C. Morton, S. P. Todd, and P. D. W. Haughton (eds.), Developments in Sedimentary Provenance Studies, *Geol. Soc. London. Spec. Publ.*, no. 57, 1991; F. J. Pettijohn, P. E. Potter, and R. Siever, *Sand and Sandstone*, Springer-Verlag, New York, 1987; R. Siever, *Sand*, W. H. Freeman, New York, 1988; G. G. Zuffa, *Provenance of Arenites*, NATO Advanced Studies Institute, Reidel, Dordrecht, 1985.

Sand dollar

An echinoderm belonging to the order Clypeasteroida in the class Echinoidea. Sand dollars have a flat, disk-shaped body (see **illus.**), with the mouth in a mid-ventral position and with the anus, also on the ventral surface, laterally displaced. There are several species, but the best known is *Echinarachnius parma*, common along the Atlantic Coast of



Sand dollar. The nearly circular body may reach 3 in. (7.5 cm) in diameter. The characteristic pattern of a five-part flower is seen on the aboral surface.

the United States, as well as the Pacific Coast south of Puget Sound. Other species are *Clypeaster subdepressus* and *Dendraster excentricus* of the Pacific Coast and *Mellita quinquiesperforata* found from Nantucket to Brazil.

Sand dollars live in sand, on the surface or partly buried, from the low-tide mark to depths of 4800 ft (1460 m). Burrowing and locomotion are assisted by the short spines which cover the body. Sand dollars ingest sand grains covered with diatoms or other algae. Little is known about their breeding habits. *See* ECHINOIDEA; SEA URCHIN.

Charles B. Curtin

Sandalwood

The name applied to any species of the genus *Santalum* of the sandalwood family, Santalaceae (see **illus.**). However, the true sandalwood is the hard, close-grained aromatic heartwood of a parasitic tree, *S. album*, of the Indo-Malayan region. This fragrant wood is used in ornamental carving, cabinet work, and as a source of certain perfumes. The odor of the wood is an insect repellent, and on this account the wood is much used in making boxes and chests. The fragrant wood of a number of species in other families bears the same name, but none of these is the real sandalwood. *See* SANTALALES.

Perry D. Strausbaugh; Earl L. Core

Sandalwood branch with foliage and fruit.

Sandstone

A clastic sedimentary rock comprising an aggregate of sand-sized (0.06–2.0-mm) fragments of minerals, rocks, or fossils held together by a mineral cement. Sandstone forms when sand is buried under successive layers of sediment. During burial the sand is compacted, and a binding agent such as quartz, calcite, or iron oxide is precipitated from ground water which moves through passageways between grains. Sandstones grade upward in grain size into conglomerates and breccias; they grade downward in size into siltstones and shales. When the proportion of fossil fragments or carbonate grains is greater than 50%, sandstones grade into clastic limestones. *See* BRECCIA; CONGLOMERATE; LIMESTONE; SAND; SHALE.

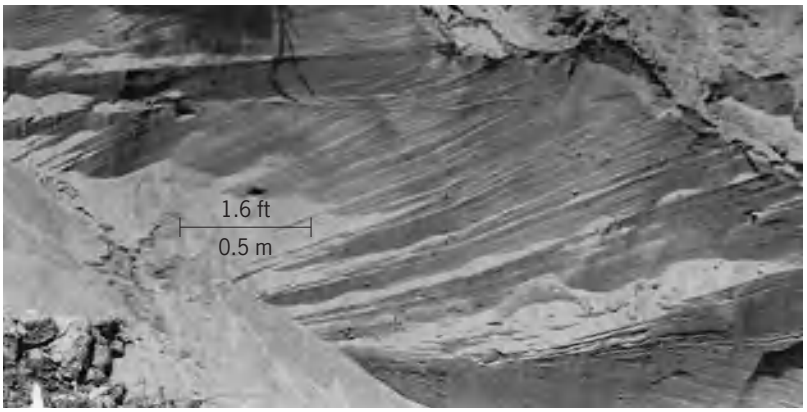
Components. The basic components of a sandstone are framework grains (sand particles), which supply the rock's strength; matrix or mud-sized particles, which fill some of the space between grains; and crystalline cement. The composition of the framework grains reveals much about the history of the derivation of the sand grains, including the parent rock type and weathering history of the parent rock. Presence or absence of mud matrix is a function of the viscosity of the transporting agent and the rate of deposition and burial of the sand. Crystalline cement is largely responsible for the color of sandstone. Very small amounts (less than 1%) of iron oxide cement, for example, can impart vivid reds to ancient sandstones. Cements can be studied to learn about the temperature and depth of burial of sand. Textural attributes of sandstone are the same as those for sand, and they have the same genetic significance. *See* SAND.

Lee J. Suttner

Classification. Sandstones are composed of mineral grains and rock fragments of many different kinds. Individual crystals of quartz and feldspar (the most common rock-forming minerals) and polycrystalline fragments of volcanic, sedimentary, and metamorphic rocks are typically most abundant. The quartz and feldspar grains are derived from the erosion of bedrock sources, such as granite, with constituent crystals large enough to yield sand grains composed of single minerals. The polycrystalline sand grains, called lithic fagments, are derived from rocks so finely crystalline that derivative sand grains are composed of multiple internal crystals. The nature of the sand grains is thus controlled partly by the nature of the bedrock from which the sand grains were derived, and partly by the degree of weathering of bedrock sources prior to erosion. Weathering tends to concentrate quartz, which is resistant to weathering, by dissolving feldspar and most other rock-forming minerals or converting them to clays at varying rates, depending upon the climate and topographic relief of source areas. In general, warm humid climates and low relief promote more

intense weathering of source rocks, while cool dry climates and high relief allow deeper erosion of fresh bedrock without the comparable weathering of susceptible minerals. Consequently, sandstones are classified according to the relative proportion of quartz to other grain types, and according to the ratio of feldspar grains to finely crystalline lithic fragments. Quartz-rich sandstones are commonly called quartz-arenite. Sandstones poor in quartz are commonly called arkose, when feldspar grains are more abundant than lithic fragments, and litharenite (or graywacke) when the reverse is true. Subarkose and sublitharenite (or subgraywacke) refer to analogous sandstones of intermediate quartz content. The interpretation of sandstone composition in terms of parent source rock is complicated by the fact that many sand grains are successively reworked by sedimentary transport during sequential cycles of weathering and erosion, meaning that the source of some sand is preexisting sandstone. Some sandstones also contain grains that are produced by either organic or inorganic processes at or near the site of deposition, and do not reflect any bedrock sources. Chief among these intrabasinal grains are fragments of fossils and other calcareous particles composed of calcium carbonate. Sandstones composed dominantly of calcareous grains are called calcarenite, and represent a special variety of limestone. Other sandstones composed exclusively of volcanic debris are called volcanic sandstone, and are gradational, through the interplay of eruptive and erosional processes, to tuff, the fragmental volcanic rocks produced by the disintegration of magma during explosive volcanic eruptions. *See* ARENACEOUS ROCKS; ARKOSE; FELDSPAR; GRAYWACKE; QUARTZ; TUFF. William R. Dickinson

Structures. Much can be learned about the history of a particular sandstone deposit through careful study of its sedimentary structures and fossils. The most common structure found in sandstones is crossbedding, which is produced by the wind- or water-driven migration of ripples or dunes (see **illus.**). Crossbeds are inclined at angles less than 35° to the primary parallel bedding visible in most outcrops. The crossbeds slope downward in the direction in which the current that formed them was flowing. Consequently they are important indicators of paleoflow directions. Detailed study of the size and shape of the crossbeds also can aid in interpreting the velocity of the ancient current flow which formed them, and in the case of subaqueously formed crossbeds, the depth of the flow.



Crossbeds in sandstone. The water which deposited the sand was flowing from right to left as indicated by the direction of inclination of the crossbeds.

Structures also can be observed on the tops and bottoms of sandstone beds. For example, many different shapes of ripples have been found on the tops of sandstone beds and have been used to distinguish between movement of the sand by water currents, tidal currents, waves, and wind. Linear groove and scour marks occur as casts on the bottoms of some sandstone beds. These features are formed by very swift currents which gouge the underlying sand or mud. The gouges are then filled with sand when the current slows down. Generally, fossil shells are not common in sandstone. However, both within some sandstone beds and on the tops and bottoms of some beds, a variety of fossil trails and tracks of organisms may be present.

Interpretation. Many consider sandstones to be the most useful sedimentary rocks to interpret Earth history. They are widespread, constituting between 10 and 20% of the total volume of sedimentary rocks in the Earth's crust, and they extend almost back to the beginning of geologic time. Furthermore, they are easier to study than fine-grained rocks such as shale and are less susceptible to change after deposition, which obscures their original properties, than limestone.

Unlike limestones, sandstones also can reveal important paleogeographic information about the areas surrounding their sites of deposition. As an example, an area where arkose or feldspathic sandstone accumulated more than likely was located near mountains, as abundant sand-sized feldspar can be derived only from crystalline rocks. Significant volumes of such rocks normally are exposed only within mountain ranges. In contrast, quartz arenites typically are formed where relief is low or where older sandstones served as parent rocks. Lithic sandstones have been deposited in a wide range of shallow- to deep-water environments in structurally active parts of the Earth's crust such as subduction zones and zones of continental collision. *See* PLATE TECTONICS.

Uses. Because sandstone can possess up to 35% connected pore space, it is the most important reservoir rock in the Earth's crust. By far the greatest volume of water, oil, and gas is extracted from sandstone, where it occurs in pore spaces. Although in the past fluid has usually been extracted from sandstone, in the future sandstone may serve as a reservoir into which hazardous fluids, such as nuclear wastes, are injected for storage. *See* HAZARDOUS WASTE.

Sandstone which is easily split (flagstone) and has an attractive color is used as a building stone. Sandstone is also an important source of sand for the glass industry and the construction industry, where it is used as a filler in cement and plaster. Crushed sandstone is used as road fill and railroad ballast. Silica-cemented sandstone is used as firebrick in industrial

furnaces. Some of the most extensive deposits of uranium are found in sandstones deposited in ancient stream channels. *See* GLASS; REFRACTORY; SEDIMENTARY ROCKS; STONE AND STONE PRODUCTS; URANIUM.

Lee J. Suttner

Bibliography. H. Blatt, *Sedimentary Petrology*, 2d ed., 1991; S. Boggs, Jr., *Petrology of Sedimentary Rocks*, 1991; S. Boggs, Jr., *Principles of Sedimentology and Stratigraphy*, 2d ed., 1996; R. L. Folk, *Petrology of Sedimentary Rocks*, 4th ed., 1980; J. T. Greensmith, *Petrology of the Sedimentary Rocks*, 7th ed., 1988; G. Klein, *Sandstone Depositional Models for Exploration*, 1988; F. J. Pettijohn, P. E. Potter, and R. Siever, *Sand and Sandstone*, 2d ed., 1987.

Santalales

An order of flowering plants, division Magnoliophyta (Angiospermae), in the subclass Rosidae of the class Magnoliopsida (dicotyledons). The order consists of 10 families and about 2000 species. The largest families are the Loranthaceae (about 900 species), Santalaceae (about 400 species), Viscaceae (about 350 species), and Olacaceae (about 250 species). A few of the more primitive members of the Santalales are autotrophic, but otherwise the order is characterized by progressive adaptation to parasitism, accompanied by progressive simplification of the ovules. In the Loranthaceae the ovules have no integument and are embedded in the large, central placenta. Some other families have a free central placenta with pendulous, terminal ovules; in any case there are only a few ovules, in contrast to the very numerous ovules of the Rafflesiales. Some members of the Santalales, such as sandalwood (*Santalum album*, a small tree of southern Asia), are rooted in the ground and produce small branch roots which invade and parasitize the roots of other plants. Others, such as mistletoe (*Viscum* and other genera of the Viscaceae), grow on trees, well above the ground. *See* FLOWER; MAGNOLIOPHYTA; MAGNOLIOPSIDA; MISTLETOE; PLANT KINGDOM; RAFFLESIALES; SANDALWOOD.

Arthur Cronquist; T. M. Barkley

Sapindales

An order of flowering plants, division Magnoliophyta (Angiospermae), in the subclass Rosidae of the class Magnoliopsida (dicotyledons). The order consists of 15 families and about 5400 species. The largest families are the Rutaceae (about 1500 species), Sapindaceae (about 1500 species), Meliaceae (about 550 species), Anacardiaceae (about 600 species), and Burseraceae (about 600 species). Most of the Sapindales are woody plants, with compound or lobed leaves and polypetalous, hypogynous to perigynous flowers with one or two sets of stamens and only one or two ovules in each locule of the ovary. A large proportion of the species have glandular cavities in the leaves, or resin ducts in the bark and wood, or other sorts of secretory structures. Some of the Anacardiaceae, including poison ivy (*Toxicodendron radicans*; see **illus.**) and the lacquer tree (*T. vernicifluum*) of the Orient, are notoriously allergenic to humans. Some other well-known members of the order are orange, lemon, lime, and grapefruit (all species of *Citrus* in the Rutaceae); the various kinds of maple (*Acer*, in the Aceraceae); and the horse chestnut (*Aesculus* hippocastanum, in the Hippocastanaceae). *See* BUCKEYE; CASHEW; CITRON; GRAPEFRUIT; KUMQUAT; LEMON; LIGNUMVITAE; LIME (BOTANY); MAGNOLIOPHYTA; MAGNOLIOPSIDA; MAHOGANY; MANDARIN; MAPLE; ORANGE; PISTACHIO; PLANT KINGDOM; POISON IVY; QUEBRACHO; SECRETORY STRUCTURES (PLANT); TANGERINE; VARNISH TREE.

Arthur Cronquist; T. M. Barkley

Poison ivy (*Toxicodendron radicans*), a characteristic member of the family Anacardiaceae in the order Sapindales, division Magnoliophyta. (*John Gerard*, *National Audubon Society*)

Sapphire

Any gem variety of the mineral corundum (Al_2O_3), except those called ruby because of their medium- to dark-red color. Sapphire has a hardness of 9 (Mohs scale), a specific gravity near 4.00, and refractive indices of 1.76–1.77.

Although blue sapphires are most familiar to the public, transparent sapphires occur in many other colors, including yellow, brown, green, pink, orange, and purple, as well as colorless and black. Some sapphires change color, from violet to purple or green to red, depending on the type of light under which they are viewed. A lovely, slightly light red-orange variety is called padparadscha. Star sapphires are fashioned from translucent corundum that contains many silklike crystals of the mineral rutile. The rutile crystals

form in three sets, and when the corundum is cut en cabochon (domed, without facets) and oriented correctly, light reflected from the rutile creates a six-rayed star. Corundum belongs to the trigonal crystal system and is isostructural with hematite, Fe_2O_3. *See* CORUNDUM; HEMATITE; RUBY.

Kashmir, in northern India, once the most famous source of blue sapphires, has become a minor source; another former source is Burma. Sri Lanka, Thailand, and Australia are now considered the most important sources. Many Sri Lankan and Australian sapphires are sent to Bangkok, where they are heat-treated to improve their color. One type, which is a cloudy white, becomes a magnificent, transparent rich blue when the heating dissolves the rutile needles, clarifying the stone. Australia produces enormous quantities of very dark blue sapphires, as well as yellow-to-golden sapphires. Sapphires are also mined in Montana in the United States and in East Africa. Most sapphires are mined either in gravels resulting from the weathering of basic igneous rocks, or in the type of marble resulting from the intrusion of an igneous mass into an impure limestone.

Blue sapphire is most valuable when it is a medium to medium-dark tone of a slightly violet-blue, or cornflower blue. The Kashmir grade has a slightly "sleepy" appearance, caused by inclusions that very slightly reduce transparency. Blue sapphires and the padparadscha variety are generally more valuable than other colors. Colorless sapphire is of little value. The golden sapphires are much in demand and thus quite expensive.

Sapphire in blue and other colors is synthesized by the Verneuil flame-fusion process, and some blue sapphires and the padparadscha variety have been made by other methods. The only flux-melt blue sapphires are made in California. *See* GEM.

Richard T. Liddicoat, Jr.

Sapropel

A term possessing genetic implications, originally defined as an aquatic sediment rich in organic matter that formed under reducing conditions (lack of dissolved oxygen in the water column) in a stagnant water body. This contrasts with the term gyttja, which is also a sediment high in organic carbon content but which formed under inferred oxygenated conditions in the water column down to the sediment-water interface (thus benthic organisms may be present). Such inferences about water-column dissolved-oxygen contents are not always easy to make for ancient environments. Therefore, the term sapropel or sapropelic mud has been used loosely to describe any discrete black or dark-colored sedimentary layers (>1 cm or 0.4 in. thick) that contain greater than 2 wt % organic carbon. Sapropels may be finely laminated (varved) or homogeneous, and may less commonly exhibit structures indicating reworking or deposition of the sediment by currents. Sapropels largely contain amorphous (sapropelic) organic matter derived from planktonic organisms (such as planktonic or benthic algae in lakes or plankton in marine settings). Such organic matter possesses a large hydrogen-to-carbon ratio; therefore, sapropelic sequences are potential petroleum-forming deposits. The enhanced preservation of amorphous organic matter in sapropels may indicate conditions of exceptionally great surface-water productivity, extremely low bottom-water dissolved-oxygen contents, or both. Some sapropels may, however, contain substantial amounts of organic matter derived from land plants. *See* ANOXIC ZONES; MARINE SEDIMENTS; ORGANIC GEOCHEMISTRY; PETROLEUM; VARVE.

Michael A. Arthur

Sarcodina

A superclass of Protozoa (in the subphylum Sarcomastigophora) in which movement involves protoplasmic flow, often with recognizable pseudopodia. Gametes may be flagellated, as in certain Foraminiferida. Most species are floating or creeping; a few, sessile. The pellicle is thin, and the body is apt to be plastic unless restrained by skeletal structures. Sarcodina live in fresh, brackish, or salt waters; soil or sand; and as endoparasites in animals and plants. A group may be limited to a specific habitat, but many have a rather wide range. *See* FORAMINIFERIDA.

Sarcodina include two major classes: Rhizopodea and Actinopodea. In addition, for supposed lack of a better taxonomic position, Piroplasmea (*Babesia*), as a third class, have been assigned to the group by some protozoologists.

Rhizopodea. Pseudopodia show two different types of activity: contractile-hydraulic movement (*Amoeba proteus*) and bidirectional movement (Foraminiferida and Heliozoia). Such pseudopodial dimorphism has led to the suggestion that a basic revision of the Sarcodina is needed. *See* RHIZOPODEA.

The Rhizopodea include five subclasses (Lobosia, Filosia, Granuloreticulosia, Mycetozoia, and Labyrinthulia), with such varied characteristics that interrelationships are somewhat uncertain. However, Rhizopodea do not show distinctly radial arrangement of slender pseudopodia. Skeletal elements range from none to complicated tests (**Fig. 1**) in many Granuloreticulosia. Simpler testate rhizopods build a one-chambered test (**Fig. 2**). In some species the test is composed entirely of secreted material (often siliceous), but many have a test composed of sand grains or other foreign particles (Fig. 2*e*) held together by a secreted cement. A few have a moderately flexible chitinous test. Rarely (for example, *Paraquadrula*), shell plates may be composed of calcium salts. Skeletal plates of *Euglypha* (Fig. 2*d*) are secreted during growth and stored until the next fission, when one organism receives them for building its test. Other rhizopods may ingest and store sand grains for use at the next fission.

Multichambered tests are formed by many Foraminiferida (Fig. 1*a–l*), and simpler tests by primitive species (Fig. 1*o*). In some genera the adult retains a chitinous test. More often, a primary test is

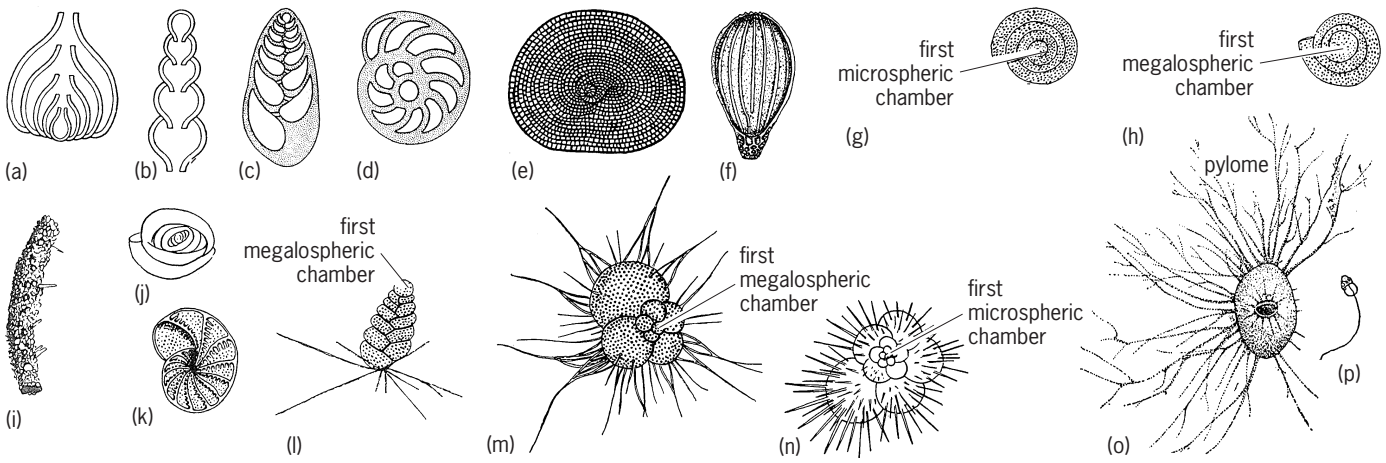


Fig. 1. Foraminiferida. (*a*, *b*) Foraminiferan tests, nodosaroid types. (*c*) Textularid test. (*d*) Spiral test. (*e*) Cycloid test (*Discospirulina*). (*f*) Test of *Lagena*. (*g*, *h*) Microspheric and megalospheric tests of *Cornuspira*. (*i*) Arenaceous test (*Saccorhiza*). (*j*) Test representative of family Miliolidae. (*k*) Spiral test (*Elphidium*). (*l*) *Textularia* (living). (*m*, *n*) *Globigerina* (living) megalospheric and microspheric. (*o*) *Gromia*. (*p*) Supposed flagellated gamete of *Gromia*. (*After L. H. Hyman, The Invertebrates, vol. 1, McGraw-Hill, 1940*)

reinforced by the addition of inorganic salts or foreign particles. In the formation of arenaceous tests (Fig. 1*i*), sand grains, sponge spicules, or other materials are cemented onto a primary chitinous wall. However, tests of most species are mainly calcareous (rarely siliceous) secretions, the materials resembling the cements used in arenaceous tests. During growth of certain simple species the original test is discarded and then replaced by a larger one. Forms with multilocular tests typically add new chambers in succession as growth continues (Fig. 1*a–b* and *j–n*). The protoplasm remains continuous from one chamber to the next, since the intervening apertures remain at least partly open. These openings (foramina) suggested the taxonomic name for the group. The reticulopodia tend to anastomose into a sticky net (Fig. 1*o*), often greatly exceeding the test in area. The net is an efficient food trap; also, partial digestion may occur outside the test. Reticulopodia also may construct the walls of new chambers or cysts. In addition, some Foraminiferida can creep at rates up to several millimeters an hour.

Actinopodea. Members of this class have slender pseudopodia, with or without axonemes, commonly extended radially with little anastomosing or branching (**Fig. 3**). Bidirectional cytoplasmic flow is known in reticulopodia of representative genera. There are four subclasses: Radiolaria, Acantharia, Heliozoia, and Proteomyxidia. *See* ACTINOPODEA.

Life cycles. Sexual activities include syngamy (the fusion of gametes) in certain Heliozoia, Mycetozoia, and Foraminiferida. Other reported cases are less well substantiated. In certain Foraminiferida there is an alternation of a haploid gamete-producing generation with a diploid generation, a plantlike cycle apparently unique in Protozoa. Species with such cycles may show dimorphism with respect to initial chambers of the tests (Fig. 1*g*, *h*, *m*, and *n*). A diploid adult has a microspheric (small) initial chamber; a haploid adult is megalospheric. It is suggested that in *Gromia* a naked multinucleate asexual

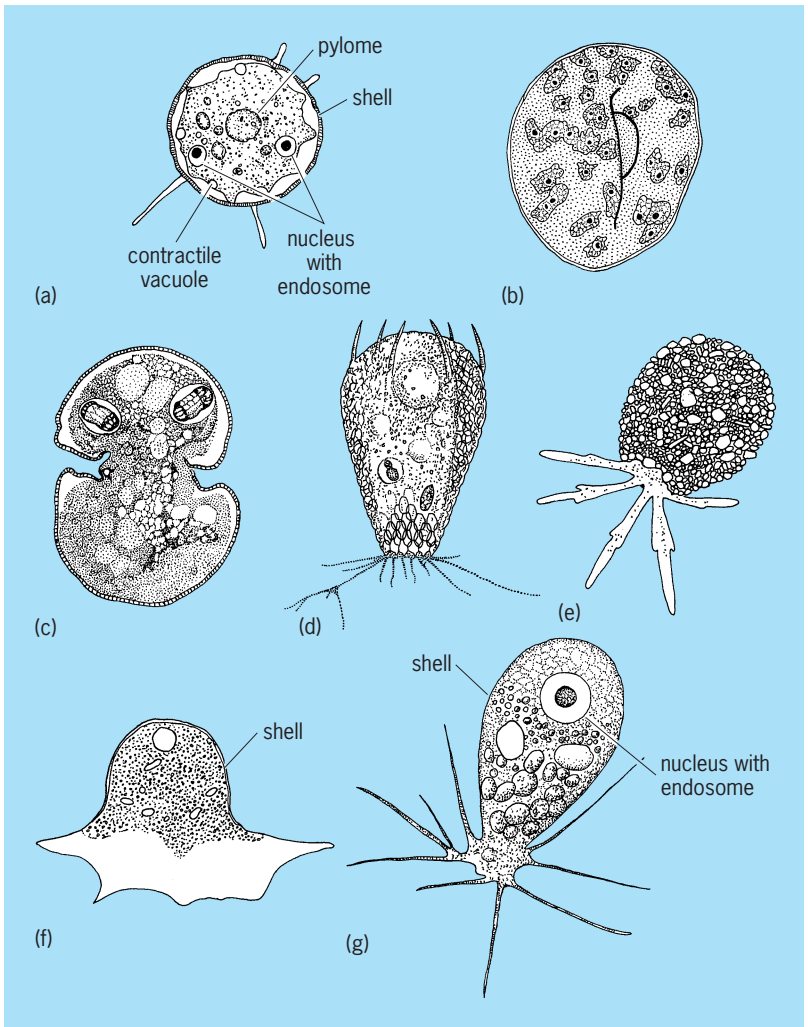


Fig. 2. Some shelled Lobosia. (*a*) *Arcella*, seen from above; (*b*) sporulating into numerous ameboid young (*c*) in binary fission, with the two nuclei in mitosis with polar caps part of the cytoplasm exudes from the pylome and secretes a new shell. (*d*) *Euglypha* with a shell of little siliceous scales. (*e*) *Difflugia*, with a case made of rock grains. (*f*) *Cochliopodium*, with a membranous case. (*g*) *Chlamydophrys*. (*After L. H. Hyman, The Invertebrates, vol. 1, McGraw-Hill, 1940*)

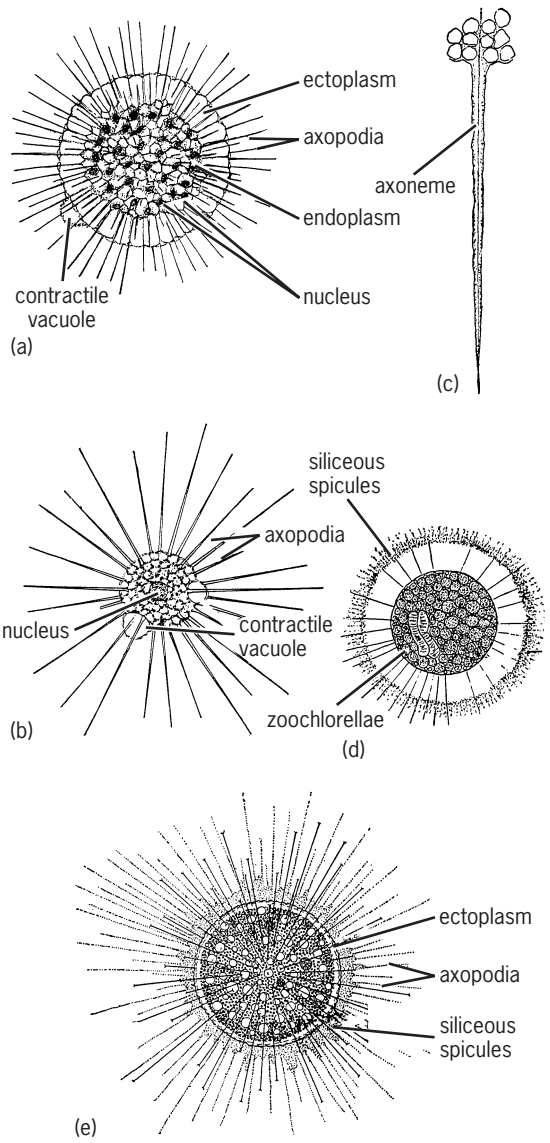


Fig. 3. Heliozoia. (*a*) *Actinosphaerium*. (*b*) *Actinophrys*. (*c*) Axopodium of *Actinosphaerium*, highly magnified. (*d*) *Heterophrys*. (*e*) *Acanthocystis*. (*After L. H. Hyman, The Invertebrates, vol. 1, McGraw-Hill, 1940*)

generation may alternate with a testate sexual generation.

In typical alternation of generation cycles, ordinary reproduction may occur in both haploid and diploid phases in sequences ranging from 2–3 weeks to 1 year or so in duration. *See* PROTOZOA; SARCOMASTIGOPHORA.

Richard P. Hall

Sarcomastigophora

A subphylum of Protozoa. Sarcomastigophora replaces Rhizoflagellata for those protozoa that possess flagella or pseudopodia or both, according to a classification set up by an international committee of the Society of Protozoologists. Included organisms have a single type of nucleus, except the developmental stages of some Foraminiferida. Sexuality, if present, is syngamy, the fusion of two gametes. Spores typically are not formed. Flagella may be permanent or transient or confined to a certain stage in the life history; this is true also of pseudopodia. Both flagella and pseudopodia may be present at the same time.

Classification. Three superclasses are included: (1) Mastigophora, commonly flagellates, contains 19 orders. (2) Opalinata includes 1 order; these organisms were once considered as ciliates, but further research has indicated flagellate kinships. (3) Sarcodina comprises organisms which normally possess pseudopodia and are flagellated only in the developmental stages; 13 orders possess irregularly distributed lobose or filose and branching pseudopodia, while 7 orders have radially distributed axopodia, often with axial filaments. *See* MASTIGOPHORA; SARCODINA.

Mastigophora. This superclass is often divided into two classes: Phytamastigophorea and Zoomastigophorea. The Phytamastigophorea, or plant flagellates, usually have green or variously colored chromatophores. However, the order Ebriida is altogether colorless, and eight other orders contain few to many species without chromatophores, especially Chrysomonadida, Dinoflagellida, and Euglenida. Besides chlorophylls, many other pigments have been identified, and reserve materials vary from order to order. Thus Volvocida elaborate starch; Euglenida, paramylum; Chloromonadida, lipids and glycogen; Dinoflagellida, starch, lipids, and unknown materials; and Chrysomonadida, leucosin and lipids. *See* PHYTAMASTIGOPHOREA.

These plant flagellates are very diverse in form and structure. Some are naked, and some live in loricas or tests which they manufacture. Some are fresh-water and others marine forms; most Cryptomonadida and colorless Euglenida tolerate either milieu. Most of them are unicellular, but some form colonies, and in the Volvocida the *Volvox* colonies are quite complex. However, this whole class shows decided relationships to one another.

Members of the class Zoomastigophorea, all nonpigmented, are referred to as the animal flagellates. Three of the orders, Choanoflagellida, Bicosoecida, and Rhizomastigida, are free-living; the remaining six orders are predominantly parasitic. The choanoflagellates have been used to relate the protozoa to sponges by virtue of the common protoplasmic collar. This group also manufactures very complex loricae, as shown by R. E. Norris for *Pleurasiga* or *Sportelloeca*. Relationships between these three orders are quite evident, and through genera such as *Mastigella* also, to the ameboid protozoa. The increase in number of flagella from one in Choanoflagellida to many in Hypermastigida conceivably leads to opalinid and ciliate relationships.

The remaining six orders are partially or wholly parasitic, the parasites (or in some cases symbionts) of wood roaches and termites attaining a very high degree of structural organization and adaptation to their mode of life. *See* ZOOMASTIGOPHOREA.

Opalinata. These are practically all inhabitants of the digestive tract of amphibians, and M. M. Metcalf, who made very extensive studies of them in 1923, referred to them as "opalinid ciliates," which they

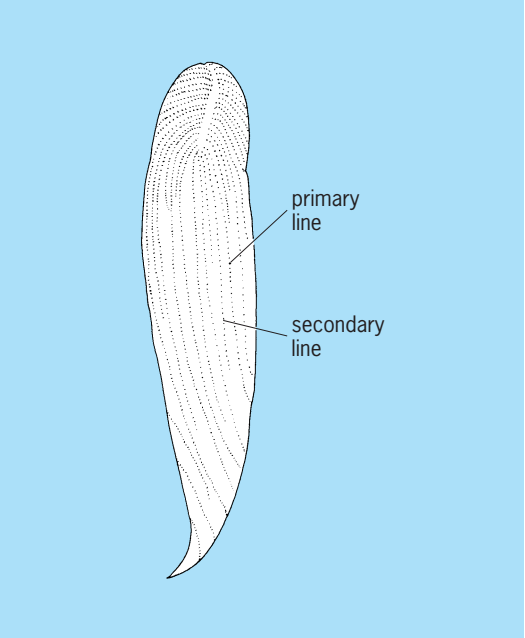


Fig. 1. An opalinid (*Protoopalina axonucleata*) showing primary and secondary lines of blepharoplasts.

certainly resemble (**Fig. 1**). Since about 1950, however, considerable questioning of their phylogenetic position has arisen, especially since anisogamy (fusion of dissimilar gametes) has been demonstrated. They are regarded as a highly specialized group, and the protozoologists' committee regards their taxonomic position as needing further clarification. In this respect they are important, but as to the size of the group and their ecology, they are less so.

Sarcodina. No common name can apply to this group, which includes the amebas, foraminiferids, slime molds, radiolarians, acantharians, and heliozoans. Each of these is well known to perhaps a few specialists. The superclass Sarcodina comprises 20 orders which appear to differ widely but commonly have pseudopodia, restriction of flagella to developmental stages, a cortical zone of relatively undifferentiated ectoplasm, and asexual reproduction by fission. Most of them are free-living. *See* SARCODINA.

The class Rhizopodea includes the large *Amoeba*, *Chaos*, and *Pelomyxa* (of beginning courses in biology), which are common in fresh waters. It also includes the large *Gromia*, usually marine, and the very abundant and often large Foraminiferida. The subclass Mycetozoia embraces slime molds, which may be parasitic (order Plasmidiophora) or free-living and which may live in an aerial instead of an aqueous environment. The Mycetozoia forms peculiar and diagnostic fruiting bodies on decaying logs and leaves; and these organisms have a complicated life cycle. *See* RHIZOPODEA.

Rhizopodea even includes blood parasites (order Piroplasmida), in which flagella are unknown. This order is placed in a separate class, Piroplasmea, according to the new classification.

The class Actinopodea contains four subclasses which are usually floating and have radiating filose pseudopodia, often with an axial filament. Included are the marine Radiolaria, which have a central capsule containing endoplasm, perforated by one to many pores. The subclass Acantharia differs in having a skeleton of regularly arranged radial spines of strontium sulfate. The central capsule is thin, membranous, and without pores. *See* ACTINOPODEA.

The three orders of Heliozoia are common freshwater genera, some large (*Actinophrys*), some small (**Fig. 2**), and all free-living. The subclass Proteomyxidia is principally parasitic and is poorly known.

Ecology and economic significance. Most of the plant flagellates will live either in fresh water or in both fresh and salt water. The orders Silicoflagellida, Coccolithophorida, and Ebriida are marine forms. The green Euglenida have only four or five species of *Euglena* in salt water, while species of the genera *Eutreptia* and *Eutreptiella* are confined to salt water. The plant flagellates may be very abundant and are of general ecologic interest in being reoxygenators, as a lower step in the food chain, and in the decomposition of dissolved organic matter, since some of them are saprophytic in nutrition. Economically, some of them, such as the dinoflagellate *Gymnodinium breve*, cause enormous fish kills. *See* FOOD WEB.

The zooflagellates are small and are not sufficiently abundant to enter markedly into the food chain. But the parasites and symbionts are of considerable interest economically and theoretically, for example, the trypanosomes and the peculiar xylophagous

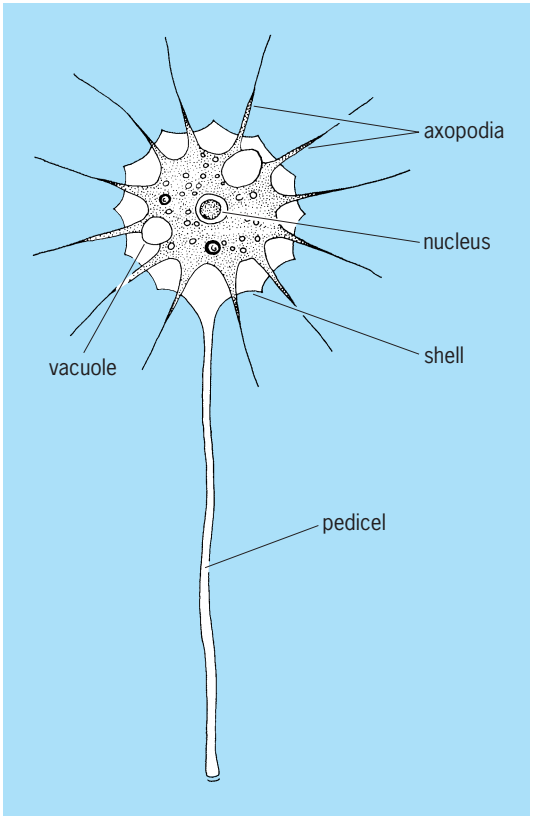


Fig. 2. *Hedriocystis reticulata*, a small heliozoan of the superclass Sarcodina.

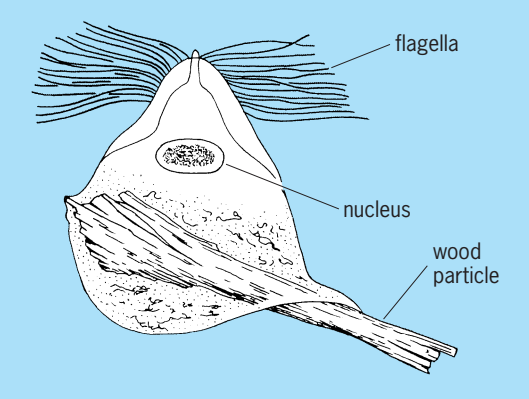


Fig. 3. A xylophagous symbiotic flagellate (*Trichonympha* sp.) from a termite.

(wood-eating) symbionts of termites (**Fig. 3**). In the termites these parasites actually digest the wood eaten by the host.

The rhizopods are typically free-living and rarely overabundant. Oil geologists make use of the shells of foraminiferids to determine what strata may have been penetrated. *See* FORAMINIFERIDA.

In the Mycetozoia are important parasites of cruciferous ciferous plants and the organism which is claimed, but not proven, to be the cause of the eelgrass (*Zostera*) decline along the Atlantic coast in the 1930s.

Conspicuous in the ecology of marine waters are the dinoflagellates, radiolarians, and acantharians, especially in tropical waters of otherwise low productivity. *See* PROTOZOA.

James B. Lackey

Bibliography. B. M. Honigberg et al., A revised classification of the phylum Protozoa, *J. Protozool.*, 11(1):7–20, 1964; M. M. Metcalf, The opalinid ciliate infusorians, *Bull. U.S. Nat. Mus.*, no. 120, pp. 1–484, 1923; M. M. Metcalf, Further studies on the opalinid ciliate infusorians and their hosts, *Proc. U.S. Nat. Mus.*, 87:465–635, 1940; R. E. Norris, Neustonic marine craspedomonadates (choanoflagellates) from Washington and California, *Protozoology*, 12(4):589–602, 1965.

Sarcopterygii

By consensus of contemporary scholars, Sarcopterygii is a class of vertebrates that contains the orders Coelacanthiformes (lobefin fishes) and Ceratodoniformes (lungfishes), as well as the Tetrapoda.

Following is a recent classification of the extant sarcopterygians:

Class Sarcopterygii
- Subclass Coelacanthimorpha
  - Order Coelacanthiformes (lobefin fishes)
    - Family Coelacanthidae
- Subclass Dipnotetrapodomorpha
  - Order Ceratodontiformes (lungfishes)
    - Family Ceratodontidae
    - Lepidoserinidae
    - Protopteridae
- Infraclass Tetrapoda

The names Amphibioidei and Choanichthyes also appear in the literature, both being equivalent to Sarcopterygii less the tetrapods. The latter name is certainly inappropriate, as not all sarcopterygians have choanae (internal nares which allow passage from the nasal cavity to the mouth). The Sarcopterygii of only what are conventionally called crossopterygians and dipnoans is also still a popular classification. Debates concerning the relationships of the sarcopterygians and the ancestry of tetrapods—that is, whether the origin of the tetrapods occurred within the lungfishes, coelacanths, or rhipidistians—have persisted for many years, and will probably continue into the future. *See* COELACANTHIFORMES; CROSSOPTERYGII; DIPNOI (CERATODONIFORMES); OSTEICHTHYES; RHIPIDISTIANS (OSTEOLEPIMORPHA); TETRAPODA.

Herbert Boschung

Bibliography. W. E. Bemis, W. W. Burggren, and N. E. Kemp (eds.), *The Biology and Evolution of Lungfishes*. J. Morphol. Suppl. 1, New York, Alan R. Liss, 1987; M.-M. Chang, Rhipidistians, dipnoans, and tetrapods, in H.-P. Schultze and L. Trueb (eds.), *Origins of the Higher Groups of Tetrapods*, pp. 3–28, Cornell University Press, Ithaca, New York, 1991; R. Cloutier, Patterns, trends, and rates of evolution within the Actinistia, *Environ. Biol. Fishes*, 32(1):23–58, 1991; J. G. Maisey, Heads and tails: A chordate phylogeny, *Cladistics*, 2(3):201–256, 1986; J. S. Nelson, *Fishes of the World*, 4th ed., Wiley, New York, 1994; D. E. Rosen, Lungfishes, tetrapods, paleontology and plesiomorphy, *Bull. Amer. Mus. Nat. Hist.*, 167:159–276, 1981; H.-P. Schultze, Dipnoans and sarcopterygians, in W. E. Bemis, W. W. Burggren, and N. E. Kemp (eds.), *The Biology and Evolution of Lungfishes*, pp. 39–47, J. Morphol. Suppl. 1, New York, Alan R. Liss, New York, 1987; K. S. Thompson, The origin of the tetrapods, in P. Dodson and P. Gingerich (eds.) *Functional Morphology and Evolution*, pp. 33–62, Amer. J. Sci. Spec. Vol. 293-A, 1993; G. C. Young, J. A. Long, and A. Richie, Crossopterygian fishes from the Devonian of Antarctica: Systematics, relationships and biogeographical significance, *Rec. Aust. Mus., Suppl. 14*, 1992.

Sarcosporida

An order of Protozoa of the class Haplosporea which comprises parasites in skeletal and cardiac muscle of vertebrates. The organisms have a very wide distribution both geographically and in host species, infecting reptiles, birds, and mammals (including marsupials). Humans are occasionally infected by the parasite referred to as *Sarcocystis lindemanni.* It is doubtful whether Sarcosporida found in different hosts are themselves always different, though this has often been assumed. Host specificity is known not to be strict; however, it is unlikely that there is only one species—*S. miescheriana*—as proposed by some investigators.

Sarcocystis. This genus is now usually placed with the Protozoa (subphylum Sporozoa), but it has given

taxonomists considerable difficulty in the past. By some it was thought to be a fungus, and by others it was simply relegated to a group of "undetermined position." The Committee on Taxonomic Problems of the Society of Protozoologists recommended that the parasites collectively referred to as the genus *Sarcocystis* be put in an order to be called Sarcosporida, class Haplosporea (this suggestion is followed here); *Sarcocystis* would then be the only genus in the order.

Species descriptions. The criteria of species are very hard to define for *Sarcocystis*. Species descriptions have usually been based on an assumed host specificity, size of spores, and cyst characteristics. But host specificity has seldom been proved, spore size is variable (degree of variation often in part dependent on the host species), and cyst morphology is frequently inconstant. Nevertheless, cyst morphology is probably more stable than anything else about the parasite.

Morphology. In general, the cyst wall is said to have one or two layers, the outer one being either smooth or provided with spines or villi. The genesis of the wall is in some dispute; some think it is formed by the host as a reaction to the parasite, but the majority opinion is that the parasite itself forms the cyst.

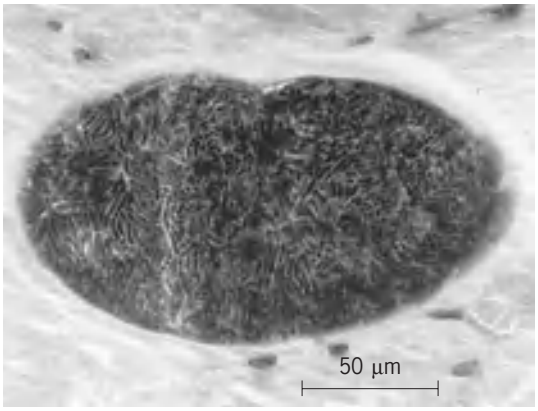
The morphology of both cyst wall and spores has been studied under the electron microscope. Examination with the ordinary optical microscope shows the cyst to be divided internally into compartments and the outer wall to be smooth or rough, depending on the species. Electron microscopy shows numerous, minute villuslike projections on the outer surface (at least for *S. fusiformis*, a parasite of cattle). The spores are crescent-shaped and have a somewhat superficial resemblance to *Toxoplasma* under the light microscope. When viewed with the electron microscope, however, this resemblance is much more marked and extends even to malaria sporozoites. All three have a minute projection, termed a conoid, at the more pointed or anterior end, from which fibrils pass anteriorly. These fibrils are called toxonemes at the point of origin, and sarconemes (in *Sarcocystis*) farther back. The nucleus and several mitochondria are in the posterior half.

Immune reactions. Additional evidence for thinking that there may be a close relationship between *Toxoplasma* and *Sarcocystis* is the frequent occurrence of cross-immune reactions when the Sabin-Feldman dye test is used. This test is ordinarily used for the diagnosis of toxoplasmosis, present or past, but is often positive in cases of sarcosporidiosis.

Life cycle. The life cycle of *Sarcocystis* is quite typical of the Sporozoa, since sexual stages are lacking and schizogony (multiple fission), though sometimes claimed, apparently does not occur. Instead, reproduction is by binary fission, with the eventual development of cysts (Miescher's tubules) in the muscles. These cysts are relatively large structures, easily visible with the unaided eye, and contain myriads of the minute crescentic spores (see **illus.**).

Infection of a new host is believed to be by the ingestion of these spores with food and water contaminated by feces from an infected animal. Most natural infections appear to be mild and to do the host little harm. Of course, some muscle tissue is destroyed but probably too little to be significant. However, sheep with heavy infections of *S. tenella*, one of the best-known species, suffer from loss of weight and give an impression of general unthriftiness. Though a quite potent endotoxin (sarcocystin) has been isolated from the parasites, there is little evidence that this usually does the host much harm. *See* HAPLOSPOREA; PROTOZOA; SPOROZOA.

Reginald D. Manwell

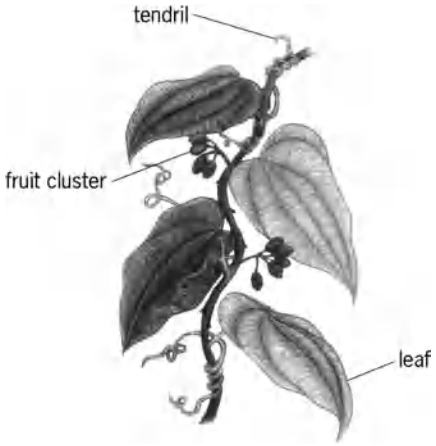


Cross section of Miescher's tube filled with *Sarcocystis* spores (probably *S. lindemanni*) from monkey.

Sarsaparilla

A flavoring material obtained from the roots of at least four species of the genus *Smilax* (Liliaceae). These are *S. medica* of Mexico, *S. officinalis* of Honduras, *S. papyracea* of Brazil, and *S. ornata* of Jamaica, all tropical American vines found in the dense, moist jungles (see **illus.**). The flavoring is used mostly in combination with other aromatics such as wintergreen. *See* LILIALES; SPICE AND FLAVORING.

Perry D. Strausbaugh; Earl L. Core



***Smilax aristolochiaefolia*, which yields Mexican sarsaparilla.**

Sassafras

A medium-sized tree, *Sassafras albidum*, of the eastern United States, extending north as far as southern Maine. Sometimes it is only a shrub in the north, but from Pennsylvania southward heights of 90 ft (27 m) or more with diameters of 4–7 ft (1.2–2.1 m) have been reported for this plant. Sassafras is said to live from 700 to 1000 years. It can be recognized by the bright-green color and aromatic odor of the twigs and leaves. The leaves are simple or mitten-shaped (hence a common name "mitten-tree"), or they may have lobes on both sides of the leaf blade (see **illus.**). The wood is soft, brittle, coarse-grained and

***Sassafras albidum*: twig, terminal bud, and leaf.**

somewhat aromatic, and the heartwood is reddish. Because it is durable in contact with moisture, it has been used for fence posts, rail fences, sills for houses, and boats. The aromatic substance, found especially in the roots, is a volatile oil (oil of sassafras) used medicinally as a stimulant and a diaphoretic, and also as a flavoring agent. *See* MAGNOLIALES; STEM; TREE.

Arthur H. Graves; Kenneth P. Davis

Satellite (astronomy)

A relatively small body orbiting a larger one that in turn orbits a star. In the solar system, all of the planets except Mercury and Venus have satellites. Over 150 were known by November 1, 2005, with 154 in well-established orbits and distributed as follows: Earth, 1; Mars, 2; Jupiter, 63; Saturn, 47; Uranus, 27; Neptune, 13; Pluto, 1. More satellites are being discovered each year as a result of deliberate searches by increasingly powerful Earth-based techniques and by the *Cassini-Huygens* mission to the Saturn system. The close flyby of the *Galileo* spacecraft (enroute to Jupiter) of the asteroid 243 Ida in 1993 revealed the presence of a 0.9-mi-diameter (1.5-km) satellite, now known as Dactyl. This unexpected discovery has been followed by the detection of several other asteroid satellites by Earth-based observers. Several Kuiper Belt objects (distant comet nuclei) have also been observed to be binaries. *See* ASTEROID; BINARY ASTEROID; KUIPER BELT.

For planets, it is customary to distinguish between regular satellites that have nearly circular orbits lying essentially in the plane of a planet's equator and irregular satellites whose orbits are highly inclined, very elliptical, or both. The former almost certainly originated with the parent planet, while the latter must be captured objects. The Earth's Moon is a special case. The most widely favored hypothesis for its origin invokes an impact with Earth by a Mars-sized planetesimal, and ejection of material that first formed a ring around the Earth and then coalesced to form the Moon. Pluto's Charon may have formed through a similar collision. *See* MOON; PLUTO; SOLAR SYSTEM.

The two satellites of Mars are tiny objects, less than 9 mi (15 km) in their longest dimension. They are irregularly shaped because the gravitational fields produced by their small masses are too weak to deform the material into spheres. The same is true of the small satellites of the outer planets. *See* MARS.

Jupiter has six groups of small outer satellites, all of which are irregular. Three groups have direct orbits; the other three have retrograde (clockwise) orbits. All six groups have high eccentricities. These six families of satellites presumably represent the capture of six separate objects in an early stage of Jupiter's history when the planet was surrounded by an extended subnebula. The capture process resulted in fragmentation, leading to objects from less than 2 km (1.2 mi) to 185 km (115 mi) in diameter. So far, 32 of these fragments have well-established orbits; there may be over 100 more minor fragments awaiting discovery. The four largest satellites were discovered by Galileo and Simon Marius in 1610. One of these, Io, exhibits an astonishing amount of volcanic activity, driven by tidal forces. Another, Europa, may have an ocean of liquid water trapped beneath an icy crust. The remaining two, Ganymede and Callisto, are both larger than the planet Mercury. Unlike the planet, however, they consist of 50% ice by mass, surrounding a rocky core. *See* JUPITER.

Saturn's inner, icy satellites have the brightest surfaces in the solar system, due in part to a coating of tiny ice crystals from the E ring, which is centered on (and presumably originates from) the satellite Enceladus. This tiny (diameter 500 km or 310 mi) satellite has an unusually warm region near its south pole, where the surface is split by a series of roughly parallel fissures. A plume of gas dominated by water vapor fills the space nearby. The leading hemisphere of Iapetus has a reflectivity of only 4%. It appears to be coated with dark, nitrogen-rich organic material whose origin remains unknown. Titan, the largest of the family, is only slightly smaller than Jupiter's Ganymede and has a nitrogen-dominated atmosphere that is denser than Earth's. Methane and nitrogen in that atmosphere are involved in photochemical reactions that produce organic compounds, some of which contribute to a ubiquitous aerosol layer that hides the surface of the

satellite at visible wavelengths. The *Huygens* probe successfully landed on Titan's surface on January 14, 2005, finding evidence of rivers of liquid methane and deposits of organic aerosols. Saturn has four distinct groups of irregular satellites, implying four discrete capture events, a situation similar to Jupiter's satellite families. *See* ALBEDO; SATURN.

Voyager 2 discovered 10 small regular satellites of Uranus close to the planet during its 1985–1986 flyby. Twelve distant, irregular satellites have subsequently been discovered from Earth. The five large previously known satellites of Uranus are unusually regular, except for Miranda, whose surface reveals evidence of a surprisingly vigorous geological history. *See* URANUS.

Neptune has two highly irregular satellites, Triton and Nereid, plus six small regular moons discovered by *Voyager 2*, in 1989, close to the planet, and five irregular satellites found from Earth. Additional observations of the latter group are required to establish their orbits. Triton has a tenuous atmosphere of nitrogen (a few millionths of the Earth's surface pressure) with traces of methane and carbon monoxide. All three of these gases, plus carbon dioxide and water, are frozen on its surface. That frigid (38 K or −391°F) surface reveals streaks of windblown material and active, eruptive plumes. *See* NEPTUNE.

Ice-covered Charon orbits Pluto, the closest member of the Kuiper Belt, with a period of revolution that is the same as the planet's period of rotation. Thus, Charon is like one of the Earth's artificial geosynchronous satellites: it would appear to hang motionless in the dark skies of Pluto. *See* PLANET.

Tobias C. Owen

Bibliography. J. K. Beatty, C. C. Petersen, and A. Chaikin (eds.), *The New Solar System*, 4th ed., Cambridge University Press, 1999; J. Burns and M. Matthews, *Planetary Satellites*, University of Arizona Press, 1987; D. Morrison and T. Owen, *The Planetary System*, 3d ed., Addison-Wesley, 2003; P. Nicholson, Natural satellites of the planets, in R. Gupta (ed.), *Observer's Handbook*, Royal Astronomical Society of Canada, published annually; D. A. Rothery, *Satellites of the Outer Planets: Worlds in Their Own Right*, 2d ed., Oxford University Press, 1999; F. W. Taylor, *The Cambridge Photographic Guide to the Planets*, Cambridge University Press, 2001.

Satellite (spacecraft)

A spacecraft that is in orbit about a planet (usually the Earth) or a moon. Spacecraft are devices intended for observation, research, or communication in space. Even those spacecraft which are on the way to probe the outer reaches of the solar system usually complete at least a partial revolution around Earth before being accelerated into an interplanetary trajectory. Devices such as sounding rockets follow ballistic (approximately parabolic) paths after fuel exhaustion, but they are not satellites because they do not achieve velocities great enough to avoid falling back to Earth before completing even one revolution. *See* ROCKET ASTRONOMY; SPACE PROBE.

Orbital motion. An observer looking from a remote body might see artificial satellites in orbit around Earth. Their paths would be circular at a nearly constant altitude above Earth's land and water surfaces, or elliptical with Earth's center of mass at one focus. Once in orbit at an altitude above perceptible atmospheric drag, without further propulsive maneuvers a satellite remains in a plane through Earth's center called the orbital plane. This plane travels with Earth in its orbit around the Sun but does not rotate with Earth about its polar axis. Earth rotates around its center in the satellite's orbital plane with a period relative to the fixed stars of one sidereal day, which is 3 min 56 s shorter than the 24-h solar day.

As seen from the remote body, the motion of a satellite in its orbital plane depends on the satellite's orbital period and the inclination of its orbital plane relative to Earth's equatorial plane. In prograde (direct) orbits the satellite moves with an eastward velocity component, in the direction of Earth's rotation. By definition, the inclination of an orbital plane is 0° when it coincides with Earth's Equator and the satellite is traveling eastward. Other prograde orbits have inclinations between 0 and 90°. In retrograde orbits, which have inclinations between 90 and 180°, the satellite moves with a westward velocity component, opposite to that of Earth's rotation. A polar orbit has an inclination of exactly 90° and is neither prograde nor retrograde. The maximum inclination of 180° characterizes an equatorial retrograde orbit. *See* CELESTIAL MECHANICS; ORBITAL MOTION.

Low Earth orbits. The space shuttle, the International Space Station (ISS), and many automated (crewless, robotic) satellites travel in low Earth orbits (LEO) about 100 mi (160 km) or more above Earth's surface. They have typical orbital periods of about 90 min. These satellites have lifetimes of days, weeks, months, or years, depending on their altitudes, their mass-to-drag ratios, and atmospheric drag variations caused by solar activity. The International Space Station would have an orbital lifetime of only a few years without the periodic orbital boosts provided by the space shuttle or other rocket-powered space vehicles such as the Russian *Progress*. Most LEO satellites spend up to nearly half of their time in Earth's shadow. The space shuttle provides its electric power from fuel cells, but almost every other LEO spacecraft depends on solar cells for its power and batteries to operate through the sixteen 35-min "nights" which occur during each 24-h terrestrial day. *See* SPACE SHUTTLE.

Sun-synchronous orbits. Earth is not a perfect sphere. Its rotation causes its equatorial diameter to be 26 mi (42 km) greater than its corresponding polar dimension. For LEO satellites at altitudes below 3700 mi (6000 km), a retrograde inclination slightly greater than 90° may be selected which will cause the orbital plane to rotate eastward at exactly one revolution

per year. This equatorial bulge phenomenon has the desirable result of permitting the plane of such Sun-synchronous orbits, as viewed from the Sun, to remain in the same apparent orientation throughout the year. In more practical terms, if such a Sun-synchronous satellite crosses the Equator in Brazil at 10:00 a.m. on January 1, it will also do so on June 30 or on any other day of the year. Since the orbital plane remains fixed relative to the Earth-Sun axis, the equatorial crossing time also occurs at the same local time at any longitude. This is ideal for weather, Earth resources monitoring, and reconnaissance purposes, because shadows will fall with the same relative length in the same direction and daily weather buildups will be imaged at essentially the same stage from each orbit to the next. *See* METEOROLOGICAL SATELLITES; MILITARY SATELLITES; REMOTE SENSING.

Geosynchronous and geostationary orbits. Earth has a period of rotation relative to the fixed stars of 23 h 56 min 4 s, which is one sidereal day. A satellite orbit of this period is said to be a geosynchronous orbit (GEO). If this orbit is also circular and equatorial, the spacecraft is said to be geostationary, because it remains in a fixed position relative to any observer on the approximately one-third of Earth from which the satellite is visible. The geostationary orbit is also known as the Clarke orbit, after the science fiction writer Arthur C. Clarke, who in 1945 first suggested the use of this particular orbit for communications purposes. Its principal advantage for communications is that, once pointed at the GEO spacecraft, an antenna on Earth never needs to be repointed. *See* COMMUNICATIONS SATELLITE.

A nearly stationary circular GEO satellite with a slight inclination appears to hover over a fixed longitude centered on the Equator and moves north and south in latitude in a very narrow symmetric figure-eight. It takes one sidereal day to complete the figure. The maximum latitude excursion, north or south, equals the orbital inclination. A perfectly positioned GEO satellite is stationed directly over Earth's Equator at its assigned longitude, but is quickly displaced from this location by the combined gravitational pull of the Moon and Sun.

Operators of a geostationary communications satellite must monitor its position regularly to be sure that it remains within its prescribed longitudinal slot, which is assigned internationally by the International Telecommunication Union (ITU) and domestically in the United States of America by the Federal Communications Commission (FCC). Typical orbital slot widths are a few tenths of a degree of longitude. This dimension must be small compared with the spacing between adjacent satellites, so that the ground antenna beamwidth can be wide enough to ignore satellite position variations within the assigned slot but narrow enough to reject interfering signals from adjacent satellite transmitters. The geostationary orbit altitude is 22,237 mi (35,786 km) above Earth's equatorial surface. Below this altitude the satellite will drift eastward; above it, westward. This fact is used when a satellite must be repositioned to a different slot along the geostationary arc.

When a GEO spacecraft approaches a boundary of its assigned position, its operator must fire station-keeping jet thrusters to prevent crossing that boundary. This is no small matter. Historically up to half of the fueled spacecraft mass, upon arrival at the assigned orbital position, has consisted of station-keeping propellant which was exhausted in about a decade, thereby ending that satellite's useful life and requiring the launch of a replacement spacecraft.

In addition to communications, the GEO arc is used for weather observation spacecraft. Three such spacecraft evenly spaced along the Equator can monitor continuously severe weather around the entire globe, with the exception of regions within about 10° of the North and South poles, where hurricanes and tornadoes are absent.

Other types of satellite orbits. There are certain space science phenomena whose investigation requires direct sampling of the local environment, notably magnetic fields and associated ionized plasma particles in Earth's magnetosphere and radiation (Van Allen) belts. Many *Explorer*-class spacecraft [special-purpose smaller satellites, typically 150–500 lb (70–230 kg) mass] have been devoted to such studies. Their orbits have been quite varied. Most have traveled in highly eccentric Earth orbits, characterized by perigees (lowest altitudes) of 100–200 mi (160–320 km) and apogees (highest altitudes) out to lunar distances. *See* SCIENTIFIC AND APPLICATIONS SATELLITES.

During the 1990s, radio navigation satellite systems assumed global importance, from missile midcourse guidance and satellite, aircraft, ship, land vehicle, and hiker position location, to precision aircraft terminal guidance. The two leading systems are the U.S. Global Positioning Satellite (GPS) constellation and the Russian GLONASS system. GPS employs eight satellites in each of three highly inclined circular orbits at altitudes intermediate between LEO and GEO. GLONASS is similar. **Table 1** enumerates and gives examples of various satellite missions and payloads. *See* SATELLITE NAVIGATION SYSTEMS.

Configuration and structure. The size and shape of a spacecraft is almost always dictated primarily by its mission requirements. The principal constraints are usually imposed by the dimensions and shape of the satellite payload provisions of the launch vehicle. An important requirement of virtually all powered automated spacecraft is sufficient solar cell mounting area both to power the payload in sunlight and to charge its batteries to continue payload operations during solar eclipse periods. Another requirement is to provide spacecraft attitude stabilization and control so that sensors and antennas can be pointed in the required directions. *See* SPACECRAFT STRUCTURE.

Spin-stabilized satellites. In the 1960s and 1970s, many smaller *Explorer* and communications satellites were launched by vehicles such as the Delta, which used a spin-stabilized (same principle as a spinning rifle bullet) upper stage. This led to a spin-stabilized cylindrical (drum-shaped) satellite, symmetric about its longitudinal axis for ease of spin balancing (similar to automobile tire spin balancing) before launch, and

TABLE 1. Types of satellite missions and payloads

Mission	Payload	Example
Communications		
Full-duplex broadband	Transceiver	Intelsat, Milstar
Message broadcast	Transmitter	DirecTV, GPS
Personal communication	Transceiver	Globalstar
Remote sensing		
Imaging	Imagers and cameras	Landsat, Hubble Space Telescope
Intensity measurement	Radiometer	Chandra X-Ray Observatory
Topographic mapping	Altimeter	TOPEX/Poseidon
Navigation		
Ranging	Transceiver	TDRS
Navigation signal	Clock and transmitter	GPS, GLONASS
In situ science		
Crewed	Physical and life sciences	International Space Station
Robotic	Sample collection and return	LDEF/space shuttle
Other		
Microgravity manufacturing	Physical plant and raw materials	International Space Station
Space medicine	Medical laboratory, centrifuge	International Space Station

SOURCE: After B. Chesley, R. Lutz, and R. F. Brodsky, Space payload design and sizing, Ch. 9 in J. R. Wertz and W. J. Lawson (eds.), *Space Mission Analysis and Design*, 3d ed., Kluwer Academic, 1999.

axially oriented in the desired direction (for example, parallel to Earth's polar axis) by the launch vehicle before separation. The usual choice for solar cell mounting was the curved sides of the cylinder, leaving the flat ends free for sensors and antennas with no obstructions over an entire hemisphere. The cylindrical satellite exterior was also ideal for locating the spacecraft inside the launch vehicle protective fairing (nose cone), which usually consists of a cylindrical section of the same diameter as the rocket's fuel tanks topped by a conical tip for thermal protection from the nose stagnation temperature.

A serious disadvantage of spin-stabilized cylindrical satellites is the limitation on solar-cell area. Because of its cylindrical shape, at any given time half of the cells are shadowed by the satellite itself, and the remainder lose about one-third of their efficiency as compared to a flat solar panel, both perpendicular to the satellite-Sun line, resulting in an overall efficiency of the cylindrical geometry of 33% of that of the flat solar panel of the same area.

Three-axis-stabilized satellites. As payloads became larger, heavier, and more power-hungry, the body-mounted solar cell area limitation became more troublesome because satellite volume and weight increase as the cube of its length or diameter while its surface area increases only as its length squared. The solution to this dilemma for the larger satellites is to switch from spin stabilization to three-axis stabilization (without spin), usually with internal gyroscopes to maintain stability about each axis. These requirements lead to a rectangular or cubic box configuration, with symmetric solar panels mounted at each end of a shaft running through the satellite's center of mass, to minimize drag-induced torques. This shaft is usually perpendicular to the orbital plane, and is rotated to increase solar cell illumination by making the line perpendicular to the solar panels fall into the plane defined by the shaft and the satellite-Sun line. This box configuration has other manufacturing advantages associated with the ability to assemble each box face individually to check out subsystems thoroughly before final assembly of the entire spacecraft with its payload. A major disadvantage of the three-axis stabilized configuration is the necessity to fold the solar panels and shafts for launch, followed by deployment after satellite injection into its final orbit. Such on-orbit mechanical deployments are a major cause of mission failures and consequent insurance cost increases.

Because the space shuttle is reusable dozens of times, it was designed with a cargo bay which can accommodate payloads in a cylindrical configuration 15 ft (4.5 m) in diameter and up to 60 ft (18 m) long.

In the free-fall conditions of orbital flight, objects are essentially weightless but possess their usual mass. Consequently it is customary to refer to spacecraft component masses in kilograms, rather than in pounds-mass. Early satellites were limited to a few kilograms, but launch vehicles such as the space shuttle, the Titan IV, and the Russian Proton can place more than 20,000 kg (44,000 lbm) into low Earth orbit.

Spacecraft construction materials. Aluminum is the most common spacecraft construction material because of its favorable combination of light weight, low cost, adequate rigidity and strength, and good electrical and thermal conductivity. Certain esoteric materials such as beryllium, magnesium, and graphite-reinforced fiber plastics (GRFP) have special properties such as lighter weight, better strength-to-weight ratio, or superior rigidity. However, they are expensive and difficult to fabricate. In addition, beryllium dust is hazardous (carcinogenic) to humans, magnesium with oxygen is quite exothermic (burns at a high temperature if ignited), and GRFPs may require special treatment to prevent electrical interference or thermal hot spots because of their relatively low conductivity.

Thermal control. Satellite exteriors are usually covered partly with thermal insulation (multiple layers

of aluminized plastic) to isolate the spacecraft from the radiated temperature extremes of the hot Sun (6000 K or 10,000°F) or cold deep space (few kelvins). The remaining area is usually covered by passive thermal control surfaces designed to maintain the satellite interior at about 68°F (20°C) for optimum operation and lifetime of solid-state electronics and about 18°F (10°C) cooler for nickel-hydrogen batteries. For example, satellite exterior white paint with its high emissivity and low absorptivity works to reduce the satellite temperature, while black interior paint with its high emissivity and absorptivity helps to maintain a more uniform temperature throughout the satellite's interior.

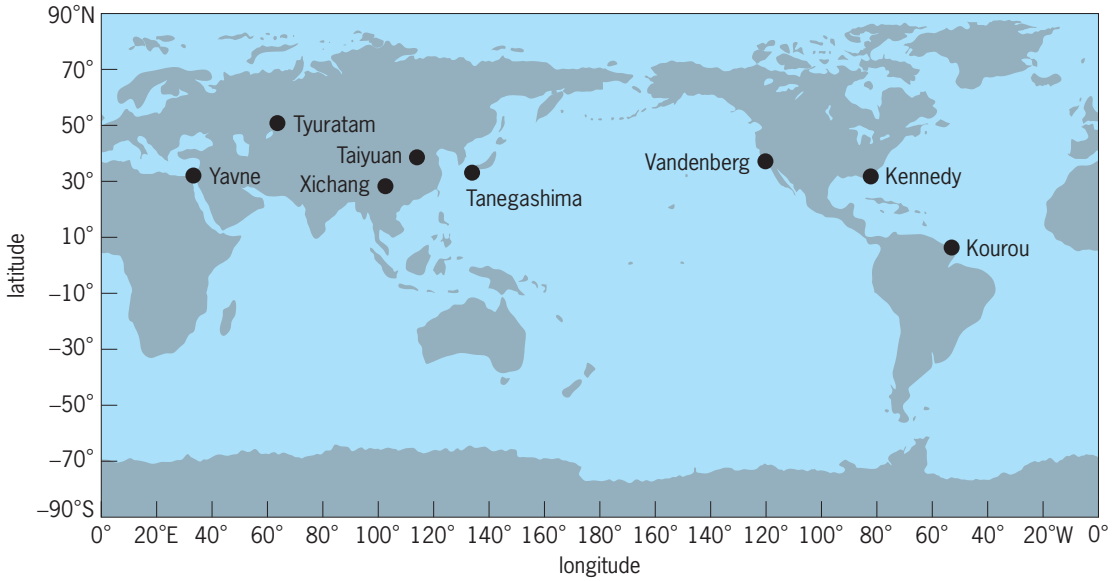
Power systems. Except for the brief trips of the space shuttle, most satellites obtain their electrical power from the Sun through large areas of solar cells. The solar panels are sized to carry the payload electrical load in sunlight plus enough additional power, stored in batteries, to power the payload during solar eclipse periods. Because solar cells degrade over time because of solar bombardment by ultraviolet radiation and energetic particles, the panels must be sized to produce the amount of power which will be required at the planned satellite end of life, rather than at the beginning of the mission. *See* SOLAR CELL; SPACE POWER SYSTEMS.

Launch vehicles. All current launch vehicles use chemical combustion of a fuel and oxidizer to generate a hot, high-pressure gas which is expanded through a supersonic nozzle to generate thrust. Solid-propellant motors cast the fuel and oxidizer together in a rubbery matrix. Liquid fuels and oxidizers are stored in separate tanks and pressure-fed into the combustion chamber just ahead of the nozzle. Liquid cryogenic hydrogen and oxygen constitute the most

TABLE 2. Launch systems characteristics[1]

Country	Launch system	Upper stage (if any)	Payload mass to low Earth orbit (LEO)[2], kg[5]	Payload mass to geosynchronous transfer orbit (GTO)[3], kg[5]	Payload mass to geosynchronous orbit (GEO), kg[5]	Payload mass to polar orbit,[4] kg[5]	Payload accommodations		
							Launch site[7]	Diameter, m[6]	Length, m[6]
United States	Atlas I	Centaur-1	—	2,255	—	—	A	3.3	10.4
	Atlas II	Centaur-2	6,580	2,810	570	5,510	A	4.2	12.0
	Atlas IIAS	Centaur-2A	8,640	3,606	1,050	7,300	A	4.2	12.0
	Delta II								
	6920/25	Star 48B	3,990	1,450	730	2,850	A, B	2.9	8.5
	7920/25	Star 48B	5,089	1,840	910	3,890	A, B	2.9	8.5
	Pegasus	—	375	—	—	—	Aircraft	1.3	4.4
	Pegasus XL	—	460	—	—	345	Launch[8]	1.3	4.4
	Shuttle		24,400	—	—	—	A	4.5	18.0
		Inertial Upper Stage	—	5,900	2,360	—			
	Taurus	STAR 37	1,400	450	—	1,060	A, B	1.4	2.8
	Titan II	No upper stage	—	—	—	1,905	B	2.8	3.7, 5.2, 6.7
	Titan IV	No upper stage	—	—	—	14,110	A, B	4.5	9.7
		Centaur	—	8,820[9]	4,540	—	A, B		12.8
		Inertial Upper Stage	—	8,350	2,380	—	A, B		15.8
		No upper stage[9]	21,645	—	—	18,600	A, B		18.0
France (European Space Agency)	Ariane 4	H-10	4,900	2,050	—	3,900	C	3.7	3.9 S
	42P	H-10	6,100	2,840	—	4,800			4.9 L
	42L[10]	H-10	7,400	3,380	—	5,900			6.5 XL
	44P	H-10	6,900	3,320	—	5,500			
	44LP[11]	H-10	8,300	4,060	—	6,800			
	44L	H-10	9,600	4,520	—	7,700			
	Ariane 5	L9	18,000	6,800	—	12,000	C	4.5	12.0
Japan	H-2	—	10,500	4,000	2,200	6,600	D	3.7, 4.8	3.5–9.0
China	Long March								
	CZ38	—	13,600	4,500	2,250	—	E	3.8	6.0
	CZ4	—	4,000	1,100	550	—	F	3.0	3.9
Russia	Proton	D1	20,900	—	—	—	G	4.1	15.6
		D1e	—	5,500	2,200	—	G	4.0	7.5
	Proton K	DM	20,100	4,815	2,100	—	G	—	—
	Proton M	Breeze M	22,000	5,100	2,500	—	G	—	—
	Zenit 2	—	13,740	5,180	1,535	11,380	G	3.4	5.9–8.4

[1]Characteristics for existing systems to 28.5° inclination, unless specified otherwise.
[2]Approximately 185 km (115 mi) circular orbit.
[3]Geosynchronous transfer orbit from the LEO altitude to the GEO altitude.
[4]90° inclination and approximately 185 km (115 mi) circular orbit.
[5]1 kg = 2.20 lbm.
[6]1 m = 3.28 ft.
[7]A = Kennedy Space Center. B = Vandenberg Air Force Base. C = Kourou Launch Center. D = Tanegashima Space Center. E = Xichang (Sichuan) Satellite Launch Center. F = Taiyuan Satellite Launch Center. G = Tyuratam (Baikonur).
[8]Carrier aircraft can stage from various locations.
[9]With solid rocket motor upgrade.
[10]With two liquid rocket boosters.
[11]With two liquid and two solid rocket boosters.
SOURCE: After J. P. Loftus, Jr., C. Teixera, and D. Kirkpatrick, Launch systems, ch. 18 in J. R. Wertz and W. J. Lawson (eds.), *Space Mission Analysis and Design*, 3d ed., Kluwer Academic, 1999.



Seven major launch sites that can do orbital launches. Their location and range safety considerations determine acceptable azimuths. (*After J. P. Loftus, Jr., C. Teixera, and D. Kirkpatrick, Launch Systems, Ch. 18 in J. R. Wertz and W. J. Larson, eds., Space Mission Analysis and Design, 3d ed., Kluwer Academic, 1999*)

powerful fuel and oxidizer combination in present use. Kerosene is sometimes substituted for hydrogen in a launch vehicle first stage, to reduce the fuel tank size because of the higher density of kerosene as compared with hydrogen.

The space shuttle uses a combination of liquid and solid motors. At liftoff the three main liquid hydrogen–liquid oxygen motors are assisted by two large solid-rocket boosters, which burn out 123 s after launch and are recovered from the ocean, refurbished, and refueled for future use. The liquid engines are used to complete the orbit insertion process, near the end of which the large external hydrogen-oxygen tank is jettisoned. Unlike the reusable space shuttle, expendable (automated, crewless) vehicles are used only once. *See* ROCKET PROPULSION.

Table 2 summarizes the capabilities of the most important launch systems from the United States, Russia, China, Japan, and France, the five world leaders in the field. The **illustration** provides a map of the launch sites referred to in the table.

A few spacecraft have been launched from international waters. Italy used *Scout* to launch *Uhuru*, an early x-ray astronomy satellite, from a converted oil-drilling platform in the Indian Ocean east of Kenya. Heavier spacecraft have been launched from the floating *Odyssey* platform in the low-latitude Pacific Ocean.

Space stations. The Russian space station *Mir* was deorbited in a controlled reentry into the South Pacific Ocean on March 23, 2001. Meanwhile the International Space Station construction proceeded so well that the station was occupied by its first permanent crew of two cosmonauts and one astronaut mission commander by the end of 2000. In March 2001 the space shuttle made the first three-person crew exchange, and it was planned to make such exchanges at similar intervals for the foreseeable future. However, the 2003 *Columbia* tragedy initiated a two-year hiatus before a shuttle returned to the International Space Station, during which the station crew was limited to one astronaut and one cosmonaut transported by the three-person Russian *Soyuz*. The crew size may be increased if and when a United States Crew Return Vehicle has been developed or an additional *Soyuz* is dedicated to this use. *See* SPACE FLIGHT; SPACE STATION.

John F. Clark

Bibliography. *Jane's All the World's Aircraft*, biennially; J. R. Wertz and W. J. Larson (eds.), *Space Mission Analysis and Design*, 3d ed., Kluwer Academic, 1999.

Satellite meteorology

The branch of meteorological science that uses meteorological sensing elements on satellites to define the past and present state of the atmosphere. Meteorological satellites can measure a wide spectrum of electromagnetic radiation in real time, providing the meteorologist with a supplemental source of data.

Since the 1960 launching of *TIROS 1*, the first satellite designed for direct visual sensing of the atmosphere, sensors on satellites have become more sophisticated. Modern satellites are sent aloft with multichannel high-resolution radiometers covering an extensive range of infrared and microwave wavelengths. Radiometers sense cloudy and clear-air atmospheric radiation at various vertical levels, atmospheric moisture content, ground and sea surface temperatures, and ocean winds, and provide visual imagery as well. *See* METEOROLOGICAL SATELLITES.

Satellite types. There are two satellite platforms used for satellite meteorology: geostationary and polar. Geostationary (geo) satellites orbit the Earth

at a distance that allows them to make one orbit every 24 hours. By establishing the orbit over the Equator, the satellite appears to remain stationary in the sky. This is important for continuous scanning of a region on the Earth for mesoscale (approximately 10–1000 km horizontal) forecasting. Geo platforms have been operating since the first *GOES* satellite launch in the late 1970s. The biggest liability to geo platforms is their cost (it takes a lot of fuel to get the satellite to this orbit), and their distance from Earth (22,000 mi; 35,000 km). At this distance, it takes much more sophisticated navigation and control to scan the Earth with accuracy to enable images to be looped (animated) smoothly and to be useful for co-locating satellite-sensed data with data gathered on the Earth. This sophistication also adds significantly to cost. The biggest advantage of geosatellites is that they can image a specific place on the Earth as frequently as every 2 minutes, allowing for unprecedented animation of meteorological phenomena, such as severe weather.

Polar satellites orbit the Earth in any range of orbital distances with a high inclination angle that causes part of the orbit to fly over polar regions. The orbital distance of 100–200 mi (160–320 km) is selected for meteorological applications, enabling the satellite to fly over a part of the Earth at about the same time every day. With orbital distances of a few hundred miles, the easiest way to visualize the Earth-satellite relationship is to think of a satellite orbiting the Earth pole-to-pole while the Earth rotates independently beneath the orbiting satellite. The advantage of polar platforms is that they eventually fly over most of the Earth. This is important for climate studies since one set of instruments with known properties will view the entire world. Another advantage is the low orbit. Since the instruments are closer to the Earth, the optics and physical constraints (especially in the microwave region) allow the instrumentation to achieve high spatial resolution at much lower cost than geo platforms. Fuel costs for launch are lower, allowing more and heavier instrumentation to be flown. The biggest liability for polar satellites is that they pass over a given spot on the Earth only twice per day. The high-resolution imagery is spectacular, but continuity is lost when the imagery is looped. For this reason, three polar satellites are flown in staggered orbits, so approximately six polar overflights occur every day over a midlatitude location. With only six passes per day over a location, this satellite is used more for climate studies than mesoscale analysis. The one exception is in the Polar Regions. There the spacecraft passes over the high latitudes on each orbit. This is very useful for high-latitude regions such as Alaska, northern Canada, and arctic regions as well as the Antarctic since geo satellites can scan only to about 75°N and 75°S latitude.

Data presentation. The enormous aerial coverage by satellite sensors bridges many of the observational gaps over the Earth's surface. Satellite data instantaneously give meteorologists up-to-the minute views of current weather phenomena.

Visual. Images derived from the visual channels are presented as black and white photographs. The brightness is solely due to the reflected solar light illuminating the Earth. Visible images are useful for determining general cloud patterns and detailed cloud structure, and they provide inherently higher resolution than the infrared channels. In addition to clouds, visible imagery shows snowcover, which is useful for diagnosing snow amount by observing how fast the snow melts following a storm. Cloud patterns defined by visual imagery can give the meteorologist detailed information about the strength and location of weather systems, which is important for determining storm motion and provides a first guess or forecast as to when a storm will move into a region. *See* CLOUD.

Infrared. More quantitative information is available from infrared sensors, which measure radiation at longer wavelengths (from infrared to microwave). Through most of the infrared spectrum, the Earth's surface and clouds absorb and reradiate as a blackbody, that is, they emit energy at infrared wavelengths with both a peak wavelength and intensity dependent on the temperature of the radiator (atmosphere, ground, or cloud). The Earth is also a very efficient absorber of infrared radiation so that infrared radiation observed by the satellite even during the daytime is coming from the Earth's radiation and is not reflected from the Sun. Thus by analyzing the infrared data, the ground surface, cloud top, and even intermediate clear air temperatures can be determined 24 hours a day. By relating the cloud top temperature in the infrared radiation to an atmospheric temperature profile from balloon data, cloud top height can be estimated. This is a very useful indicator of convective storm intensity since more vigorous convection will generally extend higher in the atmosphere and appear colder. *See* BLACKBODY; PLANCK'S RADIATION LAW.

At a few specific infrared and microwave bands, the atmospheric constituent gases absorb and reradiate. Carbon dioxide (CO_2), water vapor (H_2O), nitrous oxide (N_2O), oxygen (O_2), and ozone (O_3) have significant absorption bands at the wavelengths, shown in **Fig. 1**. These infrared radiation active regions are generally due to quantum-mechanical vibrational-rotational molecular absorption. As shown, the bands do not have sharp, steep sides; instead, absorption falls off gradually. The absorption band edges are called tails. The tails are important since they are modified by pressure through a quantum-mechanical process called pressure broadening. The result is that absorption occurring far out on a tail is associated with trace-gas absorption low in the atmosphere (at high pressures). Toward the center of the absorption band are the absorption properties of the gas higher and higher in the atmosphere. As energy radiates from the Earth's surface, it interacts with the trace gases, which modify the radiation by absorption and reradiation at each point up to the top of the atmosphere. The net result is sensed by the satellite. Infrared radiation sensors are tuned to look at different portions

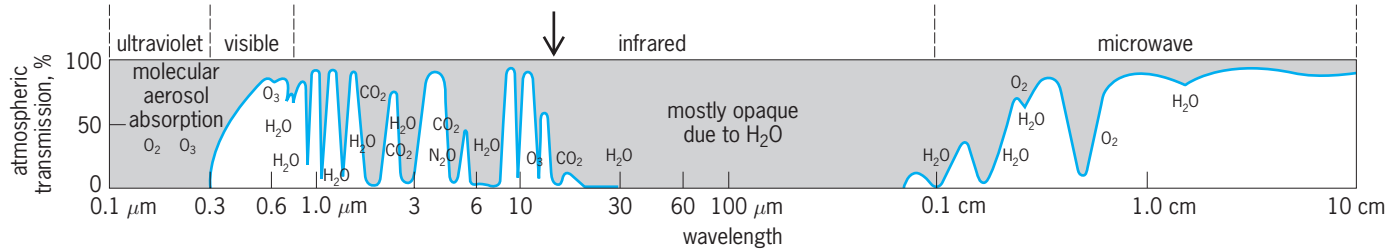


Fig. 1. Radiational transmission coefficient for a 1-km-deep (0.6-mi) atmospheric layer (100 to 0%) from ultraviolet to microwave for various atmospheric gases. 100% indicates total transmission (all wave energy passes out of the atmosphere). 0% indicates that the gas is opaque to radiation (the energy is absorbed by the gases indicated on the diagram). The arrow shows wavelengths where CO_2 is a partial absorber. These are the bands for which sensors (shown in Fig. 2) are designed to recover vertical temperature profiles. (*After D. Houghton, ed., Handbook of Applied Meteorology, Wiley, 1985*)

of the absorption tail, making it possible to establish the thermal characteristics of the atmosphere through different vertical regions or layers. Infrared radiometry is the science used to determine the thermal atmospheric profile based in part on the observed radiation intensity at the specific wavelengths that correspond to the different atmospheric levels (**Fig. 2**). Geostationary satellites have about 13 channels devoted to measuring CO_2 temperature at various atmospheric levels. The CO_2 absorption spectra is used to measure the air temperature because CO_2 exists in uniform proportion to air throughout the atmosphere. In addition to CO_2, O_2 absorption is used for thermal profiling in the microwave region. In general, the sensors mounted on both polar orbiting and geosynchronous satellites give highly useful results. *See* ELECTROMAGNETIC RADIATION; HEAT RADIATION; INFRARED RADIATION; QUANTUM MECHANICS.

Microwave. Microwave radiation is not impeded by clouds, as in the infrared bands, since radio waves can propagate through clouds without attenuation. However, the layer of atmosphere sensed by the microwave channels is greater than can be resolved by the infrared radiation, and microwave sensors have poorer horizontal resolution. The reason for the lower horizontal resolution is that longer microwave wavelengths require a large radio receiving antenna instead of optics, which are used for visible and infrared radiation measurements. Microwave sensing is constrained to polar satellites only, since it is currently impractical to fly the enormous microwave antenna to achieve comparable horizontal resolution of a polar spacecraft at geosynchronous altitudes. *See* MICROWAVE.

Wind measuring. The advent of geosynchronous satellites allowed the position of cloud elements to be traced over time. These cloud movements can be converted to winds, which can provide an additional source of data in an otherwise unobserved region. These techniques are most valuable for determination of mid- and high-level winds, particularly over tropical ocean areas. Other applications have shown that low-level winds can be determined in more spatially limited environments, such as those near thunderstorms, but those winds become more uncertain when the cloud elements grow vertically into air with a different speed and direction (a sheared environment).

Microwave data from polar satellites can measure sea surface roughness since the emissivity of the ocean changes when it contains foam, small waves, or ripples. The degree of the emissivity change can be correlated with surface wind speed. Wind direction is not possible using the microwave technique, but the speed can be inferred. *See* WIND.

Moisture. Like the other constituent gases of the atmosphere, water vapor has its unique set of absorption bands. Unlike CO_2, water vapor is not a fixed percentage of total mass in an air column. With sensors that measure radiation in the moisture bands, a complete vertical moisture profile can be developed by the same basic methods used to retrieve temperatures by studying the tails of the absorption bands. There is a very strong water vapor absorption band

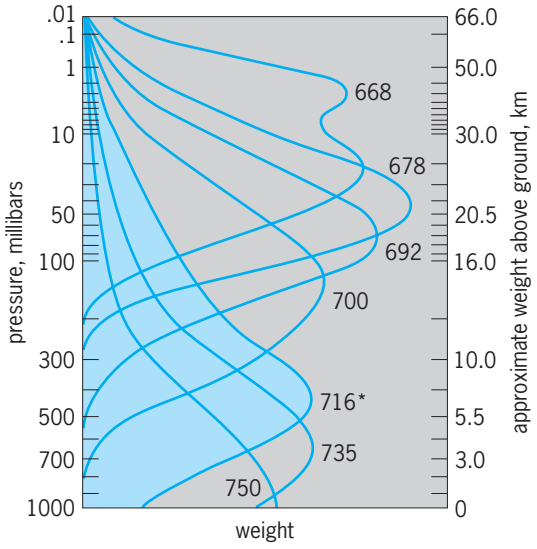


Fig. 2. Plots of transmissivity variation with height (weighting factors) for seven wavelengths (or wavenumbers, given in cm^{-1}) in the 15-m band for CO_2. The shaded weight factor representing wavelength 13.966 μm starred wavenumber 716 cm^{-1}) shows the atmospheric layers, which influence the radiance measurement at the satellite. Retrieved temperature is most representative of the layer where this weight is maximum (about 6.2 km or 3.9 mi), but note that atmospheric temperatures at levels as high as 30 km (19 mi) and as low as the surface significantly influence this value. Other curves indicate weighting factors for the other six sensor channels. Seven sensors will yield seven layer temperatures, which can be reassembled into a vertical temperature profile. 1 millibar = 102 Pa; 1 km = 0.6 mi. (*After D. Houghton, ed., Handbook of Applied Meteorology, Wiley, 1985*)

around 6.8–7.0 micrometers (Fig. 1). It is so strong that it acts like a cloud in the infrared radiation, and virtually all the sensed radiation is coming from the topmost water vapor layer in the atmosphere. Image loops of this water vapor infrared radiation imagery reveal many of the air motions in clear air. Work is under way to derive water vapor winds much like cloud drift winds only by sensing water vapor in this part of the infrared radiation spectrum.

At infrared radiation wavelengths near 12 μm, there exists a very weak water vapor absorption band that can be exploited to measure the total water vapor in cloud-free areas since water will only slightly attenuate the infrared radiation. Near this water vapor band is the infrared radiation window at 11.2 μm. At 11.2 μm, no absorption takes place from any atmospheric gas. Total water vapor in an atmospheric column can be estimated by comparing radiances from these two very close wavelengths—since if there were no water vapor, the radiation in the 12 μm region would show no absorption, as would the measurement at 11.2 μm. A decrease in intensity in the 12 μm measurements is directly related to total water vapor in an atmospheric column.

Applications. By using a wide variety of sensors, satellite data provide measurements of phenomena from the largest-scale global heat and energy budgets down to details of individual thunderstorms. Having both polar orbiting and geosynchronous satellites allows coverage over most Earth locations at time intervals from 3 minutes to 3 hours.

Global (large) scale. Polar orbiting satellites have had an impact on the monitoring of global climate and climate change. Long-term observational programs, notably the Earth Radiation Budget Experiments (ERBE), active since the late 1970s, have produced information in the form of colorized images of the outgoing long-wave radiation, a quantity that can aid in detection of long-term warming or cooling. Satellites have also been used to monitor atmospheric ozone, a gas that protects the Earth from harmful solar radiation. The ozone hole that forms over the Antarctic during October probably would not be known even today had it not been for polar orbiter satellite data.

Satellites can closely monitor land and sea temperatures and show anomalies that produce phenomena such as El Niño, which shifts the storm track of the eastern Pacific region. Satellite monitoring of the polar ice packs helps identify important seasonal variations. Ice caps and areas where vegetation is decreasing (owing to drought or human exploitation) modify the albedo (fraction of incident radiation reflected) and influence the effective solar radiation reaching the Earth. *See* ALBEDO; INSOLATION.

The great semipermanent cloud bands of the Earth, such as those along the Intertropical Convergence Zone, can be monitored by either polar orbiting or geosynchronous satellites. Long-term perturbations in the tropical regions can lead to excessive rain or drought and influence the frequency of tropical storms. Long-term averaged satellite cloud imagery is a basis for estimating rainfall in the tropics.

With observations using space-borne instruments carried out for many years, satellites are the prime tools for monitoring global climate and climatic change. *See* CLIMATE HISTORY.

Storm (synoptic) scale. Superimposed on the great atmospheric circulations are the transient eddies or cyclones, which ultimately maintain the heat balance of the Earth by moving warm air northward and cold air southward. Of course, these eddies and swirls are the major weather-producing phenomena over most of the globe. Storms are regions of concentrated air circulation (or relative vorticity), with rotation counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere. The centers of the eddies are generally associated with relative vorticity maxima. As these centers move along in the background global circulation, certain adjustments of the atmosphere produce vertical lift of the air, leading to the clouds and weather experienced at a ground location. The spiral or comma-shaped patterns often seen in visual imagery are in reality reflections of these vorticity centers moving through the atmosphere (**Fig. 3**). As the intensity of the circulation increases, the cloud material, which has been generated ahead of the circulation center by lifting, begins to deform, eventually resulting in a spiral or comma-shaped cloud band. In clear air, the spiral shape can be seen in the water vapor imagery. The cloud band or tail elongates in regions of horizontal speed variation (Fig. 3*f*). Elongated cloud bands are common in regions where winds stretch neighboring air parcels apart (equatorward of jet stream bands) in regions of atmospheric deformation. Geostationary satellites globally monitor temperature, moisture, and cloud

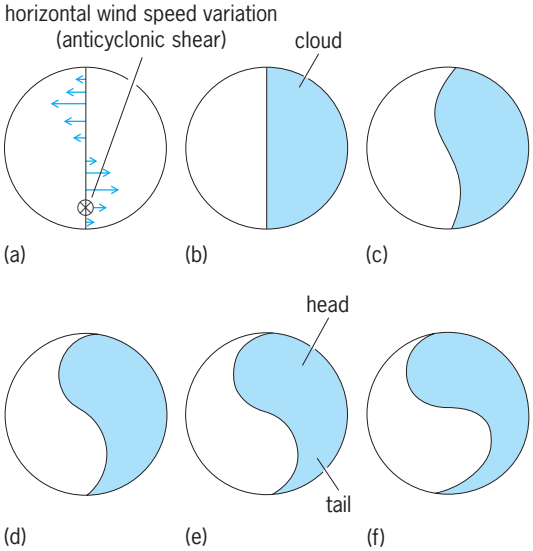


Fig. 3. Schematic diagram of a circulation center or vorticity maximum and the formation of the comma-shaped cloud pattern. (*a*) Relative wind circulation (wind vectors along a north-south axis). (*b*) Lifting forms cloud (shaded region) ahead of the center, and with time (*c*) 6 h, (*d*) 12 h, (*e*) 18 h, and (*f*) 24 h later, the circulation transforms the cloud into the comma shape seen typically in satellite imagery. The comma head is in the north, the tail is in the south. (*After R. Weldon, Cloud Pattern and Upper Level Winds, NESDIS Satellite Training Course, NOAA, 1979*)

patterns associated with these systems. The satellite measurements are combined with other weather data to initialize computerrun mathematical models, which routinely forecast major cyclones.

In tropical latitudes, widespread convective clouds or thunderstorms leave much cloud debris and moisture at very high levels. Hadley cell circulation moves this air poleward; the northward limit of this air at high levels is the subtropical jet stream. Satellite imagery typically shows extensive elongated bands of high cirrus clouds at these latitudes (10–30°N) with very sharp cloud edges. Many cloud bands seen on satellite imagery correspond closely to regions of the atmosphere where neighboring air parcels are pulled apart along the so-called axis of dilation in a zone of atmospheric deformation. Jet streams are characteristically associated with deformation zones, so the position of cloud bands can often be used to locate these strong wind currents. *See* ATMOSPHERIC GENERAL CIRCULATION; FRONT; JET STREAM.

Figure 3*f* shows how the elongation of the tail of the comma cloud appears in a region where neighboring air parcels are in horizontal shear. Thus, the same general rule applies in locating the major jet stream bands existing in the intense westerly currents at higher latitudes. Midlatitude jet stream flows indicate the position of the polar front, where there exists an intense temperature contrast through the depth of the troposphere. The importance of locating these frontal zones or jets is that the eddies or cyclones previously discussed typically have their origin here. The meteorologist, by noting past locations and apparent qualitative strength of jet streams and storms, and by using satellite-improved, model-generated forecasts, is able to make more accurate weather predictions for a particular location.

The information that can be extracted from water vapor imagery is important for analysis of synoptic-scale upper-level flow patterns. Color infrared radiation images are used primarily for cloud top and surface temperature (see **colorplate**). **Figures 4** and **5** illustrate the application of water vapor (WV) imagery in analyzing 500-hectopascal

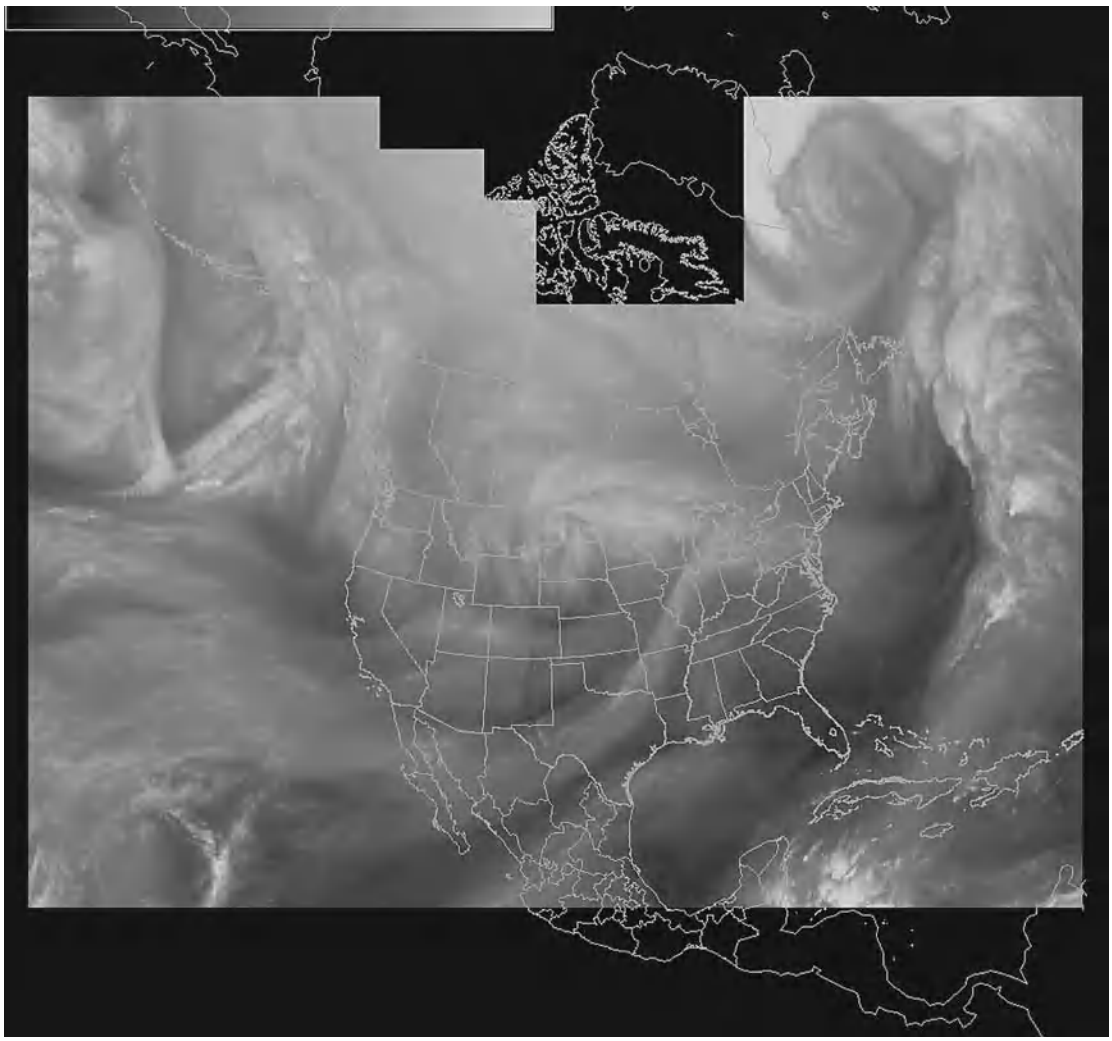


Fig. 4. Water vapor image at the same time and location as in the colorplate. Water vapor imagery can show features in clear areas such as the dark variations imaged over northern Kansas, central California, and another extending from northern Baja, southern Arizona, and New Mexico, up through the Oklahoma panhandle. Each of these areas indicates an upper-level "short wave" that is not evident at all in the cloud image in the window infrared radiation (see colorplate). The closed low pressure in southern Greenland and the system in the Gulf of Alaska are also evident in this image.

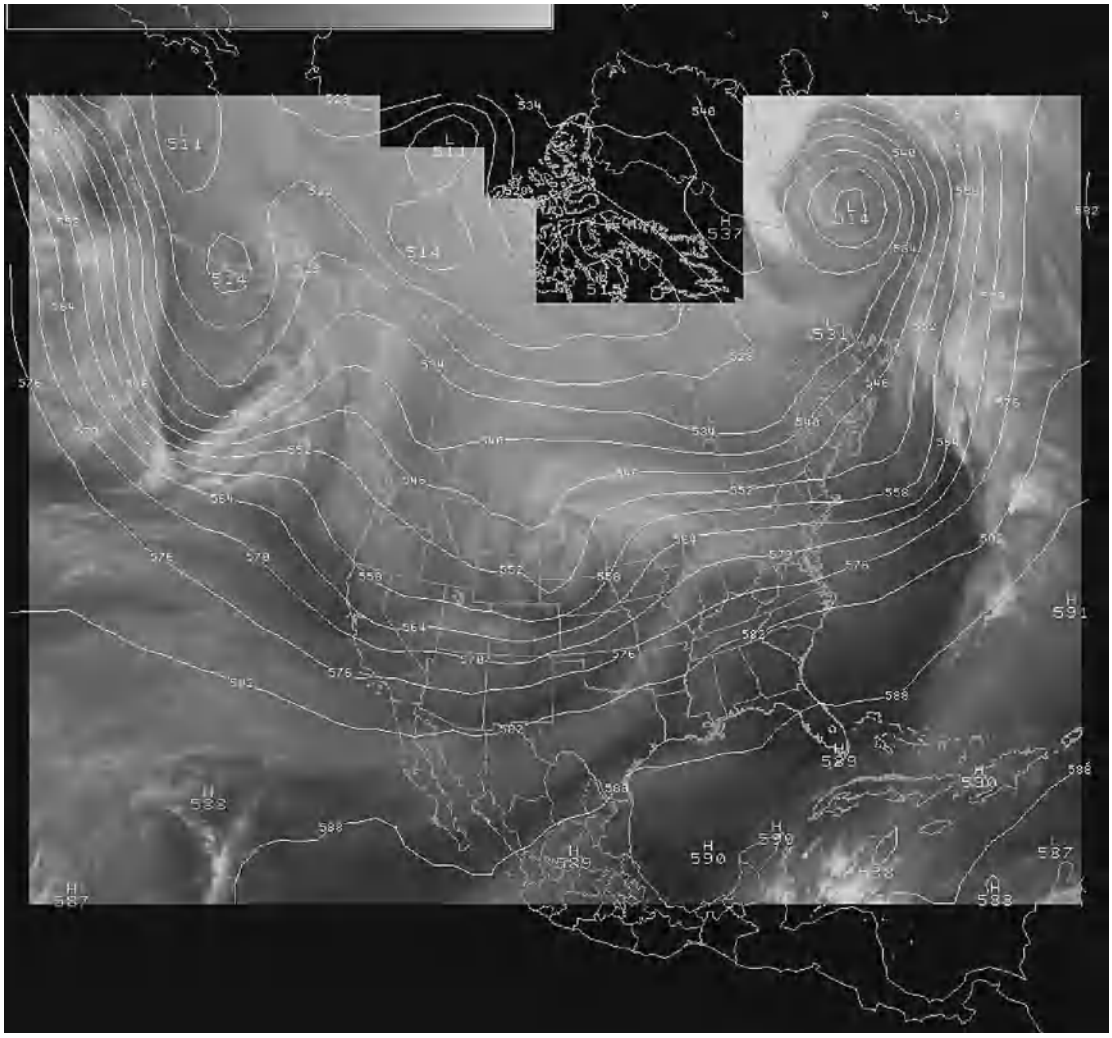


Fig. 5. Same water vapor image as shown in Fig. 4 but with contours depicting 500-hectopascal heights. The analysis has a resolution of about 60 km, whereas the image can show up to 8-km detail. The upper-level height contours align well with the water vapor imagery features. The deep trough in the Gulf of Alaska also is depicted well; the analysis indicates a very weak closed low. The imagery shows the complete circulation in the closed low over southern Greenland. South of that location can be seen another front and corresponding upper-level trough over the Atlantic Ocean. Other subtle features are the short wave over Nebraska with the dark (warm) band at its eastern edge. Another short wave is seen extending from central Missouri south through Oklahoma and the Texas panhandle. In addition, the imagery shows a similar short wave over California; in fact, the short wave might be composed of three parts, which would be nearly impossible to resolve in the height analysis. The extra detail over California is typical of the added resolution that can be inferred from satellite data. (*NOAA Forecast Systems Laboratory*)

(millibar) heights. Image and analysis overlays are very useful for determining whether the analysis is in agreement with atmospheric circulations (Fig. 5). If the image features and analysis contours are not well aligned, it implies that model forecasts initialized from the analysis may not produce an accurate forecast. This is especially important for systems over oceanic regions where conventional data are not available for the analysis. Research is focusing on methods to extract image features and automatically incorporate them into the analysis used for model forecasts. However, for now the imagery is primarily used to assess the analysis quality as has been demonstrated here.

Tropical storms. The greatest gain with the introduction of weather satellites was in early detection, positioning, and monitoring of the strength of tropical storms (hurricanes, typhoons). Lack of conventional meteorological data over the tropics (particularly the oceanic areas) makes satellite data indispensable for this task. The hurricane is one of the most spectacular satellite images. The exact position, estimates of winds, and qualitative determination of strength are possible with continuous monitoring of satellite imagery in the visible channels. In addition, infrared sensors provide information on cloud top height, important for locating rain bands. Microwave sensors can penetrate the storm to provide an indication of the interior core's relative warmth, closely related to the strength of the hurricane, and sea surface temperature to assess its development potential. *See* HURRICANE; TROPICAL METEOROLOGY.

Regional (meso) scale. Most significant weather events experienced by society—heavy rain or snow, severe thunderstorms, or high winds—are organized by systems that have horizontal dimensions of about

60 mi (100 km). These weather systems, known as mesoscale convective systems, often fall between stations of conventional observing networks. Hence, meteorologists might miss them were it not for satellite sensing.

Satellite data are critical in the period 2–3 hours before the development of regional weather phenomena. Experienced satellite meteorologists can recognize cloud patterns that mark areas of potential development (converging winds, rapidly growing clouds), and can issue forecasts or advisories long before the weather event strikes. Even in situations where no clouds exist initially, satellites can estimate the instability of air columns by measuring the thermal stability and vertical moisture distribution. Measurement of surface temperatures, as derived from infrared channels in clear-sky situations, significantly aids in accurately determining the surface air temperature, particularly when combined with a few observations from sites at ground level. This can help the forecaster locate potential severe storm regions up to 6 hours before development, such as the development of thunderstorm squall lines. Hours before storm development, upper-level systems can trigger mesoscale convective events, which are often revealed by patterns in water vapor imagery and can be tracked. By frequently monitoring the visual or infrared satellite imagery, the meteorologist can recognize lines of convergence (where air is flowing together) and note where low clouds seem to be merging and organizing. Later, as the forced upward motion allows the release of instability, vertical development of the cloud can be easily seen in visual imagery and confirmed by infrared. Finally, when the storm tops reach the tropopause, environmental winds and storm-generated pressure forces cause the tops to spread horizontally in cumulonimbus anvils. By noting the rapidity of development, cloud top heights, and anvil spreading rates, the forecaster can estimate the strength of the thunderstorm complex. At low levels, air cooled by precipitaion also spreads rapidly outward away from the squall line. Positions of associated arc cloud lines (named for their shape) are important for determining when a location will experience gusty winds and where new storms may develop. When severe storm systems remain stationary, flash floods can result. **Figure 6** shows the rapid development of a squall line.

Heavy snows or rains associated with large cyclones are often produced by mesoscale-sized thunderstorm systems that can be easily seen and tracked by (visual and infrared) satellite imagery. *See* MESOMETEOROLOGY; SQUALL LINE; STORM; THUNDERSTORM; TROPOPAUSE.

Small (cloud) scale. Although weather radar is the major source of data at this scale, individual clouds can be closely watched with respect to structure, height, and location (and hence movement) by frequent satellite scans. The geo satellites are set up to scan a particular location every few minutes. When the data are combined in computer-based weather workstations, which can rapidly process, display, and animate the imagery, near-real-time monitoring of individual storms is possible. Large thunderstorms with tornadic potential often exhibit a tilted or sheared appearance, with rapid growth on their rear flank, which often can be seen from satellites before tornado development. The rate of growth of single cells can be determined. Rain-cooled, low-level outflow from a single thunderstorm can often be seen. Intersections of these outflow air masses indicate a strong potential for new cell development. Satellite research has shown that at the anvil level, tracking

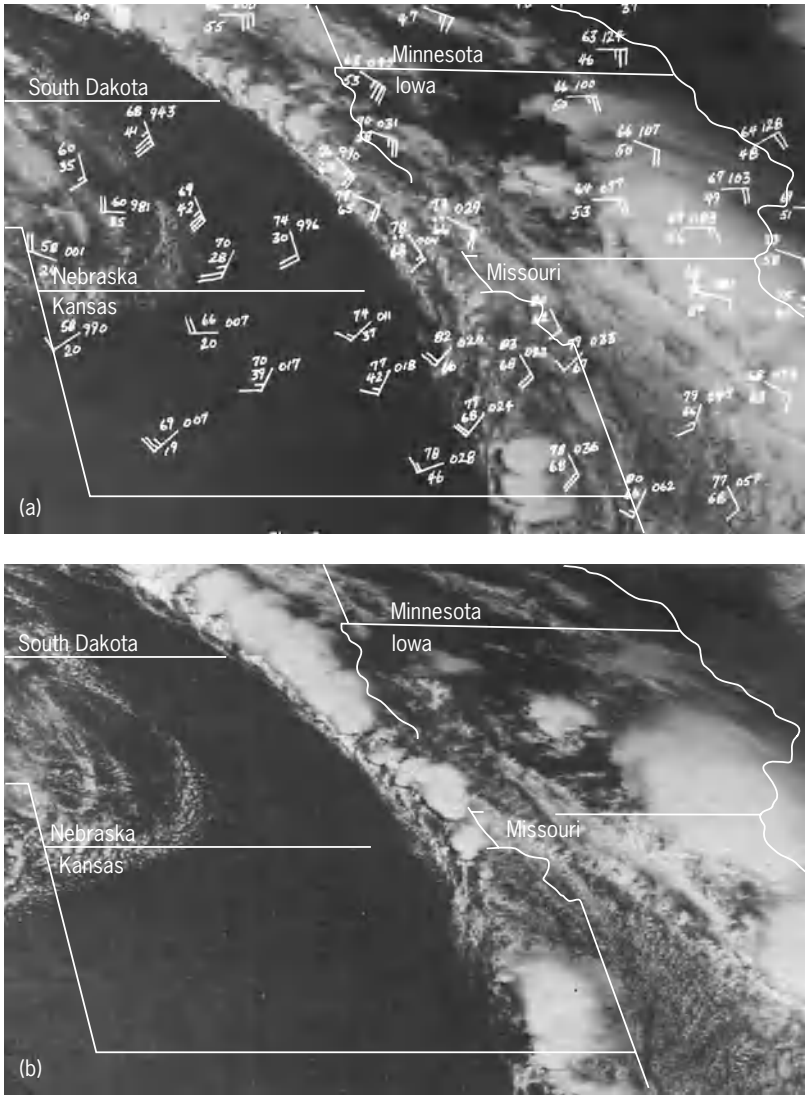


Fig. 6. Use of satellite imagery for severe thunderstorm forecasting. The two visual images were taken 1 hour apart at 18 GMT (13 EST) and 19 GMT (14 EST) on May 6, 1975. The region includes Nebraska, Kansas, Iowa, and Missouri. (*a*) Surface weather observations superimposed over the visual cloud image at 18 GMT. Surface parameters plotted are wind direction and speed [each large barb is 10 knots (5 m/s), half barb 5 knots (2.5 m/s)], temperature (°F) in upper left of a barb, dewpoints (°F) in lower left, surface pressure in upper right (for example, 029 is 1002.9 millibars or 100.29 kilopascals). Note newly developed thunderstorms in northeastern Nebraska. At least three cloud lines extend southward into a region of converging airflow. Dry air is in central Nebraska and western Kansas (shown by clear skies and low dewpoints). Forecasters issued a tornado watch for most of eastern Kansas and Nebraska based on conventional weather information and forecast models. Hardening of clumping of clouds on the line shows that this region is ready to explode. (*b*) At 19 GMT, massive storms developed over southeastern Nebraska. Soon after, the storms just northwest of the Missouri-Iowa-Nebraska juncture spawned a tornado, which devastated Omaha. In northern Nebraska, note how quickly the anvils (smooth-textured clouds) have expanded in 1 hour, indicating strong thunderstorm updrafts. Tornadoes occurred here as well. (*J. Purdom, National Earth Satellite Service, NOAA*)

cloud elements can lead to estimates of cloud top divergence (airflow spreading) and, with a few assumptions, an estimate of the vertical air speed or updraft of the storm. This is an important development since the magnitude of the updraft is related to hail production and tornadic potential. Tops that overshoot the anvil level can be seen on both visual (at low sun angles) and infrared imagery. Such occurrences indicate intense updrafts capable of forcing air well above the tropopause. Quite frequently, the anvil exhibits an enhanced V shape (open end downwind), which has been correlated with severe weather occurrences at ground level.

Rainfall estimates derived from satellite imagery, although crude, can guide the forecaster in issuing quantitative precipitation forecasts. Satellite data can be a valuable complement to weather radar by allowing simultaneous observation of the entire cloud and its associated features. Intense precipitation, like other thunderstorm-produced weather events, is related to updraft strength and hence rapid anvil spreading. The storm-relative location of the heaviest precipitation lies typically beneath strong gradients of the cloud top temperature on the upwind (usually western) edge of the infrared cloud image. Empirical quantitative methods under development show that it may be possible to estimate a rainfall rate. Research indicates that radiances measured by satellites may be used to classify cloud water drops by size. *See* HAIL; METEOROLOGY; PRECIPITATION (METEOROLOGY); TORNADO; WEATHER FORECASTING AND PREDICTION.

Daniel L. Birkenheuer; John A. McGinley

Bibliography. D. Houghton (ed.), *Handbook of Applied Meteorology*, 1985; S. O. Kidder and T. H. Vonder Haar, *Satellite Meteorology: An Introduction*, 1995; U.S. Department of Commerce, Environmental Research Laboratories, *Thunderstorms: A Social, Scientific and Technological Documentary*, vol. 3, chap. 11, 1982; R. Weldon, *Cloud Pattern and Upper Level Winds*, NESDIS Satellite Training Course, NOAA, 1979.

Satellite navigation systems

Radionavigation systems using artificial satellites as sources of radio signals and position references. Development of satellite-based systems for global positioning and navigation began almost immediately after the launch of *Sputnik I* by the Soviet Union in 1957. The U.S. Navy Navigation Satellite System, better known as *Transit*, became operational in 1964. The Soviet Union responded with its own satellite navigation systems, *Parus* and *Tsikada*. All three systems were based on measurement of Doppler shift in signals received from a satellite, and were aimed mainly at offering navigational guidance to ships and submarines. *See* DOPPLER EFFECT.

Starting in the mid-1990s, the term satellite navigation system became synonymous with the U.S. Navstar Global Positioning System (GPS), but this would change. GPS is used every day by millions around the world for accurate estimates of position, velocity, and time. These estimates can be obtained instantaneously and continuously anywhere and by anyone equipped with an inexpensive, pocket-sized GPS receiver. The only requirement is that the receiver must be able to "see" at least four GPS satellites in the sky.

The success of GPS has led to efforts by other countries to develop systems modeled after it. The development of GLONASS began in the Soviet Union a few steps behind GPS. The system was acquired by Russia following the breakup of the Soviet Union, but it is not yet ready for operational use. GPS and GLONASS were designed during the Cold War and both started out as military systems which offered a subset of their signals for civil use. By contrast, Galileo, a European system on the drawing boards in 2005, is being designed primarily as a civil system. GPS, GLONASS, and Galileo are based on the same principles and exploit the same technologies. The generic name for each is Global Navigation Satellite System (GNSS).

Principles of satellite navigation. The basic idea behind position estimation with GPS (as well as GLONASS and Galileo) is very simple: One's position can be determined if one knows, or can measure, distances to objects whose position coordinates are known. Estimation of a position based on measurement of distances is referred to as trilateration. The technology for accurate measurement of long distances did not become available until the middle of the twentieth century, and before then the more common technique of positioning was known as triangulation. It was based on measurement of angles to objects whose positions were known.

The basic idea of trilateration is easily illustrated for positioning on a plane (**Fig. 1**). An observer measures his or her distance (or range) r_1 from station S_1 whose position coordinates are known. The observer must be located somewhere on a circle with a radius of r_1 centered at S_1. Range measurement r_2 from station S_2 defines another circle, and reduces the uncertainty in the observer's position to the two points (P and P') where the circles intersect. The observer may be able to reject one of the estimates on the basis of prior information. If not, a third measurement r_3 from station S_3 determines the position uniquely.

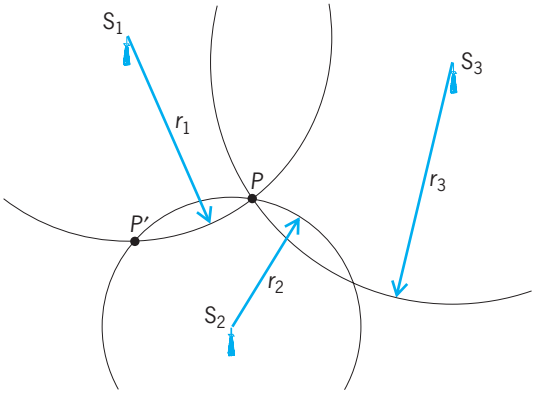


Fig. 1. Positioning by trilateration. The measured ranges to three position references, S_1, S_2, and S_3, determine the position of *P*.

This idea extends to positioning in three dimensions, with the circles replaced by spheres. The problem is formulated algebraically. Each distance r_1, r_2, and r_3 can be related to the known position coordinates of the three stations and the three unknown position coordinates (x, y, z) in the form of an equation. The three equations can be solved for the three unknowns. *See* ANALYTIC GEOMETRY.

Key enabling technologies. While the idea of trilateration is not new, a global navigation satellite system based on this idea would not have been feasible without several technologies which were developed or came to maturity in the second half of the twentieth century. The key technologies are stable space-based platforms in predictable orbits, extremely stable atomic clocks, spread spectrum signals, global coordinate frames, and microelectronics.

Recall that trilateration requires measurement of ranges to three position references (that is, objects at known positions). A satellite navigation system uses satellites as the position references. A satellite in a medium Earth orbit (for example, GPS) moves at a speed of about 4 km/s (2.5 mi/s), but its position at any instant can now be predicted 24–48 h earlier with an error no larger than a few meters.

The next two key technologies concern measurement of ranges in terms of transit times of signals from the satellites to a receiver. Roughly speaking, if meter-level positioning accuracy is required, measurement of ranges must be accurate to meter level and, therefore, measurement of transit times must be accurate to the nanosecond (10^{-9} s) level. (Radio waves travel at the speed of light, which is approximately 0.3 m/ns or 1 ft/ns.) Such precision is obtained by marking the times of transmission on the signals in accordance with ultrastable cesium and rubidium atomic clocks carried aboard the satellites. These clocks have stability of better than one part in 10^{13} and can be kept synchronized at the nanosecond level. The precise measurement of arrival times is made possible by broadcasting spread spectrum signals, which have wide bandwidths but allow each satellite to transmit its unique signal over a common frequency band. *See* ATOMIC CLOCK; SPREAD SPECTRUM COMMUNICATION.

A global positioning system requires a global coordinate frame in which to express the positions of its satellites and the users. Previously, the coordinate frames were defined at most on a country or regional basis. GPS provided the impetus for development by the U.S. Department of Defense of the World Geodetic System 1984 (WGS 84), which has now become a de facto international standard. *See* GEODESY.

Finally, the spectacular developments in microelectronics in the last quarter of the twentieth century have made possible the widespread use of GPS by producing receivers that are light, compact, and an order-of-magnitude less inexpensive than thought possible in 1980.

Pseudorange. In order to measure range to a satellite (that is, transit time of a signal from a satellite to a receiver), clearly, the clocks in the satellite and the receiver must be maintained in synchronism.

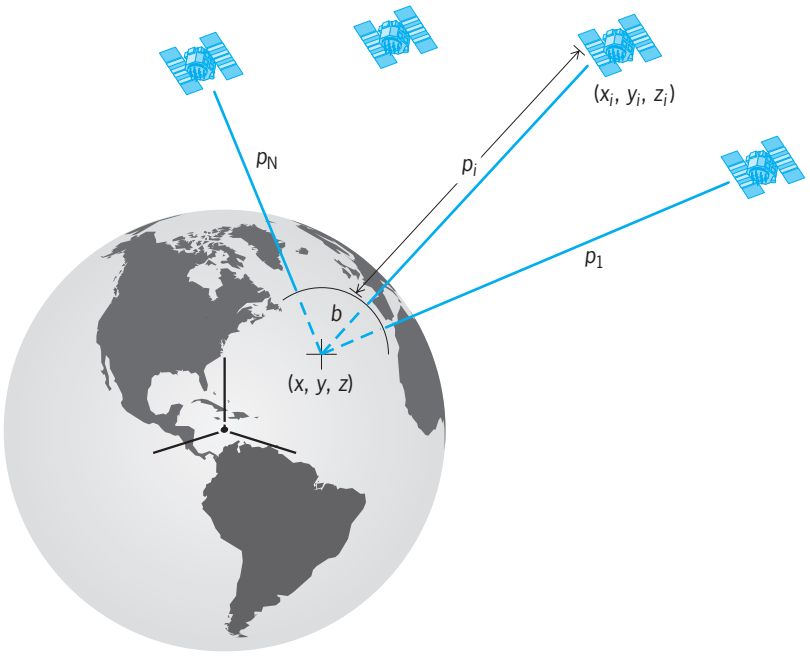


Fig. 2. Principles of satellite navigation. User-satellite range measurements based on transit times of signals are biased by a common amount and are called pseudoranges. For a constellation of N satellites, the pseudoranges p_i, $i = 1, 2, \ldots, N$, are given by $p_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} - b$, where the (x_i, y_i, z_i) are the known satellite positions. Pseudorange measurements are needed from at least four satellites ($N \geq 4$) to estimate the user position (x, y, z) and receiver clock bias b (in units of length).

Fortunately, this onerous requirement is easily sidestepped, allowing use of inexpensive quartz oscillators in GPS receivers. The bias in the receiver clock at the instant of the measurements affects the observed transit times for all satellites equally. The corresponding measured ranges are thus all too short, or all too long, by a common amount, and are called pseudoranges. The receiver clock bias, thus, becomes the fourth unknown to be estimated, in addition to the three coordinates of position. A user, therefore, needs a minimum of four satellites in view to set up four equations in the four unknowns: three coordinates of spatial position plus time (**Fig. 2**).

The pseudorange constitutes the main observable of a satellite navigation system. A second, more precise observable is the change in pseudorange, called accumulated delta range, used in applications requiring centimeter-level accuracy in relative positioning.

Range measurement errors and positioning accuracy. The quality of a position estimate obtained from a GNSS basically depends upon two factors: the accuracy of the pseudorange measurements and the spatial distribution of the satellites around the user. Multiple sources contribute to pseudorange errors. There are system errors: errors in the predicted satellite position and the satellite clock synchronization error. There are also errors in relating the measured signal transit time to pseudorange: While a radio signal travels in a straight line at the speed of light in space, the last 1000 km (600 mi) of its journey is through the variable environments of charged particles (ionosphere) and the neutral, gaseous atmosphere (troposphere), where the signal speed varies. Closer to

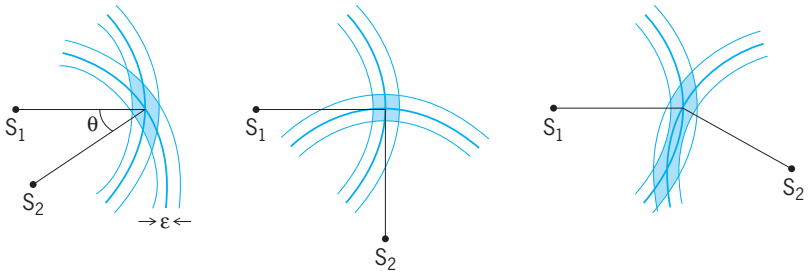


Fig. 3. Role of measurement geometry in positioning. The quality of a position estimate depends upon both the quality of the range measurements (ε) and their geometry (θ). Shaded regions represent the uncertainties in the position estimates.

the receiver, a signal may reach the antenna through multiple paths due to reflections from nearby surfaces and, finally, there is the inescapable thermal noise corrupting the signals. With careful control of the system errors and compensation for the atmospheric propagation conditions, pseudoranges can now be measured with an accuracy of several meters.

The spatial distribution (or geometry) of the satellites around the user also affects the quality of the position estimate. Satellite geometry is often characterized in terms of a parameter called dilution of precision (DOP). The more favorable the geometry, the lower the DOP. The basic idea can be illustrated simply for positioning on a plane (**Fig. 3**). For best three-dimensional positioning, a user would be "surrounded" by the satellites, several of which would be low in the sky and some high. With a 24-satellite constellation, poor satellite geometries are infrequent and change quickly.

GPS. GPS was developed in the 1970s and 1980s by the U.S. Department of Defense, and was declared ready for operational use in 1995. The system was intended primarily for the U.S. military; civil use was a secondary objective. GPS has proven to be a great success on both fronts. Its military value as a "force enhancer" surpassed expectations in the Gulf War (1990–1991), U.S interventions in Kosovo (1999), and Afghanistan (2001–2002), and the Iraq War (2003). And through its many civil applications in industry, commerce, and outdoor recreational activities, GPS is becoming essential to daily life as a new utility.

System architecture. GPS can be thought of in terms of its three elements, generally referred to as segments: space segment, control segment, and user segment. The space segment comprises the satellite constellation. The baseline constellation consists of 24 satellites in circular orbits with a radius of 26,560 km (16,504 mi) and orbital period of about 12 h. The satellites are arranged in six planes containing four satellites each and inclined at 55° relative to the equator (**Fig. 4**). With this satellite constellation, more than 99% of the GPS users worldwide with a clear view of the sky have at least four satellites in view at all times.

The first generation of GPS satellites, called Block I, was launched starting in 1978 to demonstrate the feasibility of the system. The operational system was subsequently populated with the second-generation satellites called Block II and Block IIA. The satellites launched in 2004–2005 to replace these operational satellites at the end of their service lives are called Block IIR.

The control segment encompasses the facilities and functions required for operating GPS. The facilities include crewless monitoring stations and uplink radar sites distributed around the globe, and the Master Control Station located at Schriever Air Force Base in Colorado. The main functions include satellite health monitoring, and prediction and upload of satellite orbits and clock parameters.

The user segment encompasses all aspects of design, development, production, and utilization of user equipment for the different applications.

Positioning services and signals. GPS offers two kinds of service: Standard Positioning Service (SPS) for open civil use and Precise Positioning Service (PPS) for users authorized by the Department of Defense. Access to the full capability of the system (that is, PPS) is restricted through cryptographic techniques. It rankled the civil user community that the SPS signal was degraded throughout the 1990s by introducing controlled errors to reduce its precision and limit the positioning accuracy to 60–100 m (200–330 ft). This feature, called selective availability (SA), was deactivated in 2000. Since then, the horizontal positioning error for most users has been less than 10 m (33 ft).

As of June 2005, each GPS satellite transmits continuously on two radio frequencies in the L-band (1 GHz–2 GHz), referred to as Link 1 (L1) and Link 2 (L2), with center frequencies of 1575.42 and 1227.60 MHz, respectively. Two signals are transmitted on L1, one for the civil users and the other for the Department of Defense authorized users. The lone signal on L2 is intended for the Department of Defense authorized users only (**Fig. 5**).

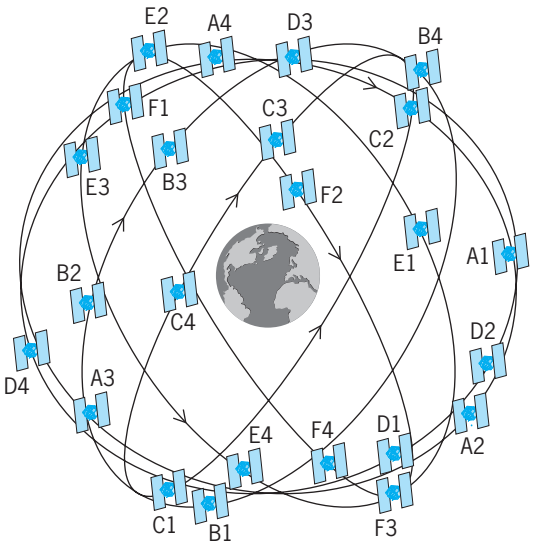


Fig. 4. GPS baseline constellation. Each satellite is identified by a two-character code: A letter identifies the orbital plane (A through F), and a number identifies the satellite slot number in the plane (1 through 4).

A signal consists of three components: the radio-frequency (RF) carrier, the ranging code, and navigation data (**Fig. 6**).

Ranging code. Associated with each Service (that is, SPS and PPS) is a family of binary codes called pseudo-random noise (PRN) codes. The PRN codes have special mathematical properties that allow all satellites to transmit in the same frequency band without interfering with each other. These codes also enable precise range measurements. The SPS codes are called coarse/acquisition codes (C/A codes) and the PPS codes are referred to as precision/encrypted codes, or P(Y) codes. Each satellite transmits a C/A code on L1 and P(Y) codes on both L1 and L2.

Each C/A code is a unique sequence of 1023 bits, called chips in this context, which is repeated each millisecond. The duration of each C/A-code chip is about 1 microsecond. Equivalently, the chip width or wavelength is about 300 m. The rate of the C/A-code chips, called the chipping rate, is 1.023 megachips per second (Mcps). A coarse/acquisition code is so-called because it is coarse as compared to a precision code, and it was originally intended to serve as a necessary steppingstone for acquisition of the much longer P(Y) code. Newer technology now allows direct acquisition of the P(Y) codes.

Navigation data. The navigation message, transmitted at a leisurely 50 bits per second (bps), consists of data on the satellite health status, ephemeris (satellite position and velocity), clock bias parameters, and an almanac giving reduced-precision ephemeris data for the entire constellation. It takes 12.5 min for the entire message to be received. The essential satellite ephemeris and clock parameters are repeated each 30 s.

Signal composition and modulation. The three components of a signal (carrier, code, and navigation data) are derived coherently from one of the atomic standards aboard the satellite. The code is combined with the binary navigation data using modulo-2 addition: If the code chip and the data bit are the same (both are 0s or both are 1s), the result is 0; and if they are different, the result is 1. The composite binary signal is then impressed upon the carrier in a process called modulation. The specific form of modulation used is called binary phase-shift keying (BPSK): a 0 bit leaves the carrier signal unchanged; and a 1 bit multiplies the carrier by −1, which is equivalent to shifting the phase of the sinusoidal signal by 180° (Fig. 5). *See* MODULATION; PHASE MODULATION.

The modulation of a carrier by a binary code spreads the signal energy, initially concentrated at a single frequency, over a wide frequency band: over 2 MHz for the C/A code and about 20 MHz for the P(Y) code, centered at the carrier frequency (**Fig. 7**). While the signal power is unchanged, this step reduces the power spectral density below that for the background RF radiation. Such signals, referred to as spread spectrum signals, have many properties which make them attractive for use in communication and navigation. The signal energy can be "despread" in a receiver if the code is known. In principle, keeping a code secret limits access to the signal.

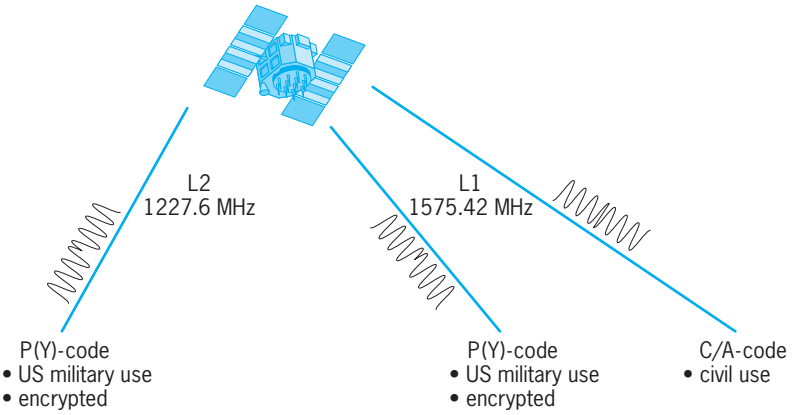


Fig. 5. GPS signals in June 2005. Each satellite transmits three signals: one for open civil use on L1, and one each on L1 and L2 for users authorized by the U.S. Department of Defense. Satellites launched starting in July 2005 transmit additional signals.

The GPS signals are extremely weak. The signal power reaching a terrestrial antenna is only about 10^{-16} W (−160 dBW). The ambient radio noise power associated with the signal is 100 to 1000 times (20–30 dB) stronger. Extraction of a signal buried so deep in noise is made possible by the knowledge of the spread spectrum ranging code. But there is little margin and attenuation (by foliage, for example) can be great enough to make a signal irretrievable.

Receiver architecture. Given a recent almanac and a rough idea of user location and time, the receiver determines which satellites are in view. Given a satellite ID, the receiver knows the structure of the C/A code transmitted by it. The receiver "tunes" to the C/A code to acquire the signal, and from then on tracks changes in it continuously.

To acquire a signal, the receiver generates a replica of the known C/A code, and attempts to align it with the incoming code by sliding the replica in time and computing the correlation between the two. The correlation function exhibits a sharp peak when the code replica is aligned with the code received from the satellite. The uncertainty in matching the replica with the incoming code is limited to only 1023 code chips, and the process of aligning them is generally quick. Direct acquisition of a P(Y) code is difficult by design due to the length of the code.

The measurement of transit time for a signal modulated by a C/A code is conceptually quite simple: The time shift required to align the receiver-generated code replica and the signal received from the satellite is the apparent transit time of the signal (modulo 1 millisecond) (Fig. 6).

The Doppler shift, caused by the relative motion of a satellite and the user, is a measure of the projection of the relative velocity on the line of sight, and can be converted into pseudorange rate. Given the pseudorange rates corresponding to four satellites, and the satellite velocity (obtained from the navigation message), a user can compute his or her velocity. A GPS receiver does this automatically, continuously, and virtually instantaneously.

Differential GPS (DGPS). The quality of the GPS position estimates can be improved by reducing

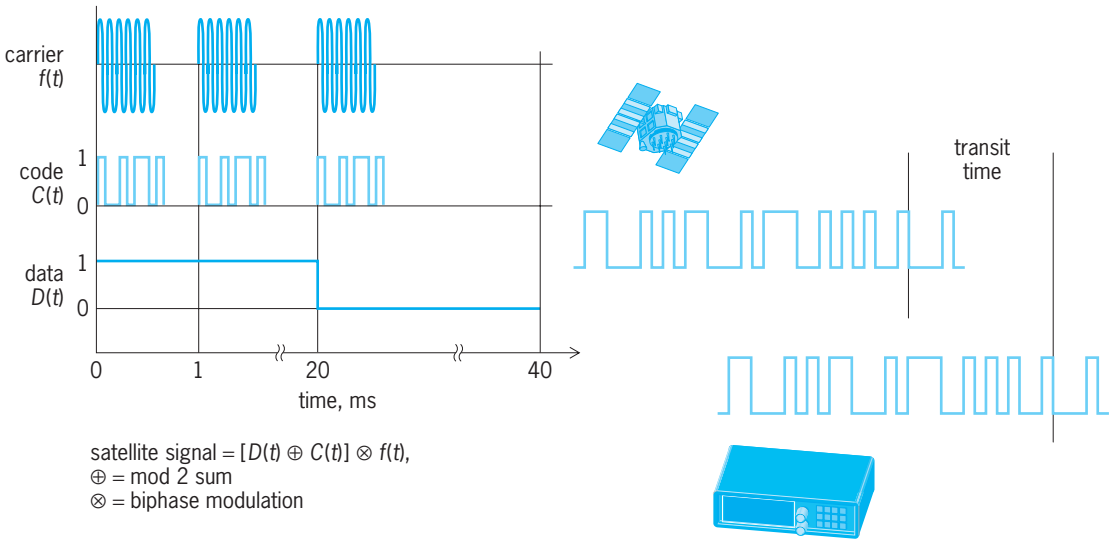


Fig. 6. Structure of the signal available for civil use and estimation of transit time from the satellite to the user.

pseudorange measurement errors. The errors associated with satellite ephemeris and clock, and atmospheric propagation delays are similar for users located near each other, and change slowly in time. These errors can be estimated from measurements with a receiver at a known location, and can be made available as differential corrections to the GPS users in the area, allowing them to mitigate errors in their measurements (**Fig. 8**).

The U.S. Coast Guard broadcasts differential corrections on marine radiobeacon frequencies (285–325 kHz) from about 50 broadcast sites in U.S. coastal areas and around the Great Lakes. This service, called Maritime DGPS, has been operational since 1999 and provides horizontal positioning accuracy of 1–3 m (3–10 ft) at a distance of 100 km (60 mi), or more, from a reference station. A number of countries have implemented systems compliant with the Maritime DGPS standards to enhance safety in their waterways.

An alternative to local-area DGPS is a centralized system which provides differential corrections usable over continent-wide areas or oceans. Such systems, referred to as wide-area DGPS, collect measurements from a network of reference stations covering the area of interest and process them to decompose the measurement errors into their constituent parts: satellite orbit, satellite clock, and ionospheric delay. The differential corrections are then transmitted in a form so that each user can determine his or her corrections on the basis of his or her (rough) position estimate.

The U.S. Wide Area Augmentation System (WAAS), operational since 2003, broadcasts corrections applicable to the United States from geostationary satellites. These corrections are coded into the navigation messages modulated on GPS-like signals on L1 which can be received with a GPS receiver with some software modifications. The Canada-Wide Differential GPS (CDGPS), which became operational in 2004, also uses geostationary satellites to broadcast corrections which are available to all who can invest in a separate radio receiver. Global Differential GPS (GDGPS), operated by the National Aeronautics and Space Administration, transmits corrections on the Internet (to authorized users). A number of commercial services also provide differential corrections via communication satellites and FM subcarrier, and

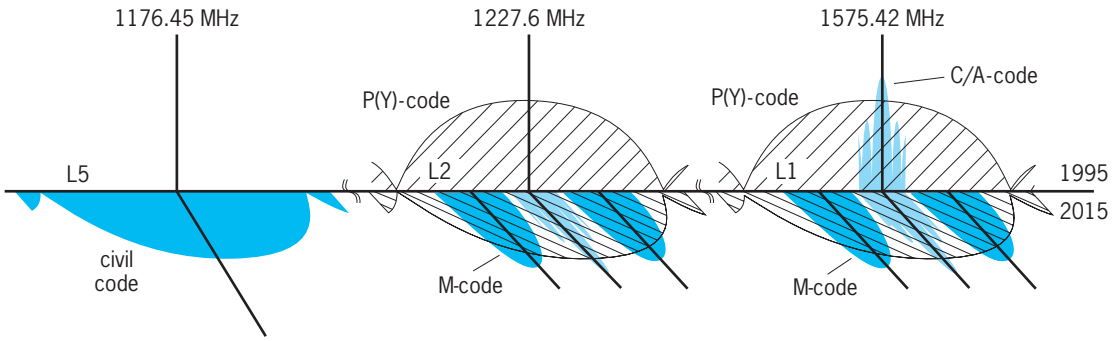


Fig. 7. Power spectral densities of the current and future GPS signals. Each satellite transmits three signals: one for civil users and two for Department of Defense–authorized users. Satellites launched starting in 2005 transmitted seven signals: three for the civil users and four for the Department of Defense-authorized users.

use of such services is common in offshore oil exploration and fleet management.

Modernization. GPS modernization was launched in the late 1990s, soon after the system became operational and its full potential for military and civil applications became clear. The civil users lobbied successfully for expanded benefits and secured two new signals: a signal on L2, and another at L5 (1176.45 MHz) to benefit civil aviation and other applications with safety-of-life considerations.

With its primary mission as a warfare support system, new measures were also deemed necessary to protect the signals for military use from interference, to prevent the use of any GPS signals by adversaries during hostilities, and to preserve peaceful civil use outside the area of military operations. The Navigation Warfare (Navwar) program of the Department of Defense addresses these requirements as a part of modernization.

There are plans to modify the design of satellites scheduled for launch during and after 2005 to include the new signals, and to design a whole new GPS of the future, called GPS III. The number of ranging signals for civil users will increase from one to three. The military will get a pair of new, Navwar-compatible signals called M-codes (Fig. 7). There will be equally important changes behind the scenes in the Control Segment to operate the modernized system. It will be 2015 or later before a full constellation of the expanded-capability satellites is deployed.

Systems under development. Besides GPS, other satellite navigation systems are under development.

GLONASS. While GPS was under development, the Soviet Union undertook to develop a similar system called GLONASS. Like GPS, GLONASS was designed primarily for the military, and a subset of its signals was offered for civil use in the late 1980s. Since the dissolution of the Soviet Union, the Russian Federation has assumed responsibility for GLONASS.

GLONASS, like GPS, fielded a 24-satellite constellation in medium earth orbits but, unlike GPS, placed the satellites in three planes with an inclination of 65°. Each satellite transmits in two frequency bands in the L-band and offers GPS-like services for the civil and military users.

The civil user community was excited at the prospect of having access to two autonomous systems as GLONASS built up to a full constellation in 1996. Unfortunately, the system has since declined as it has suffered from a lack of resources in a changed political and economic climate. Between 2000 and 2004, the Russians managed only one launch per year, each placing three new satellites in orbit, and the constellation strength has stayed mostly below 12. While a couple of GPS+GLNASS receiver models have been available since the mid-1990s, the user community and receiver manufacturers appear uncertain of the viability of GLONASS.

Galileo. A more promising GNSS is the European system, Galileo, planned as "an open, global system, fully compatible with GPS, but independent from it." Galileo was conceived as a joint public-private enterprise under civilian control, to be financed and managed by the European Commission, the European Space Agency, and industry. The deployment phase was scheduled to begin in 2006 and an operational system is planned for 2008.

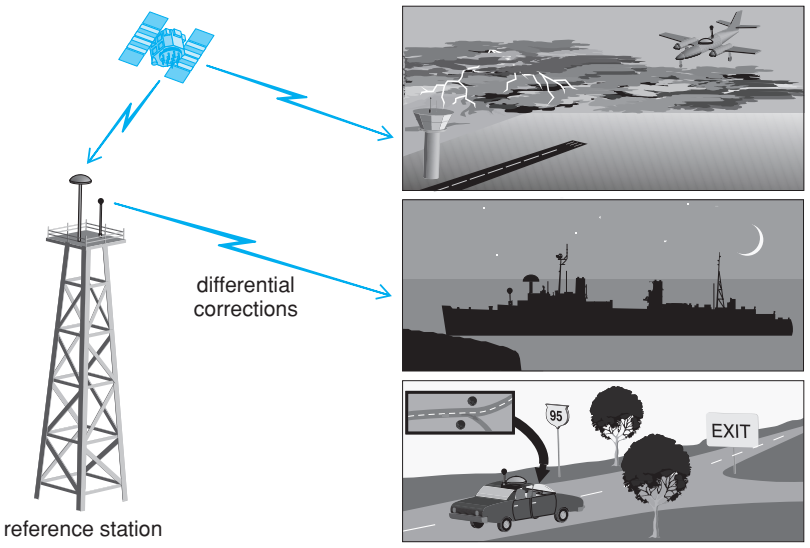


Fig. 8. Local-area differential GPS. A reference receiver at a known location measures pseudoranges, determines errors therein, and transmits corrections on a radio link to the users in the area. In this model, each user is equipped with a saucer-shaped antenna to receive the GPS signals and a bulb-shaped antenna to transmit or receive the differential corrections on a radio link. The DGPS users in the picture are (top) an aircraft in clouds preparing to land, (middle) a ship entering a harbor at night, and (bottom) a motorist on a highway in an unfamiliar city.

Galileo would be a second-generation GNSS, following the path of GPS and benefiting from technological advances since the mid-1970s when GPS was designed. Galileo would offer an Open Service without any user fees, like the GPS Standard Positioning Service, and three different, value-added services for which a user would pay.

Beidou. With the launch of two geostationary satellites in 2000 and another in 2003, China has apparently completed a regional satellite navigation system called Beidou (Big Dipper). This system provides positioning capability over China and neighboring areas. Unlike GPS, GLONASS, and Galileo, which are passive or listen-only systems, Beidou interacts with its users: The system determines the position estimate and transmits it to each user. The positioning accuracy is reported to be dozens of meters.

Compatibility and interoperability of systems. GPS, GLONASS, and Galileo are autonomous systems. From a user's perspective, at a minimum, these systems should be compatible or noninterfering: None should degrade the performance of the others in a significant way. Ideally, the systems would be interoperable, allowing a user to combine measurements from the three systems so that the performance is as least as good as the best of the three at that place and time. A user could then choose between the three independent, redundant services, or one robust integrated service.

The three systems use different time scales and coordinate frames. The offsets among the time scales can be estimated and accounted for at the user level. The differences between the coordinate frames,

however, must be characterized at system level. A user's position as determined from GPS and GLONASS was found to differ systematically by several meters because of the differences in their coordinate frames: GPS expresses the user position in WGS 84, and GLONASS in an independently realized global coordinate frame called PZ-90. A transformation between the two coordinate frames was not determined until the mid-1990s. The difference between WGS 84 and the independently realized Galileo Terrestrial Reference Frame (GTRF) is expected to be at centimeter level, insignificant for most users.

Applications. The civil applications of GPS grew rapidly in the 1990s and new applications continue to appear. The number of GPS receivers manufactured each year exceeds one million. The emergence of GLONASS and Galileo is bound to accelerate the pace of development of new applications.

While its positioning capability receives most attention, GPS is also a precise timing service. For example, before GPS, it would have been a challenge to synchronize clocks at multiple sites to conduct simultaneous observations of events (perhaps celestial) at distant locations. An inexpensive GPS receiver can now serve as a precise clock, keeping time with an accuracy of better than 0.1 microseconds relative to Coordinated Universal Time (UTC). Timing is critical in today's telecommunication systems, and GPS is being used increasingly to synchronize the elements of networks. *See* ATOMIC TIME; TIME.

The applications of GPS for navigation and fleet tracking in land transportation, aviation, and maritime commerce continue to grow. The convergence of GPS and personal computer technology has made it possible to collect vast amounts of positional data and organize them into geographic information systems (GIS), showing, for example, the location and status of every utility pole or access-hole cover in a town, or a map showing concentrations of toxic or radioactive wastes at a site. The centimeter-level positioning capability of GPS has revolutionized surveying and mapping, and has created novel applications in construction, mining, agriculture, and the earth sciences. *See* GEOGRAPHIC INFORMATION SYSTEMS.

Millions of bikers, runners, hikers, sailors, and fishermen have come to rely on GPS for their position and velocity. GPS receivers are now routine in cell phones and personal digital assistants. Indoor GPS positioning in conjunction with a cell phone, now under development, would lead to a whole new class of applications.

One's position expressed in a well-defined coordinate frame is a valuable piece of information. Determining position indoor and outside would soon become as easy as determining precise time is today, and position would come to occupy the same important place in daily life as time does today.

Acknowledgment. Figures 1–8 are reproduced by permission of Ganga-Jamuna Press, Lincoln, MA.

Pratap Misra

Bibliography. T. Logsdon, *Understanding the Navstar: GPS, GIS, and IVHS*, 2d ed., Van Nostrand Reinhold, New York, 1995; P. Misra and P. Enge *Global Positioning System: Signals, Measurements, and Performance*, 2d ed., Ganga-Jamuna Press, Lincoln, Mass., 2005; B. W. Parkinson et al. (eds.), *Global Positioning System: Theory and Applications*, vols. I and II, American Institute of Aeronautics and Astronautics, Washington, DC, 1996; D. Wells (ed.), *Guide to GPS Positioning*, Canadian GPS Associates, 1986.

Satellite radio

Direct audio broadcasting from satellites to mobile and stationary receivers. This concept originated in the late 1980s as an outgrowth of direct television broadcasting from satellites to home receivers. There are currently two satellite radio systems (also known as satellite digital audio radio service) operating in the continental United States: one is providing service to users in Japan and Korea, and another is in advanced planning to serve Europe. *See* SATELLITE TELEVISION BROADCASTING.

The useful range of frequencies for radio broadcast transmission from satellites is 1–3 GHz, the most sought-after frequency band by all transmission services due to its superior propagation characteristics. Above 3 GHz, propagation losses and rain (precipitation) attenuation pose difficulties, and below 1 GHz, satellite antenna size and galactic noise (radio-frequency noise that originates outside the solar system) pose difficulties. The Federal Communications Commission (FCC) has allocated a frequency band of 2320– 2345 MHz for satellite radio, which has been coordinated internationally. This frequency band cannot be used in the United States for any other service because the sensitivity, mobility, and antenna omnidirectionality of the ground receivers require an interference-free environment. Of the available 25 MHz, a satellite radio system requires between 10 and 15 MHz bandwidth to support a reasonable number of audio programs. As a result, the FCC auctioned two 12.5-MHz-wide bands for satellite radio service, and granted licenses for two such systems in October 1997 to Sirius Satellite Radio Inc. and to XM Satellite Radio, Inc. *See* RADIO SPECTRUM ALLOCATIONS; RADIO-WAVE PROPAGATION.

Technology. The technical challenge of satellite radio is in keeping service outages very infrequent and short. Service outages are caused by blockages, such as a building, structure, or mountain that obstructs the line-of-sight between a satellite and receiving antenna; by multipath, where received reflected signal components have sufficient amplitude and phase offset to corrupt (distort) the unreflected received signal; by foliage attenuation; and by interference, particularly the out-of-band emissions from transmitters in adjacent frequency bands.

The quality goal is to achieve continuity of service (that is, no outages) better than 99.9% of the time, excluding the time when a vehicle is indoors since the satellite transmissions do not substantially penetrate building walls. High service availability is accomplished by using high-power transmission satellites and by providing an unprecedented degree of transmission diversity in the system design.

Spatial diversity. This is accomplished by simultaneously broadcasting the same transmission from two satellites located in space so that their azimuth and elevation angles from the radio receiver are significantly different. Since the radio is designed to receive both transmissions, if the path from the radio to one of the satellites is blocked by a building or mountain, there will be no outage if the path to the other satellite is unobstructed.

Time diversity. This is accomplished by providing a 4–5-s offset either between transmissions from two satellites by a time delay or within each satellite transmission by a long interleaving span. If one or both transmissions are lost, the receiver continues to play using the portion of the transmission stored in a 4–5-s buffer memory, and no outage occurs if a transmission path is restored before the end of the stored signal. This technique is particularly effective in combating outages from roadway overpasses.

Frequency diversity. In this technique, two satellites transmit at different radio frequencies, several megahertz apart. Frequency diversity is effective in minimizing a form of multipath called frequency selective fading.

Modal diversity. Despite the above diversity techniques, it is impossible to provide high-quality service in dense urban areas since the blockage from buildings is overwhelming. However, this can be overcome by providing terrestrial repeaters. High-power broadcast transmitters located in dense urban areas are fed the same signal as transmitted from the satellite and then retransmit this signal in a frequency band between the two satellites' transmissions. A receiver with three channels, one for the terrestrial repeater and two for the diversity satellites, allows virtually uninterupted service and seamless transition between satellite and terrestrial transmissions. The three signals, which are simultaneously received when available, are added together in a maximal ratio (diversity) combiner.

In addition to diversity, satellite elevation angle and the use of equalizers are important techniques for attaining high service availability.

Elevation angle. The elevation angle from the satellite to the radio is extremely important in mitigating blockage, multipath, and foliage attenuation. For geostationary orbits, the elevation angle is reduced as the location of the radio moves further north or south of the Equator. An alternative is to use a constellation of satellites in highly inclined, elliptical geosynchronous orbits so that high elevation angles can be provided to radios located at latitudes above 30°N (**Fig. 1**). In the Sirius geosynchronous satellite constellation, three satellites follow each other at 8-h intervals and spend approximately 16 h north of the Equator and 8 h below. *See* COMMUNICATIONS SATELLITE.

The importance of elevation angle with regard to foliage attenuation is shown in **Fig. 2**. The difference between a 30° and a 60° elevation angle for a 99% service availability is over 13 dB, which means a satellite must transmit more than 20 times additional power at 30° than at 60° for equal service availability through foliage. **Figure 3** shows the importance

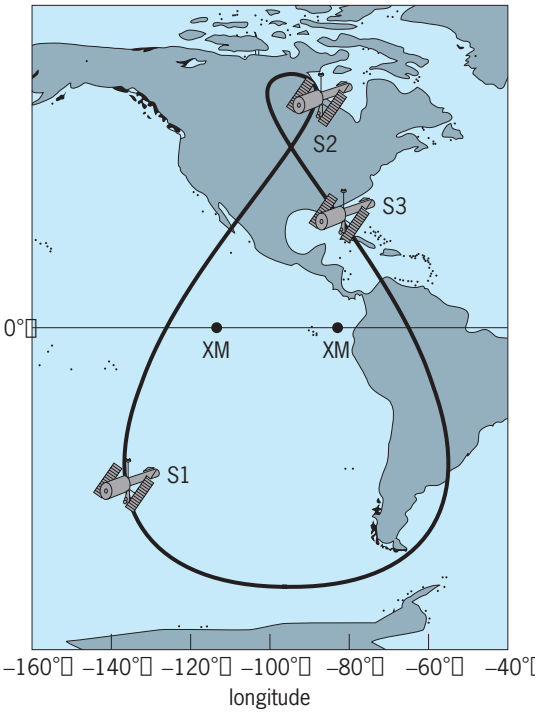


Fig. 1. Orbital locations of the XM geostationary satellites located at longitudes 85°W and 115°W, and the satellite ground track of the Sirius (S) geosynchronous satellite constellation.

of higher elevation angles for blockage reduction, where the blockage distance with a 60° elevation angle is only one-third of that with a 30° elevation angle. A high elevation angle also alleviates blockage outages from passing trucks.

Equalization. A technique for reducing outage from multipath is to have an equalizer in the receiver that compensates for a significant portion of the signal distortion in amplitude and phase caused by the multipath components. This is particularly effective in satellite radio for handling short multipath propagations that are caused, for example, by reflections from large, overhead roadway signs. *See* EQUALIZER.

System design. For satellite radio, each satellite system generally provides 65 music channels and 55 voice channels. This mix can be changed and, if

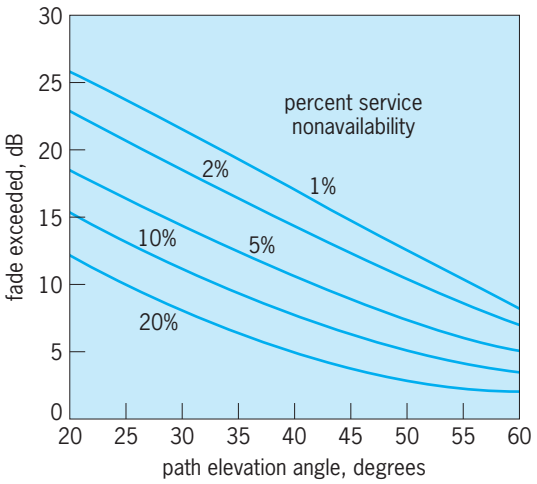


Fig. 2. Foliage attenuation.

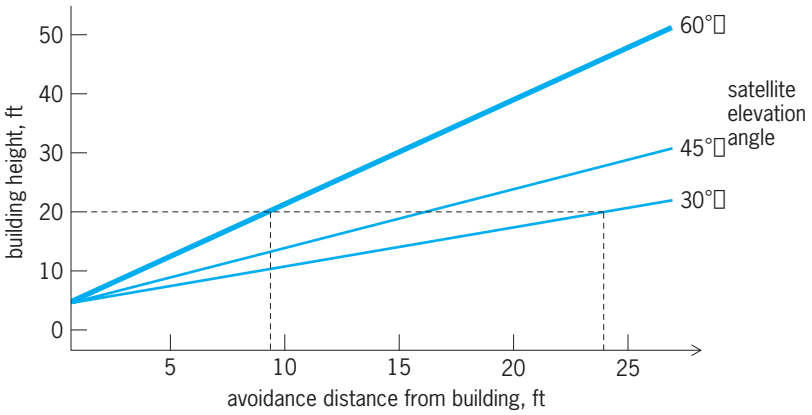


Fig. 3. Blockage avoidance. A vehicle (with antenna height 5 ft or 1.5 m) must be a minimal distance from a building to avoid blockage in the worse case. 1 ft = 0.3 m.

a market develops, one-way data channels can be substituted for one or more audio channels. The major elements of the system are the satellites, radios, and the broadcast studio/uplink (**Fig. 4**). The telemetry, tracking, and command (TT&C) subsystems that take care of the satellites are standard and use no new technology. Also, the terrestrial repeater network uses standard digital audio broadcasting (DAB) techniques, which include coded or orthogonal frequency division modulation (COFDM). *See* MODULATION.

Satellites. The satellites used are modifications of existing high-power communications satellites. The important technical modifications are conversion to 2.3-GHz transmitter operation, and concentration of the transmitter power to achieve a few megawatts (63–65 dBW) of effective isotropic radiated power (EIRP) over the relatively narrow transmission bandwidth (2.5–4.2 MHz). The high transmitter power is achieved by paralleling many traveling-wave-tube amplifiers.

Radios. Radios have been produced for aftermarket installation in motor vehicles, boats, planes, and homes, as well as for installation during the assembly of motor vehicles. The receiver is a three-channel type with intelligent combining of the two satellite and terrestrial transmissions, when available. An application-specific integrated circuit (ASIC) chipset contains most of the required electronic circuits, excluding the antenna (omnidirectional with 3–5 dBi gain), radio-frequency amplifiers (typical system noise temperature of 160 K or −172°F), bulk memories, and power supply. The received digital signal is encrypted at the broadcast studio to prevent theft of service. In addition to having a decryptor in the receiver, each receiver has an embedded unique digital address. This allows the broadcast studio, through a service channel, to send messages and commands, which are recognized solely by the addressed receiver. Among other functions, this is

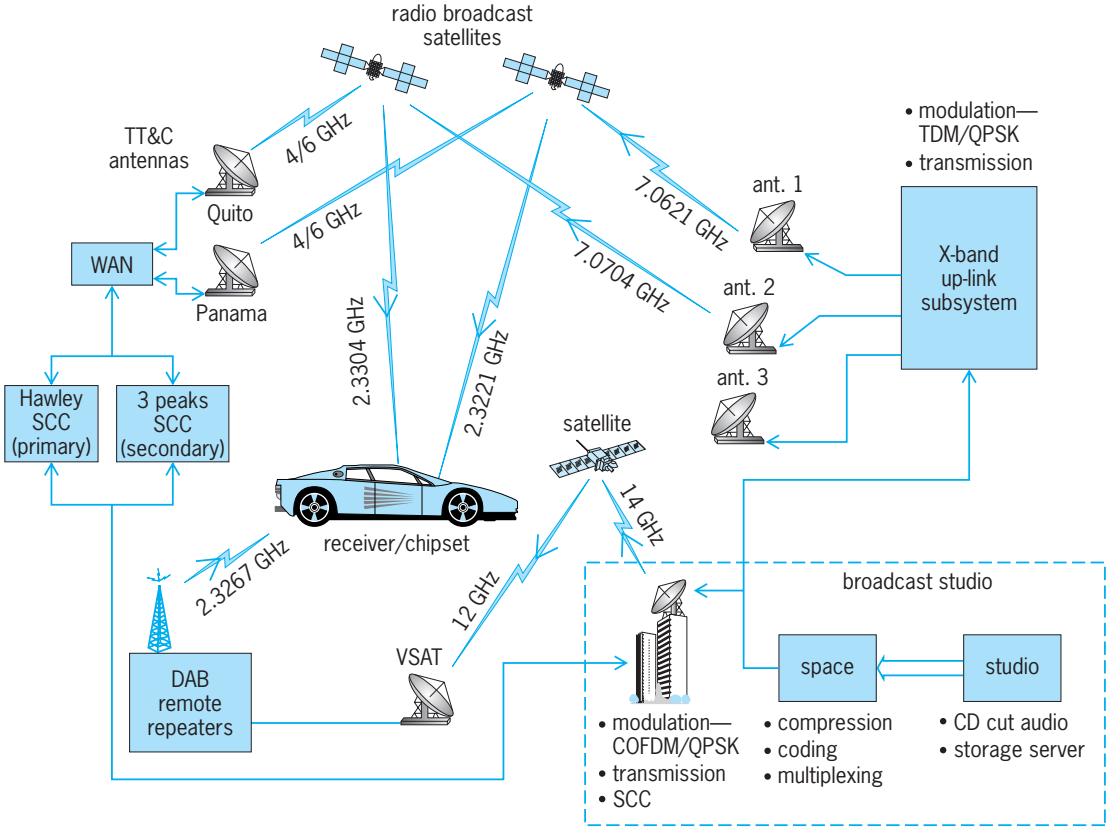


Fig. 4. Typical system for providing satellite radio service. TT&C = telemetry, tracking, and command; DAB = digital audio broadcasting; VSAT = very small aperture terminal; SCC = satellite control center; COFDM = coded othogonal frequency division modulation; QPSK = quadrature phase shift keying; TDM = time division modulation.

used to turn subscriber service on and off. *See* INTEGRATED CIRCUITS; RADIO RECEIVER.

Broadcast studio/uplink. These facilities include those used for generating the audio programming, as well as the resulting program signals' multiplexing, coding, modulation, and transmission (Fig. 4). The programming facilities are audio studios, using both recorded and live performances, which output in a digital form as a music or a voice channel. The channels are compressed, coded (the inner code is convolutional, the outer code is a Reed Solomon block code), multiplexed together using time division into one digital stream, modulated using quadrature phase shift keying, split into two identical signals, one of which is then time-delayed, translated to 7 GHz, and transmitted to the two active satellites. Other than the audio compression, the facilities and techniques are those used by newer communications satellite, terrestrial broadcast, and satellite television systems.

Audio compression. An audio compression technique, called perceptual audio coding (PAC), is the first extensive use of a technology developed over the past decade without which satellite radio would not be practical or economic. It allows compact disk recordings (whose digital output is 1.2 mb/s) to be compressed to approximately 50 kb/s at the transmit studio and expanded in the radio receiver in a manner that the listener hears a music quality close to the uncompressed music. This is accomplished by compression/expansion algorithms which take advantage of the characteristics of the human ear (for example, its inability to hear certain sounds and artifacts). Similarly, a voice channel is compressed to approximately 25 kb/s. *See* DATA COMPRESSION.

Outlook. Satellite radio is fast developing with regard to commercial success, which requires a few million subscribers for each system. During mid-2005, the two United States satellite radio systems had 7 million subscribers combined. New services and uses will unquestionably develop over the next several years. The possibilities of providing data, multimedia, and personalized services are under investigation. Navigation services were scheduled to be introduced in 2005 and television to the rear seats of sports utility vehicles in 2006. The Canadian government authorized (with some conditions on programming content) the extension of the Sirius and XM satellite radio service to subscribers in that country during 2005. *See* RADIO; RADIO BROADCASTING.

Robert D. Briskman

Bibliography. R. D. Briskman and R. J. Prevaux, S-DARS broadcast from inclined, elliptical orbits, *52d International Astronautical Congress*, IAF-01M.5.05, October 1–15, 2001; R. D. Briskman and S. P. Sharma, DARS satellite constellation performance, *AIAA 20th International Communications Satellite Systems Conference*, AIAA-2002-1967, May 12–15, 2002; Efficient high latitude service arc satellite mobile broadcasting systems, U.S. Patent and Trademark Office, 6,223,019, April 24, 2001; R. A. Michalski, An overview of the XM satellite radio system, *AIAA 20th International Communications Satellite Systems Conference*, AIAA-2002-1844, May 12–15, 2002; Mobile radio receivers using time diversity to avoid service outages in multichannel broadcast systems, U.S. Patent and Trademark Office, 5,592,471, January 7, 1997; Radio frequency broadcasting systems and methods using two low-cost geosynchronous satellites, U.S. Patent and Trademark Office, 5,319,673, June 7, 1994.

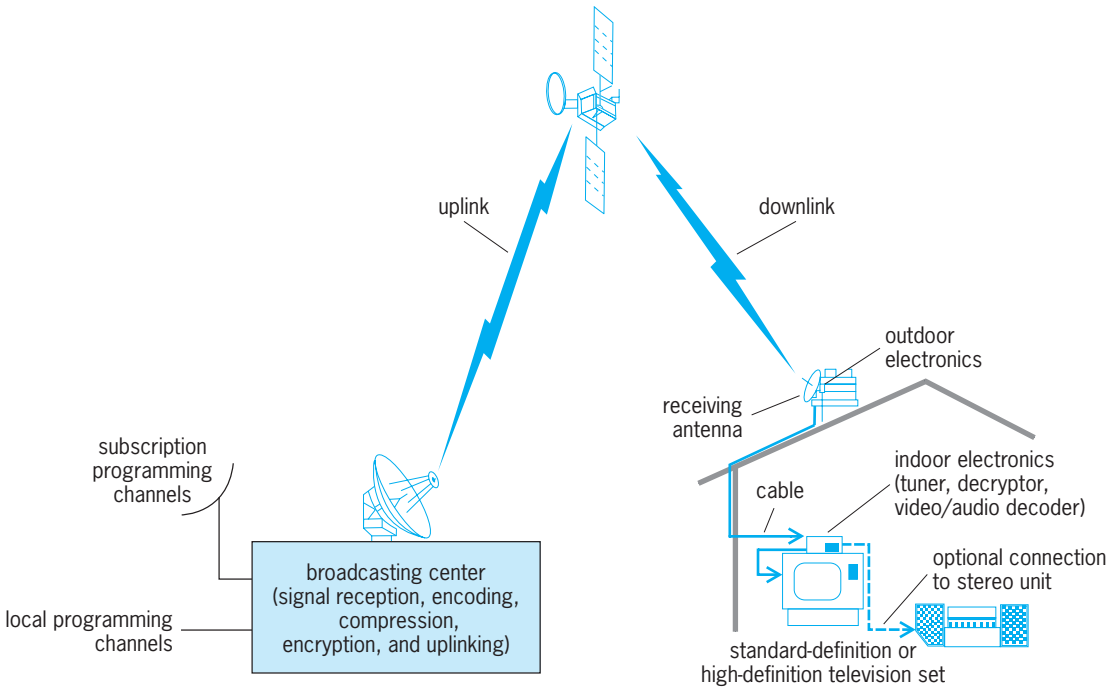
Satellite television broadcasting

The transmission of television and other program material via satellite directly to individual homes and businesses. Direct broadcasting satellite (DBS) systems, or direct-to-home (DTH) satellite systems, are operational in many nations and regions around the world. The highly successful "small dish" DTH systems are characterized by all-digital transmission with the links to the customer dishes at frequencies above 10 GHz using radio spectrum sometimes referred to as Ku-band. Two major small-dish systems in the United States, DIRECTV and Dish, have been operational since the mid-1990s and serve over 20% of all American households.

Direct television broadcasting is the most widely used consumer application for communications satellites. The related application, direct audio broadcasting, has also shown significant growth since its launch in the United States in 2001. Other "direct" consumer applications include voice, data, and Internet access, but these services have not yet achieved mass market success. *See* SATELLITE RADIO.

Technology. DBS systems use a satellite in geostationary orbit to receive television signals sent up from a broadcasting center, amplify them, and transmit them down to the customer receiving antennas called dishes. The satellite also shifts the signal frequency, so that, for example, a signal sent up to the satellite in the 17.3–17.8-GHz uplink band is transmitted back down in the 12.2–12.7-GHz downlink band. The downlink signal is picked up by a receive antenna located atop an individual home or office; these antennas are usually in the form of a small parabolic dish. The receive antenna is permanently pointed at the satellite, which is at a fixed point in the sky, in an equatorial geosynchronous or geostationary orbit, a circular orbit in the plane of the equator with a period synchronized to the 24-h rotational period of the Earth. *See* ANTENNA (ELECTROMAGNETISM).

Receiver technology. The electronics for handling microwave frequencies is expensive, so the signal from the dish antenna is first passed to a downconverter, mounted outdoors on the antenna, which shifts it to (typically) the 0.95–1.45-GHz band. This signal is then conducted by cable to the receiver near the television set. The receiver contains the channel selector or tuner, a demodulator, and a decryptor (descrambler) as well as a video-audio decoder to permit the user to view authorized channels. With the broadcasting industry's evolution to high-definition television (HDTV), the more advanced indoor



Direct broadcasting television satellite system.

receivers can now output either standard-definition or high-definition television. Additionally, the receivers have the capability to output monaural, stereo, or multichannel audio to a television set, home audio system, or home theater system (see **illustration**). *See* TELEVISION RECEIVER.

Spectrum allocations. Television broadcasting by satellite may use radio-frequency spectrum licensed as either the broadcasting satellite service (BSS) or the fixed satellite service (FSS). In most cases the system design differences due to the use of these alternative bands are due to international and domestic regulation and not technical or environmental differences. The two major U.S. DBS systems have used BSS spectrum with a downlink frequency of 12.2–12.7 GHz. Other major systems in the Western Hemisphere use FSS spectrum with a downlink frequency around 11 GHz. The term DBS is generally used to denote systems using the BSS band. The term DTH is more general and applies to television delivery systems employing any band. *See* RADIO SPECTRUM ALLOCATION.

Satellite technology. A direct broadcasting satellite typically contains 32 or more transponders, each with a radio-frequency power output in the range 120–240 W. Each transponder acts as a separate amplifier, frequency translator, and transmitter chain for a signal uplinked from the broadcasting facility. The transmitter of each transponder drives one of the satellite's downlink antennas. Each downlink antenna creates either a broad national beam or spot beams. In the major U.S. DBS systems, each service provider uses multiple satellites, with a given satellite broadcasting on a national beam or on one or more "local market" beams. Each local spot beam delivers television signals from the very high frequency (VHF) and ultrahigh frequency (UHF) stations of a particular urban area back into that geographic area.

As each satellite orbits the Earth, its communications antennas are maintained precisely pointed toward the target service areas. Large panels of photovoltaic cells are continuously oriented toward the Sun to provide power for the signal transmitters. Each satellite also carries rechargeable batteries for continuous operation during occasional "eclipse" periods when the satellite is in the Earth's shadow.

Digital signal technology. All recently deployed DBS and DTH systems use digital signals. Typically in these systems, a single 24-MHz-wide satellite transponder can carry an error-corrected digital signal of 30 megabits per second or greater. A wide variety of communications services can be converted to digital form and carried as part of this digital signal, including conventional television, high-definition television, multichannel audio, information services (such as financial and news retrieval services), and other forms of digital data.

Modern digital signal compression technology greatly increases the capacity of a satellite transponder and hence the entire system. It is possible to compress up to perhaps 10 or more standard-definition television signals into the bandwidth of a single DBS transponder. The technology used for this signal compression, from the Motion Picture Experts Group (MPEG) standards body, is very similar to that used in production of DVD optical disks. In a satellite transponder the bandwidth is fixed, but multiple signals can be "statistically multiplexed" into this bandwidth in real time. In a DVD the total storage capacity is fixed, but the processing of the movie can

be modified to assure that the entire capacity is used, but not exceeded, to achieve the target playback quality. *See* COMPACT DISK; DATA COMPRESSION; OPTICAL RECORDING.

Signal attenuation by rain. DBS systems, like all satellite systems operating above 10 GHz, are subject to attenuation of their signals by rain. Rainfall absorbs some of the signal power transmitted by the satellite. For a small percentage of the time, the rainfall may be so heavy in a particular geographical area that the remaining power is not sufficient for reception of the signal. The amount of rain-induced attenuation can be predicted, and the system is usually designed to work for the highest percentage of the time that is economically practical. The combination of satellite power and receive-dish antenna size is chosen to enable reception for all but the heaviest rainfall periods of the year at any particular location. The DBS customer can further reduce this expected outage period by purchasing a slightly larger dish antenna.

Uplink centers. Each DBS digital uplink center accepts multiple video and audio channels, processes and compresses the signals, and transmits them to the digital broadcasting satellites for relay to individual home-receive systems. Typically the subscription programming channels arrive at the broadcasting center via satellite systems provided by other firms and intended for national channel distribution to satellite and cable companies. The local channels are typically delivered to an uplink center using dedicated digital terrestrial facilities. Each incoming channel is converted to a common digital form and is then compressed to reduce the number of bits per second required for its transmission. These compressed channels are then individually encrypted to prevent unauthorized reception and combined into composite digital bit streams by digital multiplexers. Finally, each bit stream is protected by forward error control (FEC) coding and modulated, amplified, and transmitted to the satellite via a large uplink antenna.

Comparison with other satellite systems. Other satellite systems are used to provide "big dish" DTH services, notably communications satellites operating in the 3.7–4.2-GHz band (the C band). These satellites are principally designed to offer general telecommunications services, including the transmission of television programming to cable television companies and television broadcast stations. These satellites operate at lower power levels, and thus much larger antennas, typically 6–10 ft (1.8–3 m) in diameter, are needed to receive the signals. Direct-to-home service was initiated during the 1980s via these systems using analog transmission, but has since been largely supplanted by small dish systems.

Comparison with other technologies. DBS systems have expanded very rapidly in the United States and elsewhere because of their superior quality, expansive programming lineup, and centralized, efficient technical and administrative facilities. DBS technical quality is due to its all-digital delivery and the use of microwave frequencies and highly directive antennas. Cable companies are now achieving wide deployment of their own all-digital solutions and competing favorably with direct broadcast. Telephone companies have announced new video delivery networks using fiber-optic cable or Digital Subscriber Line (DSL) transmission over telephone wiring. All these competing systems use similar underlying technology for television signal processing and for the consumer's set top box. The primary network differences are in the means of transmission to the home. *See* CABLE TELEVISION SYSTEM; COMMUNICATIONS SATELLITE; OPTICAL COMMUNICATIONS; TELEVISION.

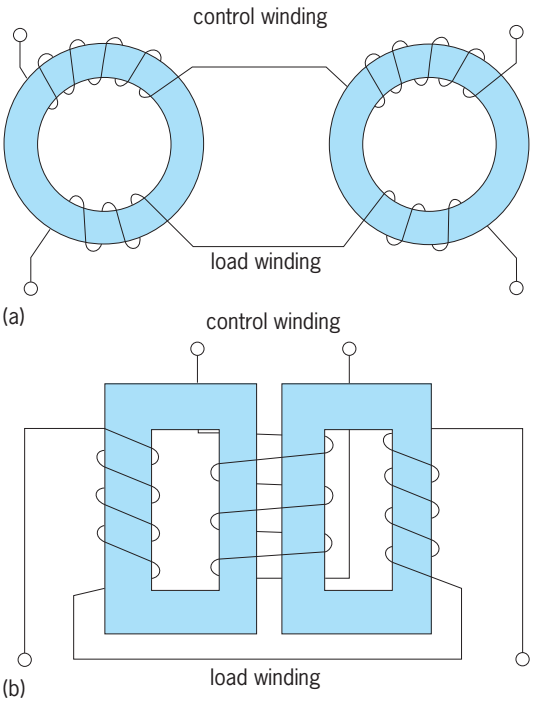
John P. Godwin

Bibliography. T. T. Ha, *Digital Satellite Communications*, 2d ed., McGraw-Hill, 1990; G. Maral and M. Bousquet, *Satellite Communication Systems*, 4th ed., Wiley, 2002; M. Robin and M. Poulin, *Digital Television Fundamentals*, 2nd ed., McGraw-Hill, 2000; B. Sklar, *Digital Communications: Fundamentals and Applications*, 2d ed., Prentice-Hall, 2001.

Saturable reactor

An iron-core inductor in which the effective inductance is changed by varying the permeability of the core. Saturable-core reactors are used to control large alternating currents where rheostats are impractical. Theater light dimmers often employ saturable reactors.

In the **illustration** of two types of saturable-core reactors, illus. *a* shows two separate cores, while in illus. *b* a three-legged core is formed by placing two two-legged cores together. The load winding, connected in series with the load, carries the alternating



Typical construction of saturable-core reactors. (*a*) Two separate cores. (*b*) Three-legged cores.

current and acts as an inductive element. The control winding carries a direct current of adjustable magnitude, which can saturate the magnetic core. When the core is saturated by the direct current, the effective inductance (and therefore the reactance) of the coils in the ac circuit is small. Little voltage can therefore be induced in the coil to oppose the applied voltage and therefore reduce the voltage that is supplied to the load.

Reducing the magnitude of the control current reduces the intensity of saturation. This increases the reactance of the load winding. As the reactance increases, the voltage drop in the load winding increases and causes a reduction in the magnitude of the voltage applied to the load. *See* INDUCTANCE; MAGNETIZATION. Wilson S. Pritchett

Bibliography. T. Croft, J. Watt, and W. I. Summers, *American Electricians' Handbook*, 13th ed., 1996; D. G. Fink and H. W. Beaty (eds.), *Standard Handbook for Electrical Engineers*, 14th ed., 2000.

Saturation

The condition in which, after a sufficient increase in a causal force, further increase in the force produces no additional increase in the resultant effect. Many natural phenomena display saturation. For example, after a magnetizing force becomes sufficiently strong, further increase in the force produces no additional magnetization in a magnetic circuit; all the magnetic domains have been aligned, and the magnetic material is saturated.

After a sponge has absorbed all the liquid it can hold, it is saturated. In thermionic vacuum tubes thermal saturation is reached when further increase in cathode temperature produces no (or negligible) increase in cathode current: anode saturation is reached when further increase in plate voltage produces substantially no increase in anode current. *See* DISTORTION (ELECTRONIC CIRCUITS); VACUUM TUBE.

In an *npn* transistor when the collector voltage becomes more negative than the base voltage, the base-to-collector diode becomes forward-biased: the collector then emits and the transistor is saturated. Current flow is controlled by a stored charge of carriers in the base region. *See* SEMICONDUCTOR; TRANSISTOR.

In colorimetry the purer a color is, the higher its saturation. Radiation from a color of low saturation contains frequencies throughout much of the visible spectrum. *See* COLORIMETRY.

In an induced nuclear reaction, saturation exists when the decay rate of a given radionuclide is equal to its rate of production. In an ionization chamber, saturation exists when the applied voltage is high enough to collect all the ions formed by radiation but not high enough to produce ionization by collision. When addition of the dissolved species produces no further increase in the concentration of a solution at a certain temperature and pressure, the solution is said to be saturated. *See* SOLUTION.

Frank H. Rockett

Saturation current

A term having a variety of specific applications but generally meaning the maximum current which can be obtained under certain conditions.

In a simple two-element vacuum tube, it refers to either the space-charge-limited current on one hand or the temperature-limited current on the other. In the first case, further increase in filament temperature produces no significant increase in anode current, whereas in the latter a further increase in voltage produces only a relatively small increase in current. *See* VACUUM TUBE.

In a gaseous-discharge device, the saturation current is the maximum current which can be obtained for a given mode of discharge. Attempts to increase the current result in a different type of discharge. Such a case would be the transition from a glow discharge to an arc discharge. *See* ELECTRICAL CONDUCTION IN GASES.

A third case is that of a semiconductor. Here again the saturation current is that maximum current which just precedes a change in conduction mode. *See* SEMICONDUCTOR. Glenn H. Miller

Saturn

The second largest planet in the solar system and the sixth in order of distance to the Sun. The outermost planet known prior to 1781, Saturn is surrounded by a beautiful system of rings, more than 30 moons, and a complex magnetosphere. Despite the planet's huge size, its mean density is so low it could float in water. Saturn is also the only planet that has a satellite (Titan) with a dense atmosphere. This distant planetary system has been visited by four spacecraft: *Pioneer 11* in 1979, *Voyagers 1* and *2* in 1980–1981, and *Cassini-Huygens* since 2004. The first three were flybys; the last is a joint United States–European mission consisting of a Saturn orbiter and a probe into the atmosphere of Titan.

Orbit and physical elements. The orbit of Saturn has a semimajor axis or mean distance to the Sun of 8.95×10^8 mi (1.43×10^9 km), an eccentricity of 0.056, and its plane is inclined to the plane of the ecliptic at an angle of 2.5°. With a mean orbital velocity of 6 mi/s (9.67 km/s), Saturn makes one revolution about the Sun in 29.42 years. *See* PLANET.

The equatorial diameter of Saturn is about 75,000 mi (120,540 km), and the polar diameter about 67,600 mi (108,700 km). The volume is 769 (Earth = 1) with a few percent uncertainty. The polar flattening caused by the rapid rotation is the largest of all the planets, and the ellipticity, $(r_e - r_p)/r_e = 0.098$, is 30 times the value for Earth (r_e is the equatorial radius and r_p is the polar radius).

The mass, about 95.2 (Earth = 1) or 1/3500 (Sun = 1), is accurately determined from the motions of the planet's brighter satellites. The mean density is 0.70 g/cm^3, the lowest mean density of all the planets. The corresponding value of the mean gravity

at the visible surface (superimposed cloud layers) is 1.15 (Earth = 1) or 11.3 m/s^2.

Photometric properties. The apparent visual magnitude of Saturn at mean opposition is +0.7 when the ring is seen edgewise, and the corresponding value of the reflectivity (geometric albedo) is 0.47. *See* ALBEDO.

Appearance. Observed through a telescope, Saturn appears as an elliptical disk, darkened near the limb and crossed by a series of barely discernible bands parallel to the equator. Frequently only the bright equatorial zone and the two darker tropical bands on either side of it are visible (**Fig. 1**). The rings may be seen even with a relatively small telescope, but their visibility changes with the position of the planet in its orbit, because the rotation axis (of both the planet and the rings) is inclined 27° to the perpendicular to the orbital plane. Thus Saturn has seasons similar to those on Earth, but each season lasts 7.32 years (**Fig. 2**).

The rotation period of Saturn's interior is 10^h40^m, as determined by radio emissions controlled by the planet's magnetic field. The visible cloud layers of Saturn are much more homogeneous in appearance than those of Jupiter. There is no feature comparable to the Great Red Spot, and the contrast of the features (huge storm systems in the cloud decks) that are visible is very low. The contrast may be enhanced by the use of special filters. By studying the movement of these storms it has been possible to determine that the circulation patterns on Saturn are also very different from those on Jupiter. At a latitude of ±40°, the atmosphere rotates with the same velocity as the interior, but wind velocities increase smoothly toward the equator, where, at the time of the *Voyager* observations (1980–1981), they reached a value of 1100 mi/h (500 m/s), about four times faster than Jupiter's equatorial jet. Twenty-five years later, the *Cassini Orbiter* found the wind speed at the equator to be considerably reduced, for reasons that are not yet clear. There is no alternation of easterly and westerly currents corresponding to transitions between belts and zones, as there is on Jupiter, except at latitudes above ±40°. This difference between Saturn and Jupiter represents a fundamental difference in global circulation, perhaps related to the relative sizes of the cores of the two planets. *See* JUPITER.

Atmosphere. The optical spectrum of Saturn is characterized by strong absorption bands of methane (CH_4) and by much weaker bands of ammonia (NH_3). Absorption lines of molecular hydrogen (H_2) have also been detected. The estimated quantities of these gases that are present above the clouds are equivalent to STP path lengths of about 30 mi or 50 km (H_2), 200 ft or 60 m (CH_4), and 6 ft or 2 m (NH_3). Here STP refers to standard temperature (273 K or 32°F) and pressure (1 atm or 1.01325×10^5 pascals). In these units, the STP path length from sea level through the Earth's atmosphere would be 5 mi (8 km). The presence of about 5 mi (8 km) of helium (He) has been deduced indirectly from infrared observations of pressure-broadened hydrogen emission lines.

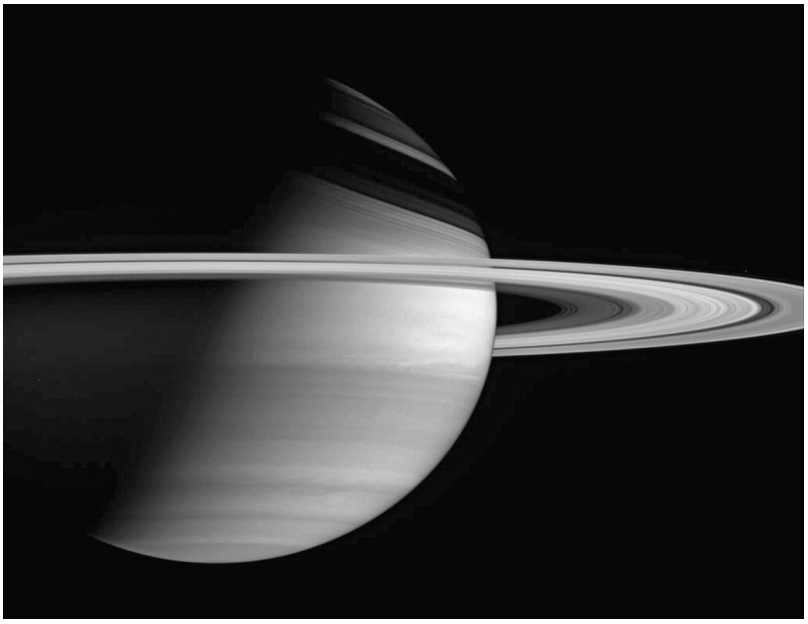


Fig. 1. Saturn viewed from the *Cassini Orbiter*. The soft velvety appearance of the low-contrast banded structure is due to scattering by the haze layer above the planet's cloud deck. The image has been rotated so that north is up. The Sun illuminates Saturn from below and Saturn's tilt throws shadows of the rings onto the northern hemisphere during the current season. The image was taken on January 23, 2005 at a distance of 1.7×10^6 mi (2.8×10^6 km) from Saturn. (*NASA/JPL/Space Science Institute*)

The temperature the planet should assume in response to solar heating is calculated to be about 76 K (−323°F), somewhat lower than the measured value of 92 K (−294°F). This suggests that Saturn has an internal heat source of roughly the same magnitude as that on Jupiter. Relatively intense infrared emission lines near a wavelength of 12 micrometers have been identified, resulting from ethane (C_2H_6) and acetylene (C_2H_2) formed in the upper atmosphere from the dissociation of methane. As in the case of Jupiter, a thermal inversion exists in this region of Saturn's atmosphere. Emission bands of methane and phosphine have also been detected here, as well as bands of water (H_2O) and carbon dioxide (CO_2), thought to arise from in-falling material. This region is well above the main cloud layer, which should consist primarily of frozen ammonia crystals, with an admixture of some other substances such as sulfur to provide the yellowish color sometimes observed in the equatorial zone. The ammonia cirrus on Saturn is apparently denser and more ubiquitous than on Jupiter, since one does not see

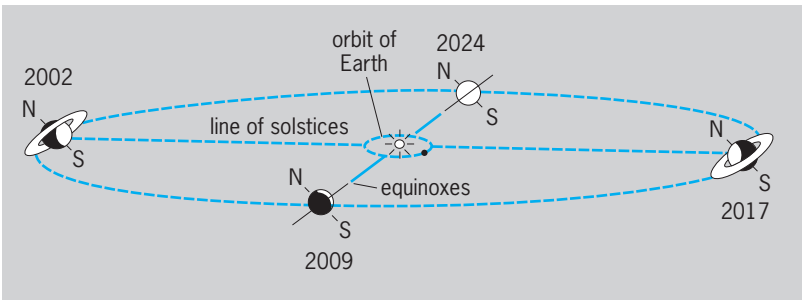


Fig. 2. Presentations of Saturn's ring system.

through it to lower cloud layers (Fig. 1). This difference probably results from the lower atmospheric temperature and smaller gravity of Saturn, which will act together to spread out and increase the depth of the cloud layer.

Internal structure and radiation belts. Observations of Saturn at radio frequencies indicate that the temperature steadily increases with depth into the atmosphere. Theoretical models for the internal structure of Saturn are similar to those for Jupiter, that is, a dense core surrounded by hydrogen compressed to a liquid metallic state that gradually merges into an extremely deep atmosphere. As on Jupiter, the gradual precipitation of helium from solution in the metallic hydrogen surrounding the core liberates heat from Saturn's interior and produces the subsolar abundance of helium (relative to hydrogen) found in Saturn's atmosphere. The existence of a magnetic field and belts of trapped electrons was initially deduced from observations of nonthermal radiation at dekameter wavelengths and was mapped out in detail by in situ measurements from instruments on the spacecraft that have visited the planet. The *Cassini Orbiter* is continuing these investigations. Saturn's magnetic field has the same polarity as Jupiter's; a terrestrial compass taken to either planet would point south instead of north. The magnetic moment of Saturn is 4.3×10^{28} gauss $\cdot$ cm^3 (4.3×10^{18} teslas $\cdot$ m^3), 500 times Earth's and 34 times smaller than Jupiter's. But the magnetic field at the cloud tops near the equator is only 0.2 gauss (2×10^{-5} tesla), two-thirds the value of Earth's field, because of Saturn's greater size. The magnetosphere contains a plasma of charged particles, but no distinct torus like the one associated with Io in the Jupiter system. The charged particle belts are absent in the region of the rings. Radio signals resembling radiation from lightning discharges in the atmosphere were detected by the *Voyager* spacecraft. *See* MAGNETOSPHERE.

Origin. Jupiter and Saturn are relatively similar bodies. Both seem to have bulk compositions close to that of the Sun and the other stars, and both are rich in hydrogen and helium. In that sense, a large fraction of the mass they contain consists of the same primitive material from which the entire solar system was formed, whereas the inner planets such as Earth and Mars never accumulated a solar proportion of the light gases.

This conclusion has been strengthened by the detection of the heavy isotope of hydrogen known as deuterium, in the form of HD (mixed with ordinary hydrogen, H_2) and CH_3D (mixed with ordinary methane, CH_4). The derived value of D/H appears to be identical to the value of this ratio measured in Jupiter's atmosphere by the *Galileo* entry probe, which is 10 times lower than the value in seawater on Earth and twice as high as the value in local interstellar hydrogen. This demonstrates that the hydrogen that makes up most of the mass of both Jupiter and Saturn was captured directly from the solar nebula when these planets formed 4.6 billion years ago. The hydrogen in seawater came from water delivered to the Earth by comets and rocks and has a higher value of D/H, while the current interstellar ratio reflects the result of nucleosynthesis in stars, which converts deuterium into helium and has therefore depleted this isotope during the last 4.6 billion years by a factor of 2. *See* DEUTERIUM; NUCLEOSYNTHESIS.

However, both Jupiter and Saturn show an enhancement of C/H (as determined from methane and hydrogen) compared with the Sun. On Jupiter, C/H is three times the solar ratio, while on Saturn the enhancement is a factor of 6. This suggests that both planets formed in a two-stage process that led initially to formation of a large core of approximately the same size that generated an outgassed, secondary atmosphere, followed by the attraction of an envelope of gases from the surrounding nebula. *See* SOLAR SYSTEM.

Ring system. The most remarkable feature associated with Saturn is the complex ring system that surrounds the planet (**Fig. 3; color plate 1**). The system is divided into four main regions, designated A through D. The narrow F ring is located just beyond the edge of ring A, and there are G and E rings still farther out. Each of the four main regions is subdivided into many individual "ringlets," so that Saturn is actually surrounded by thousands of rings.

The ring system is made up of myriad separate particles that move independently in flat, mostly circular orbits in Saturn's equatorial plane. The discontinuous, meteoric nature of the rings is demonstrated directly by spectroscopic observations, which show that the inner edge of the ring revolves about the planet faster than the outer edge, and the material in each ringlet moves precisely with the velocity that independent satellites would have at the same distance from the planet.

Appearance from Earth. The nearly circular ring system looks like an ellipse whose appearance changes with

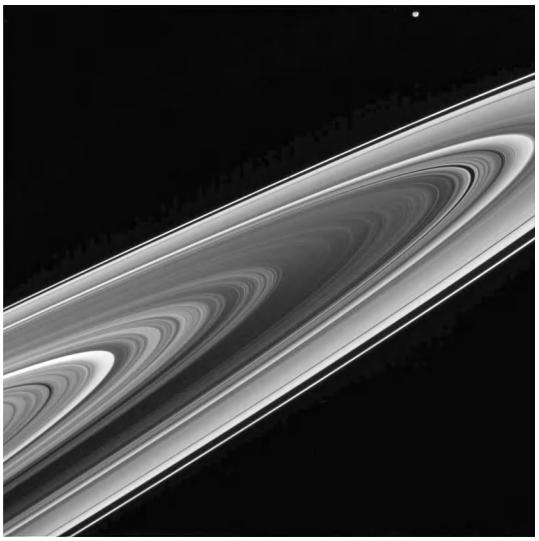


Fig. 3. Visible-light image of Saturn's rings from the *Cassini Orbiter*. The Sun is shining through the rings so that they take on the appearance of a photonegative; the dense B ring at the center blocks much of the incoming light, while the less dense regions scatter and transmit light. The image was taken on April 15, 2005 at a distance of 350,000 mi (570,000 km) from Saturn. (*NASA/JPL/Space Science Institute*)

the relative positions of Earth and Saturn (Fig. 2). The maximum opening of the rings occurs when the system is tilted 27° to the line of sight. In contrast, the rings appear as a thin line when Earth crosses Saturn's equatorial plane. Depending on the position of Earth in its orbit, the slightly variable point of view causes a slight annual oscillation of the tilt angle. Thus, when the tilt is near 0°, Earth may cross the plane of the ring system either once or three times. During these edge-on configurations, it is possible to search for faint inner satellites and distant rings that ordinarily are hidden by the bright light reflected by the main ring system.

Nomenclature and structure. The bright outer ring, A, has an outside diameter of 170,000 mi (274,000 km) and an inner diameter of 150,000 mi (244,000 km). This ring contains many wavelike patterns, indicating gravitational disturbances by the satellites. A dark gap in the A ring, often called the Encke division, was found to contain a narrow, elliptical, discontinuous and kinky ring, together with a small satellite, Pan. The broad (2800 mi or 4500 km wide) division separating ring A from B is called the Cassini division after its discoverer and contains material comprising at least five discrete rings, each of which shows some internal structure.

Ring B, the brightest component of the system, has an outer diameter of 146,000 mi (235,000 km) and an inner diameter of 114,000 mi (184,000 km). This ring contains the largest density of material in the system. It is slightly less bright in its inner regions, where it is also more transparent.

Ring C, sometimes called the crepe ring, is much fainter and more transparent. It appears as a dusky band in projection against the disk of the planet and only faintly against the sky. Its outer diameter is 114,000 mi (184,000 km) and its inner diameter 92,000 mi (149,000 km). This means the inner edge of ring C is just 8700 mi (14,000 km) above the visible cloud deck on the planet.

A fourth zone, D, of the rings of Saturn between C and the globe is fainter than the crepe ring, and still fainter rings lie beyond the A ring.

The E ring begins inside the orbit of Enceladus at a distance of about 112,000 mi (180,000 km) from the center of Saturn, and extends outward past the orbit of Dione, fading from view at about 300,000 mi (480,000 km) from the planet's center. The ring is very much brighter just at the orbit of Enceladus, suggesting that this satellite is somehow responsible for the production of the material seen in the ring.

The F ring (**Fig. 4**) lies just outside the A ring with an average diameter of 174,600 mi (281,000 km), and actually consists of more than five separate strands, which may not all be in the same plane. The two brightest strands deviate markedly from ellipses, showing the effects of nongravitational forces and the influence of embedded moonlets. The F ring is held in place by two small satellites, one inside and one outside. A third guards the outer edge of the A ring, and a fourth is responsible for the Encke gap.

Finally, the tenuous G ring has an average diameter of 211,000 mi (340,000 km), 21,000 mi (33,400 km) outside the A ring. Although this feature resembles the E ring, no known satellites are associated with it, so both its origin and its ability to persist are not understood.

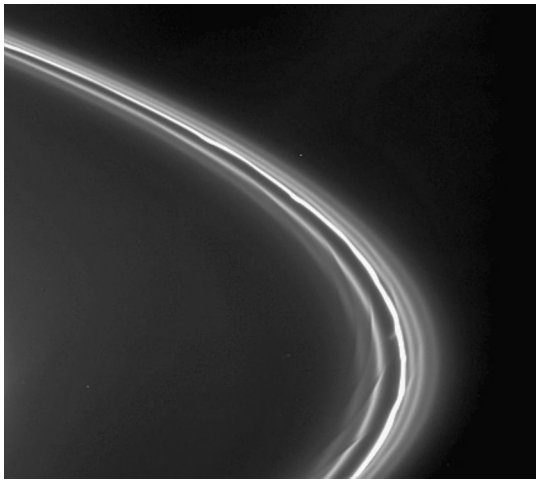


Fig. 4. Saturn's F ring viewed from the *Cassini Orbiter*. Features resembling drapes and kinks are visible, and several distinct ringlets are present, in addition to the bright knotted core of the ring. The structure in the ring and its strands has been caused by Prometheus, the inner F ring shepherd moon that swept by this region shortly before the image was taken on January 19, 2005 at a distance of 1.2×10^6 mi (1.9×10^6 km) from Saturn. (*NASA/JPL/Space Science Institute*)

To place this complex system in perspective, the mean distance from the Earth to its Moon is 240,000 mi (384,000 km), so Saturn and its rings (except for the outer edge of the E ring) would just fit in the space between the Earth and the Moon.

In addition to these nearly circular rings, transient radial features exist in the B ring. These features, called spokes, appear to be clouds of micrometer-size charged particles that are initially controlled by the planet's magnetic field but soon begin moving in keplerian orbits like the larger particles. The change in solar illumination, hence charging, of the ring particles between the *Voyager* and the beginning of the *Cassini* missions changed the charging process sufficiently that initially no spokes were visible to the *Cassini* cameras.

Origin and nature. Édouard Roche proved in 1849 that a liquid satellite with the same density as its planet cannot form if it is closer than 2.44 times the planetary radius. Within this distance, which is known as the Roche limit, disruptive tidal forces (the gradient of the planet's gravitational field) will be greater than the self-gravity of the satellite. But a rocky or icy satellite is held together by stronger forces, and will, therefore, not be disrupted by the planet's field outside a distance of 1.35 radii. In fact, the F ring and its guardian satellites are at 2.33 radii, just inside the Roche limit. If dispersed matter exists within the Roche limit, for example, in the form of a planetary ring, the gravitational attraction of individual ring particles for one another will be insufficient to overcome the planet's tidal forces. Thus a satellite cannot form from a ring that lies within the Roche limit. The ring will persist. *See* ROCHE LIMIT.

In 1859 J. C. Maxwell was able to show that a

TABLE 1. Saturn's satellites[a]

Satellite	Mean distance from center of Saturn, 10^3 km	Revolution period d	h	m	Mean density, g/cm^3	Diameter, km	Visual magnitude at mean opposition
XVIII Pan	134	0	13	49	?	(20)	(19)
XV Atlas	137	0	14	26	?	30[b]	(18)
XVI Prometheus	139	0	14	43	?	100[b]	(16)
XVII Pandora	142	0	15	05	?	90[b]	(16)
X Janus	151	0	16	41	?	190[b]	(14)
XI Epimetheus	151	0	16	41	?	120[b]	(15)
I Mimas	187	0	22	37	(1.2)	390	12.9
XXXII Methone	194	1	0	13	?	3	(23)
XXXIII Pallene	211	1	3	27	?	4	(22)
II Enceladus	238	1	08	53	1.6	500	11.7
III Tethys	295	1	21	18	(1.2)	1060	10.2
XIII Telesto	295	1	21	18[c]	?	25[b]	(19)
XIV Calypso	295	1	21	18[d]	?	25[b]	(19)
XXXIV Polydeuces	377	2	17	41[e]	?	3.5	(23)
IV Dione	378	2	17	41	1.4	1120	10.4
XII Helene	378	2	17	41[f]	?	30[b]	(18)
V Rhea	526	4	12	25	1.3	1530	9.7
VI Titan	1221	15	22	41	1.88	5550[g]	8.3
VII Hyperion	1481	21	06	38	?	255[b]	14.2
VIII Iapetus	3561	79	07	56	1.2	1460	10.2–11.9
IX Phoebe	12,940	550	11		?	220	16.4

[a]Values in parentheses still uncertain. The 13 irregular satellites discovered in 2000 and 2004 are listed separately in Table 2. Four satellites have not been observed well enough for their orbits to be determined, and therefore they are not listed in this table. 1 km = 0.62 mi.
[b]Irregular in shape
[c]Librates about trailing (L_5) lagrangian point of Tethys's orbit.
[c]Librates about leading (L_4) lagrangian point of Tethys's orbit.
[e]Librates about trailing (L_5) lagrangian point of Dione's orbit.
[f]Librates about leading (L_4) lagrangian point of Dione's orbit.
[g]This diameter refers to top of haze layer. Diameter of solid body is 5150 km.

ring system of small mass, formed of a large number of particles, is quite stable against external perturbations such as those caused by the larger satellites. The aggregate mass of Saturn's ring system is quite small, probably less than one-half the mass of the Earth's Moon. Periodic perturbations by the major satellites are responsible, in part, for the main divisions of Saturn's rings in the same way that perturbations by Jupiter cause the Kirkwood gaps in the asteroid belt. The cause of many small divisions (Fig. 3) is still obscure. It has been proposed that tiny moonlets may be present in these gaps, stabilizing individual rings even as the F ring is stabilized by its two satellites. But a careful search has failed to reveal any small satellites (with diameters of 5–10 km) in any of the gaps except Encke's. *See* ASTEROID; PERTURBATION (ASTRONOMY).

The reflection spectrum of most of the ring system is identical with that of water ice, indicating that these ring particles are either ice or covered with ice. However, there are discrete ringlets with differing reflective properties, suggesting the presence of other materials such as silicates or organic compounds. The ring system is a surprisingly good reflector of radar waves, suggesting that some relatively large (diameter approximately 0.6 mi or 1 km) particles may be present. An average effective particle size of 3 ft (1 m) has been indicated. The mean thickness of the ring system is not well determined, but *Voyager* observations of the occultation of a star by the rings demonstrated that the outer edge of the A ring is only 500 ft (150 m) thick.

Satellites. As of June 1, 2005, Saturn had 34 well-established satellites, with at least 13 more whose orbits were being confirmed. Still more are expected to be discovered (**Tables 1** and **2**). The largest and brightest, Titan, is visible with small telescopes; the other satellites are much fainter. Two inner satellites (X Janus and XI Epimetheus) are very nearly in the same orbit. Two small satellites are, respectively, in the leading and trailing lagrangian points of Dione's orbit, and two are in the leading and trailing lagrangian points of Tethys's orbit. All of these objects are small (less than 60 mi or 100 km long) and irregular in shape. The *Cassini* spacecraft found two even smaller objects (0.6–1.2 mi or 1–2 km) in independent orbits between Mimas and Enceladus (Table 1). More discoveries may follow. *See* CELESTIAL MECHANICS.

Titan. Titan shows a measurable disk in large telescopes; the mean apparent diameter corresponds to a linear diameter of approximately 3440 mi (5550 km). But this diameter refers to the satellite's atmosphere, which contains several layers of smog consisting of aerosols produced photochemically by incident sunlight. The solid surface of Titan has a diameter of 3200 mi (5150 km), making this satellite larger than Mercury but slightly smaller than Jupiter's giant Ganymede. This large satellite has a mass about two times that of the Moon, with a corresponding mean density of 1.9 g/cm^3. The low density (Moon = 3.3) means that this object contains a large fraction of icy material and is thus quite different from the Moon or the inner planets in composition. Furthermore, it is large and cold enough to retain an atmosphere of gases with relatively high molecular weights.

In fact, Titan has a thick, nitrogen (N_2)-dominated atmosphere that contains a few percent of methane

Fig. 5. Image from the *Huygens* probe from an altitude of about 5 mi (8 km) above Titan's surface, showing the boundary between lighter uplifted terrain, marked with what appear to be drainage channels, and darker, lower areas. (*ESA/NASA/JPL/University of Arizona*)

(CH_4) and exerts a surface pressure of 1.5 bars (1.5×10^5 Pa), or 1.5 times the sea-level pressure on Earth. Photochemical reactions in the atmosphere produce traces of many interesting compounds, such as acetylene (C_2H_2), hydrogen cyanide (HCN), ethane, and propane (C_3H_8), which condense and polymerize in the stratosphere, forming thick surface-shielding haze layers. The atmosphere is sufficiently thick to have allowed the safe entry and 2 h, 20 min descent by parachute of the *Huygens* probe on January 14, 2005. Instruments on the probe successfully measured the atmospheric temperature, pressure, and composition along the descent trajectory while obtaining pictures (**Fig. 5**) and spectra of the surface. After landing, the probe continued to transmit data for another hour (**color plate 2**). The surface of Titan is so cold (94 ± 2 K or -290 ± 4°F) that rivers and lakes of liquid methane must be present. Branching channel systems resembling rivers on Earth were seen by the probe, which also detected liquid methane in the surface material at the landing site. The cameras and the radar on the orbiter have found very few craters, indicating that Titan has some active processes at work that are burying and erasing them. A search for lakes, seas, and geysers of methane is continuing. The atmospheric chemistry and the solid surface on which liquids and aerosols can accumulate make Titan an attractive natural laboratory in which future investigators will be able to test ideas about chemical evolution on the primitive Earth.

Other inner satellites. The other inner satellites encompass a variety of characteristics. Most have densities near unity, indicating a predominantly icy composition. The variations in density that do occur appear random, rather than showing a radial trend as is the case for the Jupiter system. These density differences must reflect variations in the proportion of rocky material embedded in the ices. The surfaces of these objects are covered with impact craters, with the exception of Enceladus. Large regions of the surface of Enceladus are free of craters, indicating reworking of the surface in relatively recent times. Coupled with the unusually high reflectivity of this satellite (close to 100%), the modified surface suggests internal activity leading to partial melting or pulverization of the icy surface, and the production and expulsion of the tiny ice grains that populate the E ring. The *Cassini* spacecraft discovered the emission of water vapor from fissures at the satellite's south pole. Iapetus is unique in the solar system in that its trailing hemisphere is six times brighter than the leading one. The dark material on the leading hemisphere appears to be a mixture of organic compounds, but the origin of this material and the process that coated the satellite's surface with it remain unknown. All of these satellites except Hyperion keep the same hemisphere facing Saturn as they revolve around it, meaning that their rotational periods are equal to the periods of revolution as a result of tidal friction, just as is the case for the Moon. Hyperion exhibits an aperiodic, "chaotic" rotation, which, with its irregular shape, may be an indication of a relatively recent collision. *See* MOON.

Irregular satellites. Saturn also has a number of more distant satellites that are classified as "irregular" because of the large eccentricities or inclinations (or both) of their orbits. These satellites must have been captured by Saturn early in its history rather than forming together with it from the solar nebula. The largest of these is Phoebe (**Fig. 6**), which moves in a retrograde direction (opposite to that of the inner satellites and to the direction of the planets about the Sun) in an orbit of relatively high eccentricity (0.16).

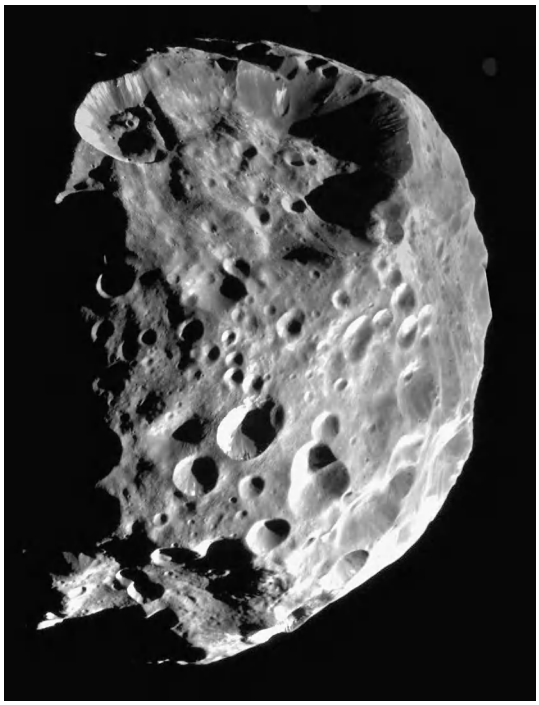


Fig. 6. Mosaic of two images of Saturn's outer satellite Phoebe taken during *Cassini*'s flyby on June 11, 2004. The image shows evidence for the theory that Phoebe is an ice-rich body coated with a thin layer of dark material. (*NASA/JPL/Space Science Institute*)

TABLE 2. Irregular satellites of Saturn*

Satellite	Mean distance from center of Saturn, 10^6 km†	Eccentricity	Inclination	Revolution period, years	Visual magnitude at mean opposition	Radius, km†
XXVI Albiorix	17.9	0.38	33.1°	2.43	20.5	13
XX1 Tarvos	18.2	0.54	33.5°	2.53	22.1	7
XXVIII Erriapo	18.2	0.48	33.8°	2.50	23.0	4
XXIX Siarnaq	16.8	0.26	44.8°	2.18	20.1	16
XX Paaliaq	15.2	0.36	45.1°	1.88	21.3	10
XXIV Kiviuq	11.3	0.33	46.1°	1.23	22.0	7
XXII Ijiraq	11.4	0.32	46.7°	1.24	22.6	5
XXXI Narvi	18.8	0.43	145.8°	2.76	24.0	3.3
XXVII Skathi	15.5	0.27	152.6°	1.97	23.6	3
XXV Mundilfari	18.4	0.23	167.4°	2.56	23.8	3
XIX Ymir	23.1	0.34	173.1°	3.59	21.7	8
IX Phoebe	12.94	0.16	174.7°	1.51	16.4	110
XXX Thrymr	20.1	0.44	175.9°	2.92	23.9	3
XXIII Suttungr	17.6	0.12	176.0°	2.40	23.9	3

*Data provided by the leader of one of the discovery teams, Dr. Brett Gladman.
†1 km = 0.62 mi.

Phoebe's rotation period is only 9 h, completely out of synchrony with its 550-day period of revolution. *See* RETROGRADE MOTION (ASTRONOMY).

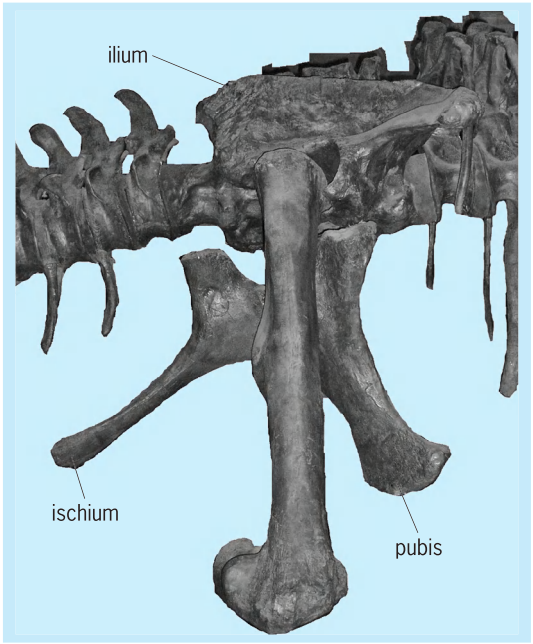
Twelve additional irregular satellites were confirmed in 2000 and one in 2004 (Table 2) as a result of deliberate searches with sensitive new detectors on Earth-based telescopes. More discoveries are expected. These small objects (the largest is only 20 mi or 32 km across) can be classified into four distinct groups by the inclinations of their orbits: three near 34°, four near 46°, two near 150°, and four near Phoebe's inclination of 175°. All the orbits have high eccentricities, and this plus the high inclinations mark these satellites as irregular. The distinct groups suggest separate capture events in which the original object was broken apart during the capturing process. Similar families exist in the Jupiter satellite system. The exploration of the Saturn system by the *Cassini Orbiter* is continuing. *See* SATELLITE (ASTRONOMY).

Tobias C. Owen

Bibliography. J. K. Beatty, C. C. Petersen, and A. Chaikan (eds.), *The New Solar System*, 4th ed., Sky Publishing/Cambridge University Press, 1999; K. R. Lang, *The Cambridge Guide to the Solar System*, Cambridge University Press, 2003; D. Morrison and T. Owen, *The Planetary System*, 3d ed., Addison Wesley, 2003; Special issue on *Voyager 1* encounter with the Saturnian system, *Science*, 212(4491):159–243, 1981; Special issue on *Voyager 2* encounter, *Science*, 215(4532):499–594, 1981; Special issue on *Cassini* at Titan, *Science*, 308(5724):969–995, 2005.

Saurischia

One of two major monophyletic clades of Dinosauria, the other being Ornithischia. Saurischian dinosaurs take their name (meaning "lizard hipped") from the structure of their pelvis. In saurischians, the pubis bone points downward and forward from the hip joint, and the ischium extends downward and backward, so that the two bones are separated by a wide angle in side view (see **illustration**). This feature allows the majority of saurischians to be easily distinguished from ornithischian dinosaurs (in which the pubis has rotated so that it points downward and backward and thus lies alongside the ischium). However, the saurischian condition is actually a retention of the primitive reptilian condition (which is also present in crocodilians, squamates, as well as many extinct groups) and, therefore, cannot be regarded as a unique feature of saurischians. To add to the confusion, some saurischian dinosaurs (various advanced theropod groups and their direct descendants, the birds) have independently evolved an ornithischian-type pelvis. However, Saurischia can be satisfactorily defined as a monophyletic clade of animals on the basis of many other detailed anatomical features shared by all members of the group, largely pertaining to the structure of the hand and vertebral column, such as the possession of a stout, robust thumb and the presence of



Pelvic girdle of the sauropod dinosaur *Camarasaurus*, showing the hip orientation typical of saurischian dinosaurs in which the pubis and ischium diverge from the hip joint. The front of the animal is to the right.

incipient lamination on the vertebrae. *See* DINOSAURIA; ORNITHISCHIA.

The earliest saurischians—including *Staurikosaurus*, *Herrerasaurus*, and *Eoraptor*—are known from deposits of Late Triassic age (approximately 225 million years ago) in Brazil and Argentina. Early in its evolutionary history, Saurischia split into two major monophyletic clades: the Sauropodomorpha and Theropoda. Both groups diversified rapidly in the Late Triassic–Early Jurassic interval, each increasing in species richness and abundance to become the dominant vertebrates in terrestrial ecosystems. They quickly achieved a global distribution, and the remains of early saurischians are known from all continents.

Sauropodomorpha. Sauropodomorpha can be defined on the basis of many anatomical characteristics, including a relatively small skull (length is less than 50% of femur length) and various features of the skull and dentition (for example, tooth crowns with coarse marginal serrations set at 45° to the long axis of the tooth) and the postcranial skeleton (for example, a "funnel-shaped" fifth metatarsal). Traditionally, sauropodomorphs have been divided into two groups: Sauropoda and Prosauropoda. Sauropods are generally recognized as a monophyletic group on the basis of numerous synapomorphies (shared derived traits), which are largely concentrated in the vertebral column (for example, the presence of a complex system of bony laminae on the cervical vertebrae), and limbs (for example, elongated forelimbs relative to other dinosaurs). The relationships of prosauropods are more controversial: some paleontologists suggest that they form a monophyletic clade, whereas others believe they form an array of forms that are ancestral to the later sauropods. In general, all sauropodomorphs share the same body plan: a small head, elongate neck, barrel-shaped body, and long, counterbalancing tail.

Prosauropods and basal sauropodomorphs first appeared early in the Late Triassic, represented by small (probably bipedal) animals such as *Saturnalia* and *Thecodontosaurus*, and quickly achieved a global distribution (excluding Australia). These forms were largely herbivorous, but may have taken animal prey and/or carrion occasionally. Later prosauropods increased in size, with some forms reaching 10 m (33 ft) in length: these larger animals were probably entirely herbivorous. The majority of prosauropods were facultatively bipedal, spending most of their time on all fours, but occasionally walking bipedally: very large prosauropods (such as *Riojasaurus*) were obligate quadrupeds. Prosauropods were confined to the Late Triassic–Early Jurassic, an interval in which they were the dominant terrestrial herbivores. The reason for their extinction is unknown, but coincides with the diversification of the ecologically similar sauropods.

Sauropods also appeared in the Late Triassic, but initially were rare members of early dinosaur faunas. In general, sauropod evolution through the Jurassic and Cretaceous followed several trends, including increased body size [culminating in animals such as *Argentinosaurus*, from the Late Cretaceous of Patagonia, which may have reached a weight of 70 metric tons (77 tons)] and increased neck length, the latter allowing feeding over wide vertical and horizontal ranges. All sauropods were obligate quadrupeds and were more advanced herbivores than prosauropods, having acquired a number of cranial and dental features (such as interlocking dentitions) that allowed more thorough processing of plant food than could be achieved by prosauropods. By the Middle Jurassic, sauropods had become the dominant terrestrial vertebrates and had achieved a global distribution (excluding Antarctica). Sauropods reached their acme in the Late Jurassic, after which they declined in diversity and abundance, though a second peak in diversity also occurred in the middle to Late Cretaceous. The group survived until the end of the Cretaceous Period. *See* CRETACEOUS; JURASSIC; TRIASSIC.

Theropoda. All theropods possess an intramandibular joint within their lower jaws, which allowed the front end of the mandible to flex slightly with respect to the back of the jaw. In addition, various features of the hands and vertebrae (such as elongated processes on the tail vertebrae) help to define Theropoda as a monophyletic group, which also includes birds. All nonavian theropods were bipedal, and the majority were carnivorous, as evidenced by the presence of recurved teeth with fine, steak knife–like serrations, grasping hands, and large claws on both hands and feet. A few Cretaceous forms (such as therizinosaurs and ornithomimosaurs) may have been omnivores or herbivores. Some authors regard *Eoraptor* and *Herrerasaurus* as the earliest theropods, though others argue that these animals are more primitive saurischians. Nevertheless, it is generally accepted that the earliest definite theropods (for example, *Coelophysis*) appeared in the Late Triassic and quickly radiated to become the most abundant and diverse predators of Late Mesozoic terrestrial ecosystems. Soon after its origin the group achieved a global distribution. Nonavian theropods survived until the end of the Cretaceous Period; avian theropods survived the Cretaceous/Tertiary (K/T) extinction event and, therefore, represent the only group of living dinosaurs.

Theropods ranged in size from animals less than a meter in length (for example, *Microraptor*), which were probably largely insectivorous, to gigantic superpredators (such as *Tyrannosaurus rex*, *Giganotosaurus*, and *Spinosaurus*) over 12 m (39 ft) long that weighed up to 6.5 metric tons (7.2 tons) and preyed upon other dinosaurs. A few theropods, such as *Deinonychus*, are believed to have been pack-hunters.

The dominant theme in theropod evolution relates to their transformation from scaly, fast-running, terrestrial bipeds to flying birds, with the subsequent transformation of forelimbs into wings and many other modifications associated with flight, including changes in brain anatomy, modifications to the hindlimbs and pelvic girdle, and the origin of feathers. The evolution of feathers is now known to have occurred in nonavian theropods, with filamentous "protofeathers" (as in *Sinosauropteryx*)

giving rise to true feathers (with vanes and barbs, as in *Sinornithosaurus*) to asymmetrical flight feathers (as in *Archaeopteryx*, the earliest-known bird). It is likely that feathers initially appeared for intraspecific display, or egg brooding or insulation, and were later co-opted for flight. *See* AVES; FEATHER.

Paul M. Barrett

Bibliography. P. J. Currie, and K. Padian (eds.), *The Encyclopedia of Dinosaurs*, Academic Press, San Diego, 1997; J. O. Farlow and M. K. Brett-Surman (eds.), *The Complete Dinosaur*, Indiana University Press, Bloomington, 1997; D. B. Weishampel, P. Dodson, and H. Osmólska (eds.), *The Dinosauria*, 2d ed., University of California Press, Berkeley, 2004.

Sauropterygia

An infraclass of Mesozoic reptiles that are, without exception, adapted to the marine environment. The infraclass includes the nothosaurs, plesiosaurs, and placodonts. These reptiles, along with the ichthyosaurs, played a significant role as predators within the marine animal community of the Mesozoic Era.

Placodontia. The placodonts are a distinctive but highly varied assemblage of aquatic reptiles. They had short bodies, paddlelike limbs, and flat cheek teeth designed for crushing hard-shelled prey. The genus *Helodus* was covered by a dorsal bony armor and a roofing of dermal scutes, but most other genera lacked armor. Placodonts have come only from rocks of the Middle and Upper Triassic of Europe, North Africa, and the Middle East.

Nothosauria. The Nothosauria are the relatively generalized stem group from which the plesiosaurs evolved. With the exception of a single New World species and a record from Japan, nothosaurs are known primarily from Europe and the Near East (Israel) in rocks of Triassic age. The nothosaurs are notably diverse in the mode and degree of secondary aquatic modification. The early phase of this process manifests itself (1) by a change in the histological structure of the bones, characterized by a total absence of Haversian systems and often a swollen appearance of the bones (pachyostosis); (2) by a notable reduction in size of the lateral girdle bones, scapulae (shoulder), and ilia (pelvis); and (3) by enlargement of the ventral elements and by the loss of digital claws. The directions of aquatic specialization involve shortening, or more often lengthening, of the neck (see **illus.**); enlargement of the orbits or the temporal fenestrae; and reduction, or more commonly increase, in the number of phalanges in manus (hand), pes (foot), or both. A feature of considerable evolutionary significance in the light of plesiosaurian differentiation is the great individual variability in the number of presacral vertebrae (32–42) in *Pachypleurosaurus edwardsi*. *See* SKELETAL SYSTEM.

Although most genera of nothosaurs are not in direct line of ancestry of the pleisosaurs, such a relationship has been suggested for *Pistosaurus*. The North American form, *Corosaurus alcovensis*, is probably the most advanced nothosaur known. *See* NOTHOSAURIA.

Plesiosauria. The Plesiosauria are the successful, compact, and highly specialized offshoot of the nothosaurs that attained worldwide distribution. The early steps of aquatic adaptation, initiated by the nothosaurs, have led to extensive anatomical modifications: The region comprising chest and abdomen became short, stout, and inflexible; the ventral bones of shoulder girdle and pelvis increased in area enormously; and the limbs, transformed into large flippers, became the principal organs of propulsion.

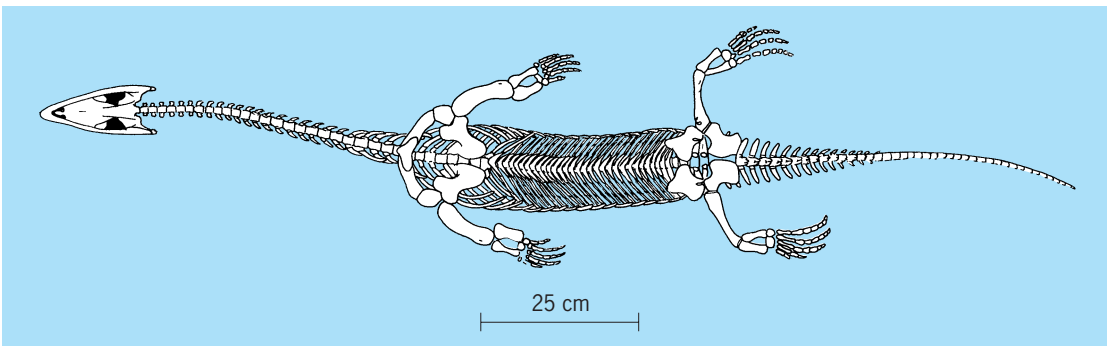
Two major trends of plesiosaur evolution may be discerned since the Early Jurassic: In the one group there was a tendency toward a shortening of the neck from 27 to 13 vertebrae and an increase in skull size: in the other group the opposite trend led to forms of bizarre body proportions, for example, *Elasmosaurus*, with a neck containing 76 vertebrae. The plesiosaurs were carnivorous. The precise affinities of placodonts, nothosaurs, and plesiosaurs remain uncertain. No earlier group of reptiles from which they might have come is known, and they left no descendants. *See* PLACODONTIA; PLESIOSAURIA; REPTILIA.

Rainer Zangerl; Everett C. Olson

Bibliography. R. L. Carroll, *Vertebrate Paleontology and Evolution*, 1987.

Savanna

The term savanna was originally used to describe a tropical grassland with more or less scattered dense tree areas. This vegetation type is very abundant in tropical and subtropical areas, primarily because of



***Ceresiosaurus calcagnii*, ventral view.**

climatic factors. The modern definition of savanna includes a variety of physiognomically or environmentally similar vegetation types in tropical and extratropical regions.

In the widest sense savanna includes a range of vegetable zones from tropical savannas with vegetation types such as the savanna woodlands to tropical grassland and thornbush. In the extratropical regions it includes the "temperate" and "cold savanna" vegetation types known under such names as taiga, forest tundra, or glades. The physiognomically savannalike extratropical vegetation types differ greatly in environment and species composition.

During the growing season the typical tropical savanna displays a short-to-tall, green-to-silvery shiny cover of bunch grasses, with either single trees or groups of trees widely scattered. This is followed by a rest period of several months during which, because of severe drought, the vegetation appears quite different, with the brown-gray dead grasses bent over and the trees either without leaves or with stiff or wilted gray-green foliage. The heat and drought during this season of the year exert a high selective pressure upon the floral and faunal composition of the savanna.

Floral and faunal composition. The physiognomic similarity of the tropical savannas is underlined by the similarity among certain floristic components. All savannas contain members of the grass family (Gramineae) in the herbaceous layer. Most savannas of the world also have one or more members of the tree family (Leguminosae), particularly of the genus *Acacia*. Also included among the trees are the families Bombacaceae, Bignoniaceae, Palmae, and Dilleniaceae, and the genera *Prosopis* and *Eucalyptus*; these are abundant when they occur. One of the most outstanding savanna trees is *Adansonia digitata* (Bombacaceae), which achieves one of the biggest trunk diameters known for all trees (**Fig. 1**). The grass species, although mostly from the genera *Panicum*, *Paspalum*, and *Andropogon*, include numerous other genera such as *Aristida*, *Eragrostis*, *Schmidtia*, *Trachypogon*, *Axonopus*, *Triodia*, and *Plectrachne*, all of regional importance.

Fig. 1. ***Adansonia digitata*** **of Senegal. (*Courtesy of H. Lieth*)**

The fauna of the savannas is among the most interesting in the world. Savannas shelter herds of mammals such as the genera *Antelopus*, *Gazellus*, and *Giraffus*, and the African savannas are especially famous for their enormous species diversity, including various members of the Felidae (for example, the lion), and the elephant.

Numerous species of birds are indigenous to the savannas. Among these is the biggest bird, the ostrich (*Struthio camelus*), found in Africa. Many birds from extratropical regions migrate into the tropical savannas when the unfavorable season occurs.

Among the lower animals, the ants and termites are most abundant. Termite colonies erect large, conical nests above the ground, which are so prominent in some savannas that they partly dominate the view of the landscape, especially during the dry season.

Environmental conditions. The climate of the tropical savannas is marked by high temperatures with more or less seasonal fluctuations. Temperatures rarely fall below 32°F (0°C). The most characteristic climatic feature, however, is the seasonal rainfall, which usually comes during the 3–5 months of the astronomic summertime. Nearly all savannas are in regions with average annual temperatures from 59 to 77°F (15 to 25°C) and an annual rainfall of 32 in. (81 cm).

The soil under tropical savannas shows a diversity similar to that known from other semiarid regions. Black soils, mostly "chernozem," are common in the moister regions. Hardpans and occasional surface salinities are also found, and lateritic conglomerations occur along the rivers. The soils vary in mineral nutrient level, depending on geologic age, climate, and parent material. Deficiencies of minerals, specifically trace elements, as well as aluminum toxicity, are reported from many grassland areas in the continents of Africa, South America, and Australia. *See* SOIL.

Certain savanna areas suffer from severe erosion and no soil can be accumulated. Plant growth in these areas is scarce, often depending on cracks and crevices in the ground material to support tree roots. The herbaceous cover provides only a thin cover for the otherwise rocky surface.

Water is the main limiting environmental factor in the majority of the tropical savannas. Total amount and seasonality of precipitation are unfavorable for tree growth. Additional stresses to forest vegetation are caused by frequent fires, normal activities of animals, excess of salt, or nutrient deficiencies. Wherever a river flows, most of these factors change in favor of tree growth. This explains the existence of extensive gallery forests along the rivers and creeks (**Fig. 2**). The gallery forest is missing only where severe local floods after storms cause soil erosion and the formation of canyons.

Geographic distribution. Tropical savannas exist between the areas of the tropical forests and deserts; this is most apparent in Africa and Australia. In the

Fig. 2. Aerial view of savanna with gallery forest north of Guiaba, Brazil. (*Courtesy of H. Lieth*)

New World tropics, different conditions exist because of the circumstances created by the continental relief, and the savannas are situated between tropical forests and mountain ranges. In Madagascar and India there is a combination of both conditions.

The transitional (ecotonal) position of the savannas between forest and grassland or semidesert is the basis for the differences in opinion among authors about the size and geographical distribution of savannas. Most authors include, however, the savanna in East Africa and the belt south of the Sahara, the bush veld in South Africa, the Llanos in northern South America, and some types of scrub vegetation in Australia. Some areas in southern Madagascar, on several tropical islands (in leeward position), in Central America, in southern North America, and in India are usually included. The two latter regions, however, are subtropical. Still other areas included in the savanna concept are the Campo Cerrado and parts of the Chaco in South America; portions of the Miombo in southern Africa; wide portions of northern Africa; south of the Sahara; and wide portions of Madagascar, the Indian peninsula, and Australia. Because of the variations in the savanna concept, it is difficult to give a correct estimate of the total surface area covered by savanna vegetation. The Food and Agriculture Organization (FAO) considers that about one-third of the total land surface is covered by predominantly grassland vegetation. Of this area, one-third can be assumed to be tropical grassland, most of which can be called savanna.

Agricultural practices. Most of the original savanna areas throughout the world are currently farmed. Ranch farming is the predominant type, with sheep being raised in the drier areas and cattle in the moister regions. In the hotter regions zebus are raised, along with several hybrids of zebus and European cattle, which are the preferred stock for this climate. The yield in meat per unit area is low, and even under extensive management it seems to be lower than the meat production of the natural animal herds of the savanna, including antelopes, giraffes, and zebras. Ranch farming does not change the character of the vegetation very much if it is well managed. The adjacent dry woodlands very often resume a physiognomy similar to the natural savannas if good farm management is applied.

Agricultural crops are of many varieties in the savanna areas, where with careful protection and management any crop can be cultivated, provided that enough water is available. Drought-tolerant crops are usually preferred among the perennials.

The majority of human settlements are small in the savanna regions of the world; the hot temperatures during part of or the entire year, together with problems of water supply, limit interest in larger settlements. Settlements are usually found along the rivers, close to the coast, or in the higher elevations. Much of the land is managed by small tribal villages or large plantation or ranch owners, with separate groups of tenants, sharecroppers or employees, or single families. *See* AGROECOSYSTEM.

Extratropical types. The main structural character of tropical savannas is the scattered trees standing within a close cover of herbaceous vegetation. This structural character is also found in several extratropical vegetation types, but these differ greatly in the forces that limit a close tree cover, including drought (areas intermediate between steppe or prairie and forest); excess water or water combined with soil that has a shortage of oxygen and nutrients (peatbogs, marshes, or glades); short vegetation periods because of extended cold temperatures below the freezing point; excessive, long snow covers; and low light intensity (forest tundra, taiga, and cold savanna).

An intermediate condition is exhibited by the savannas and everglades of the southeastern United States. The intermittent soaked or dry conditions of a peaty soil, the tropically hot summers and mild winters (with frost periods, however), and the generally low nutrient level of the soil give these areas the characteristics of the cold savannalike vegetation types and the tropical and subtropical types.

The economic potential of the three extratropical savannalike areas also varies greatly. The conditions in the southeastern United States allow orchards, tree plantations, and cattle ranching. Economic considerations control whether a given area should be developed.

Some savanna regions in many parts of the world are the last survival territories for many plant and animal species. This implies the need for conserving some savanna pieces, both tropical and extratropical. Some of the most interesting wildlife sanctuaries are set aside for this reason (such as Serengeti, Tanzania; Ruhung, Sri Lanka; and Krüger National Park, South Africa). *See* FOREST AND FORESTRY; GRASSLAND ECOSYSTEM; TAIGA; TUNDRA. Helmut Lieth

Bibliography. M. Cole, *The Savannas: Biogeography and Geobotany*, 1986; R. T. Coupland (ed.), *Grassland Ecosystems of the World*, 1979; P. M. Dansereau, *Biogeography*, 1957; A. Engler and O. Drude (eds.), *Vegetation der Erde*, 15 vols., 1976; L. T. Gutierrez and W. R. Fey, *Ecosystem Succession: A General Hypothesis and a Test Model of a Grassland*, 1980.

Savory

A herb of the mint family in the genus *Satureja*. There are more than 100 species, but only *S. hortensis* (summer savory) and *S. montana* (winter savory) are grown for flavoring purposes. A number of varieties of each species are available, although no varietal names have been assigned to them. *See* LAMIALES.

Summer savory, an annual herb with an upright growth habit reaching about 2.5 ft (0.75 m), is characterized by long thin wiry stems with long internodes between small leaves. Winter savory is a perennial in most climates and, unlike summer savory, will tolerate some freezing weather. It can become woody after one or two growing seasons.

Savory is indigenous to areas surrounding the Mediterranean Sea. *Satureja montana* occurs wild from North Africa to as far north as Russia but is little cultivated. *Satureja hortensis*, which is widely cultivated, is native to Europe.

Savory is a rapid grower that home gardeners can plant any time during the growing season up to 60 days before frost. Winter savory will resume growth in the spring, but summer savory must be replanted. Commercial cultivation employs direct seeding into the field at 10–20 lb/acre (9–18 kg/hectare). It has relatively large seeds, and germination of 60–80% can be expected. Both types of savory are harvested or cut two or three times a year, after which the leaves are dehydrated and separated from the stems to be used as a spice in various foods, including poultry seasoning and beans. *See* SPICE AND FLAVORING. Seth Kirby

Bibliography. S. Arctander, *Perfume and Flavor Materials of Natural Origin*, 1960; L. H. Bailey, *Manual of Cultivated Plants*, rev. ed., 1975.

Saxifragales

An order of flowering plants near the base of the eudicots; DNA sequence analysis has indicated that this order includes few families that have been associated together in previous classifications. The order's exact relationships to other orders are still under study, but Saxifragales appear to be related to the rosid eudicots. They include 13 small to moderately sized families and perhaps around 2300 species. There are two categories: woody species that are sometimes wind-pollinated, and insect-pollinated herbs. The members are difficult to characterize morphologically but fall into several sets of families that have been considered closely related by most authors. *See* MAGNOLIOPHYTA; MAGNOLIOPSIDA; POLLINATION; ROSIDAE.

The families that belong here were formerly considered to be in either Saxifragales in the broad sense or the largely wind-pollinated Hamamelidales. None of these are of particular economic importance except as ornamentals, although *Cercidiphyllum* (Cercidiphyllaceae), *Liquidambar* (sweet gums), *Altingia* (both Altingiaceae), and a few others produce timbers. The largest family is Crassulaceae (1500 species), which are plants of arid zones, including the stonecrops (*Sedum*) and kalanchoes (*Kalanchoe*). A number of species are commonly cultivated ornamentals, including peony (*Paeonia*, Paeoniaceae), saxifrages (*Saxifraga*, Saxifagaceae), and witch hazel (*Hamamelis*, Hamamelidaceae); a few, the currants and gooseberries (*Ribes*, Grossulariaceae), are cultivated as fruits. *See* HAMAMELIDALES; ORNAMENTAL PLANTS. Mark W. Chase

Scalar

A term synonymous in mathematics with "real" in real number or real function. The magnitude of a vector two units in length is the real number or scalar 2. The dot or scalar product of two vectors is the product of three real numbers associated with them (the magnitude of the first vector times the magnitude of the second, times the cosine of the angle between them) and is therefore a scalar. If in the functional relationships $S = S(s,y,z)$, $\mathbf{F} = \mathbf{F}(x,y,z)$, S is a real number and $\mathbf{F}$ is a vector, then $S(s,y,z)$ is a scalar function but $\mathbf{F}(x,y,z)$ is a vector function. More generally, if V is a vector space over a field F, the elements of F are often referred to as scalars. *See* CALCULUS OF VECTORS; LINEAR ALGEBRA. Homer V. Craig

Scale (music)

As a piece of music progresses, it typically outlines a set of pitches by repeatedly sounding a subset of all the possible notes. When these notes are rearranged into ascending or descending order, they are called a musical scale.

There is often a range of pitches that will be heard as the same. Perhaps the trumpet plays middle C a bit flat, while the guitar attacks a bit sharp, in accordance with the artistic requirements of the musical context. The mind hears both pitches as the same note, C, and the limits of acceptability are far cruder than the ear's powers of resolution. This suggests a

kind of categorical perception, where a continuum of possible stimuli (in this case, pitch) is perceived as consisting of a small number of disjoint classes. Thus scales partition pitch space into disjoint chunks. *See* MUSICAL INSTRUMENTS; PITCH; SENSATION.

Western tradition. In the modern Western tradition, scales are standardized subsets of the 12-tone equal temperament (abbreviated 12-tet, and also called the chromatic scale) in which each octave is divided into 12 (approximately) equal sounding divisions. These are further classified into major and minor depending on the exact ordering of the intervals, and are classified into modes depending on the starting point. Historically, however, scales based on 12-tet are fairly recent. *See* TUNING.

The ancient Greeks theorized about scales which are based on what they called pure or just intervals, those in which the fundamental frequencies of two tones are related by simple integer ratios such as 2:1 (the octave), 3:2 (the just fifth), and 4:3 (the just fourth). Pythagoras is often credited with designing the scale that contains the largest possible number of just fourths and fifths, though there are a variety of other ancient scales based on tetrachords (where a pure fourth is divided into three subintervals). Just intonation refers to a family of scales which utilize a variety of intervals based on small integer ratios. Harry Partch championed a scale with 43 unequal tones per octave, and created a family of instruments on which to perform his music. Other historical scales, such as the well temperaments and various meantone tunings, can be viewed as different ways of approximating the just intervals while attempting to maintain the ability to modulate between keys.

Worldwide variety. Other cultures use musical scales that are quite different. For example, traditional Thai music is played in a scale that is close to 7-tone equal temperament. Various Arab musics use scales with a large number of tones; one with 22 unequally spaced tones per octave is frequently cited. The gamelan orchestras of Indonesia consist of a large family of inharmonic metallophones that are tuned to either the 5-note slendro or the 7-tone pelog scales; neither lies close to 12-tet. Moreover, each gamelan is tuned differently, and tunings often have "octaves" that differ significantly from the standard 2:1 frequency ratio.

One explanation for the large variety of intervals and scales used throughout the world is cultural: each musical tradition has followed a unique path that reflects its own history. Another explanation observes that the overtones of string and wind instruments form a harmonic series, and that these overtones overlap best when two notes are played in simple integer ratios. This is one basis for the just intonation approach. Similarly, it has been observed that the inharmonic overtones of certain gamelan instruments overlap best when two tones are played in intervals of the slendro and pelog scales, suggesting that the instruments and scales of a musical tradition may have coevolved to their present state. *See* MUSICAL ACOUSTICS.

William A. Sethares

Bibliography. D. Deutsch (ed.), *The Psychology of Music*, 2d ed., Academic Press, San Diego, 1999; M. Harrison, *Contemporary Music Theory*, 2 vols., Hal Leonard Publishing, 1999; H. Partch, *Genesis of a Music*, 2d ed., Da Capo Press, New York, 1974; W. A. Sethares, *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag, 1998.

Scale (zoology)

The fundamental unit of the primary scaled integument of vertebrates, of which the epidermal or dermal component may be the more conspicuous or elaborated. Scale may have different meanings in different contexts; the semantic confusion has arisen partly for historical reasons, partly because of inherent problems in understanding fossil material, and partly due to increased knowledge of the development of the integument and teeth.

If a scale is defined as an overlapping fold of the epidermal-dermal boundary of the body, it is clear that some vertebrates (for example, a herring or a snake) should be described as having a scaled integument. Similarly, portions of the integument of some vertebrates comprise similar units, such as the leg of a bird or the tail of a rat. However, comparable integumentary units are absent from vertebrates such as sharks, eels, amphibians, and over most of the body of birds and mammals. The integumentary organization of vertebrates is very diverse; to understand the evolutionary relationships between scaled and nonscaled integuments, it is necessary to combine data from paleontology and studies of integumentary morphogenesis.

The fossil record implies that the vertebrate dermis has always possessed the capacity to house bone-forming cells (osteocytes) which give rise to dermal bone by direct deposition of extracellular calcium salts, without a cartilaginous precursor. Much of the skull roof of all vertebrates is formed in this manner. Studies of amphibian and avian embryos have shown that these osteocytes derive from the neural crest. In amphibians it has been shown that the odontoblasts, which form the dentine of the teeth, also derive from the neural crest. Studies of mammalian tooth development demonstrate that odontoblasts, lying beside dermal bone, interact with the overlying buccal ectoderm which then gives rise to ameloblasts, the cells responsible for enamel formation. Because neither odontoblasts nor ameloblasts grown separately from one another are capable of synthesizing dentine or enamel, respectively, it is recognized that odontogenesis involves a complex series of interactions between mesenchyme of neural crest origin and ectoderm. The amount and precise histology of the enamel and dentine (dentomorphic tissues) in various mammalian teeth varies in different mammals. However, a wide variety of investigations has shown that odontogenesis is basically similar in all toothed vertebrates and also that integumentary histogenesis in many fishes is characterized by similar, if not identical, cellular activities. When the histology of the

integumentary sclerifications in Paleozoic fish fossils is examined, there seems little doubt that the odontogenic capacities now confined to the oral region in tetrapods once characterized the entire body surface of their piscine ancestors.

Fishes. There is no way of knowing the precise epidermal structure of fossils. Available data suggest that in the majority of extant fishes the epidermis is extremely thin, consisting largely of mucus-synthesizing cells. The term "scale" when applied to fishes usually refers to the relatively large prominent, dermal ossification, on the outer surface of which lie so-called dentomorphic tissues. However, these sclerifications either may be parts of overlapping units in a scaled integument or may be discrete integumentary appendages where overlapping units are absent.

A full complement of bone plus both types of dentomorphic tissue characterizes the integumentary sclerifications of many "primitive" fishes: many jawless ostracoderms, the acanthodians, and certain bony fishes. The dentomorphic tissues may be seen as tuberculate units protruding irregularly from the surface of the sclerifications as in ostracoderms (**Fig. 1**) or as flattened sheets in acanthodians (**Fig. 2**) or certain bony fishes. The bony tissues consist of deep laminated materials (basal layer) and a superficial spongy vascular region (middle layer). Growth of the integumentary sclerifications involves periodic resorption and redeposition of the dental tissues, as well as the addition of laminate bone. *See* ACANTHODII; OSTRACODERM.

The thelodont ostracoderms and contemporary sharks, two completely unrelated lineages, do not have overlapping scales. Their body surfaces are covered by "dermal denticles" which strongly resemble small teeth. As such, they are true integumentary appendages, although they are frequently referred to as placoid scales.

The evolutionary history of dermal sclerifications varied in different lineages. In general, the tendency is to reduce or lose various of the fundamental constituents. Thus sharks have no bony constituents in their "dermal denticles," while in teleosts only a thin layer of acellular, bonelike material is seen in the dermis. A variety of terms describing "fish scales" appear in the literature; some, for example, cosmoid and ganoid, refer to the detailed histological structure (in sarcopterygians and primitive actinopterygians respectively; **Fig. 3**), while others, for example, rhomboid, cycloid, or ctenoid (**Fig. 4**), refer to the overall shape of the units.

The scaled integument of fishes shows a definite pattern which is governed by the orientation of the underlying myomeres and by changes in body depth and length. The general evolutionary tendency

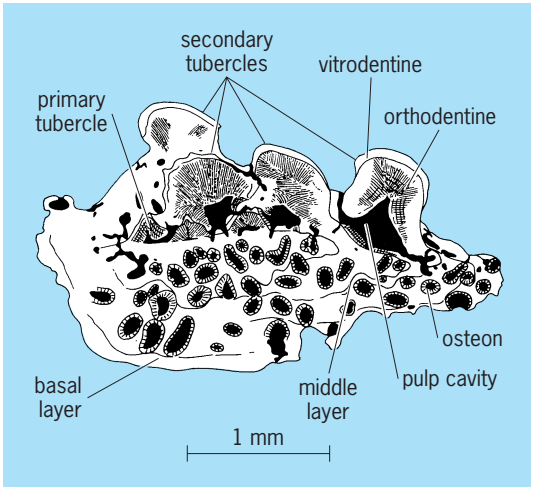


Fig. 1. Cross section of dermal plate of Ordovician ostracoderm. (*After R. H. Denison, The early history of the vertebrate calcified skeleton, Clin. Orthop., no. 31, 143, 1963*)

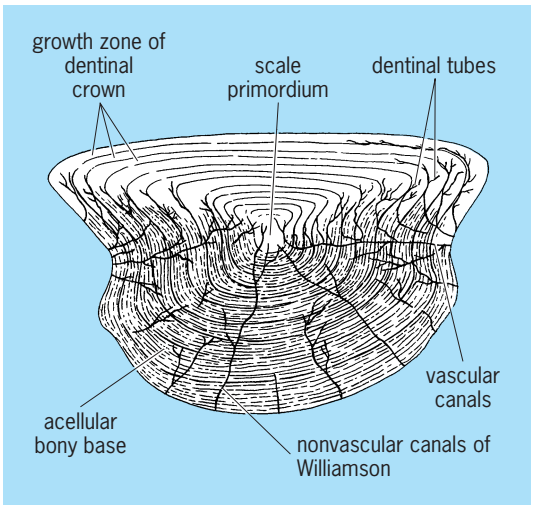


Fig. 2. Cross section of Permian acanthodian scale, showing structural details.

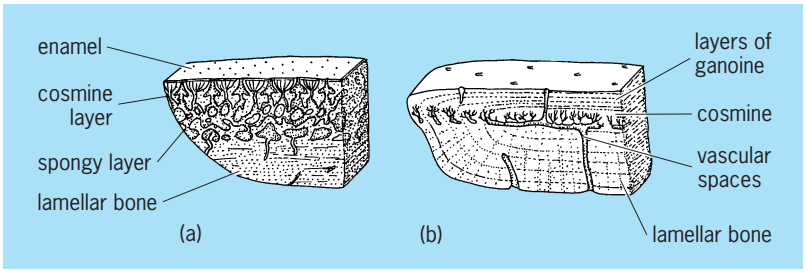


Fig. 3. Cross sections of primitive vertebrate scales. (*a*) Cosmoid scale. (*b*) Ganoid scale. (*After A. S. Romer, The Vertebrate Body, Saunders, 1962*)

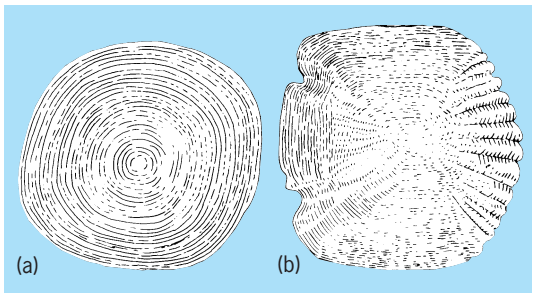


Fig. 4. Scale types. (*a*) Cycloidal scale from burbot. (*b*) Ctenoid scale from carp. (*After K. F. Lagler et al., Ichthyology, John Wiley and Sons, 1962*)

toward thinning, or size reduction, of integumentary sclerifications seen in various fish lineages is probably associated with weight reduction and increased locomotor efficiency.

Tetrapods. Scaled integuments are absent from modern amphibians, although nonoverlapping dermal ossifications are seen in some apodans. It seems probable that the absence of scales in other modern species is associated with the secondary utilization of the integument as a respiratory surface. While direct evidence for the presence of patterned arrangements of dermal ossification is available from a variety of Paleozoic amphibian genera, suggesting a scaled integument, it is uncertain if the overlying epidermis was heavily cornified, a feature which might have been expected in terrestrial forms. Certain chemical features of the matrix surrounding the fossilized remains of the lower Triassic *Uranocentrodon* have been interpreted as indicating a localized decomposition of sulfur-containing epidermal proteins.

While dermal ossifications are known from many early reptiles (captorhinomorphs), it cannot be stated with certainty whether these were retained from their amphibian ancestors or independently evolved. This doubt springs from the fact that in the well-documented lineages of captorhinomorph to therapsid to mammal, there is early reduction and loss of dermal ossifications (suggesting a loss of a scaled integument), and yet numerous fossil and extant mammals, for example, ground sloths, glyptodonts, and armadillos, have very well-developed dermal ossifications over much of the body. The characteristic mammalian hairs did not evolve from scales, but their patterned arrangement has been interpreted as indicating an evolutionary origin associated with developmental fields surrounding scales in the ancestral reptiles. In general, it would appear that the evolutionary history of the mammalian integument involved early loss of scales, concomitant with increasing reliance on initially a thickened epidermis and later a pelage, for both physiological and mechanical protection. *See* HAIR.

In all other reptilian lineages, data from fossils and from study of living species suggest a ubiquity of scaled integuments, with or without dermal ossifications, but always with elaboration of epidermal tissues. This fact has led to the formulation of the generic terms "reptilian" and "epidermal" scales, both of which are best avoided. The distribution of keratinaceous proteins is highly variable in both reptilian and avian scales. Because modern crocodilian scales exactly resemble the scutate (overlapping) units on a bird's leg, an excellent case can be made for the direct evolution of feathers from archosaurian scales. However, the reticulate scales on the plantar surface of the avian foot do not overlap, and their epidermal cells synthesize alpha rather than beta keratin.

Birds, of course, totally lack the dermal ossifications which reached such impressive proportions in their Mesozoic archosaurian cousins such as *Ankylosaurus* and *Stegosaurus*. Scaled integuments prompted the subclass name for lepidosaurs (*lepidos* is Greek for flake or scale)—the tuatara, lizards, and snakes, the last two constituting the order Squamata (*squama* is Latin for scale).

Lepidosaurian scales may or may not possess dermal ossifications. Among extant reptiles the extraordinary development of the dermal skeleton (obviously associated with epidermal patterning) in turtles forms the characteristic carapace, a remarkable parallel with the mammalian glyptodonts, but the carapace is reduced or lost (along with visible scalation) in a variety of aquatic forms.

P. F. A. Maderson

Bibliography. L. B. Halstead and R. Hill, *Vertebrate Hard Tissues*, 1974; E. Jarvik, *Basic Structure and Evolution of Vertebrates*, 2 vols., 1980.

Scandium

A chemical element, Sc, atomic number 21, atomic weight 44.956. The only naturally occurring isotope is ^{45}Sc. The electronic configuration of the ground-state, gaseous atom consists of the argon rare-gas core plus three more electrons in the 3*d*14*s*2 levels. It has an unfilled inner shell (only one 3*d* electron) and is the first transition metal. It is one of the elements of the rare-earth group. *See* PERIODIC TABLE; RARE-EARTH ELEMENTS; TRANSITION ELEMENTS.

1																	18
1 H	2											13	14	15	16	17	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

The principal raw materials for the commercial production of scandium are uranium and tungsten tailings and slags from tin smelters or blast furnaces used in cast iron production. Wolframite (WO_3) concentrates contain 500–800 ppm scandium. *See* WOLFRAMITE.

Scandium is the least understood of the 3*d* metals. The major reason has been the unavailability of high-purity scandium metal, especially with respect to iron impurities. Many of the physical properties reported in the literature vary considerably, but the availability of electrotransport-purified scandium has allowed measurement of the intrinsic properties of scandium (see **table**).

Scandium increases the strength of aluminum. It also strengthens magnesium alloys when added to magnesium together with silver, cadmium, or yttrium. Scandium inhibits the oxidation of the light

Room-temperature properties of scandium metal (unless otherwise specified)

Property	Value
Atomic number	21
Atomic weight	44.9559 ($^{12}C = 12$)
Lattice constant (hcp, α–Sc), a_0	0.33088 nm
c_0	0.52680 nm
Density	2.989 g/cm^3
Metallic radius	0.16406 nm
Atomic volume	15.041 cm^3/mol
Transformation point	1337°C (2439°F)
Melting point	1541°C (2806°F)
Boiling point	2836°C (5137°F)
Heat capacity	25.51 J/mol K
Standard entropy, $S^\circ_{298.15}$	34.78 J/mol K
Heat of transformation	4.01 kJ/mol
Heat of fusion	14.10 kJ/mol
Heat of sublimation (at 298 K)	377.8 kJ/mol
Debye temperature (at 0 K)	345.3 K
Electronic specific heat constant	10.334 mJ/mol K^2
Magnetic susceptibility, χ_A^{298} (*a*)	297.6×10^{-6} emu/mol
χ_A^{298} (*c*)	288.6×10^{-6} emu/mol
Electrical resistivity, ρ_a^{300}	70.90 μohm-cm
ρ_c^{300}	26.88 μohm-cm
Thermal expansion, $\alpha_{a,i}$	7.55×10^{-6}
$\alpha_{c,i}$	15.68×10^{-6}
Isothermal compressibility	17.8×10^{-12} m^2/N
Bulk modulus	5.67×10^{10} N/m^2
Young's modulus	7.52×10^{10} N/m^2
Shear modulus	2.94×10^{10} N/m^2
Poisson's ratio	0.279

rare earths and, if added along with molybdenum, inhibits the corrosion of zirconium alloys in high-pressure steam. The addition of ScC to TiC has been reported to form the second-hardest material known. Sc_2O_3 can be used in many other oxides to improve electrical conductivity, resistance to thermal shock, stability, and density. Scandium is used in the preparation of the laser material $Gd_3ScGa_4O_{12}$, gadolinium scandium gallium garnet (GSGG). This garnet when doped with both Cr^{3+} and Nd^{3+} ions is said to be $3^1/_2$ times as efficient as the widely used Nd^{3+}-doped yttrium aluminum garnet (YAG:Nd^{3+}) laser. Ferrites and garnets containing scandium are used in switches in computers; in magnetically controlled switches that modulate light passing through the garnet; and in microwave equipment. Scandium is used in high-intensity lights. Scandium iodide is added because of its broad emission spectrum. Bulbs with mercury, NaI, and ScI_3 produce a highly efficient light output of a color close to sunlight. This is especially important when televising presentations indoors or at night. When used with night displays, the bulbs give a natural daylight appearance. Scandium metal has been used as a neutron filter. It allows 2-keV neutrons to pass through, but stops other neutrons that have higher or lower energies. *See* YTTRIUM.

Jennings Capellen; Karl A. Gschneidner, Jr.

Bibliography. E. W. Abel et al., *Comprehensive Organometallic Chemistry II: Scandium, Yttrium, Lanthanides and Actinides, and Titanium, Zirconium, and Hafnium*, Pergamon Press, 1995; C. T. Horovitz et al., *Scandium, Its Occurrence, Chemistry, Physics, Metallurgy, Biology and Technology*, Academic Press, 1975; D. R. Lide, *CRC Handbook Chemistry and Physics*, 85th ed., CRC Press, 2004.

Scanning electron microscope

An electron microscope that builds up its image as a time sequence of points in a manner similar to that employed in television. While the idea of the scanning electron microscope (SEM) dates back to the 1930s, a great deal of the pioneering development of the instrument took place at Cambridge University, England, during the years following World War II. The SEM became commercially available in 1966 and has rapidly taken its place as a useful addition to the battery of electron optical instruments used in both biological and physical research. A part of the attractiveness of the SEM lies in its ability to utilize both analytic and subjective or intuitive channels in attempts to understand the microscopic world.

Imaging method. The imaging method of the scanning electron microscope allows separation of the two functions of a microscope, localization and information transfer. The SEM utilizes a very fine probing beam of electrons which sweeps over the specimen and at each instant excites the material of the specimen to emit a variety of radiations (**Fig. 1**). The signal, which is proportional to the amount of radiation leaving an individual point of the specimen at any instant, can be used to modulate the brightness of the beam of the display cathode-ray tube as it rests on the corresponding point of the image. When the radiation from the point on the specimen is high, the intensity of the point on the display cathode-ray tube is also high. If the probing beam is then moved to an adjacent spot on the specimen, information from that point can be used to modulate the corresponding point on the display cathode-ray

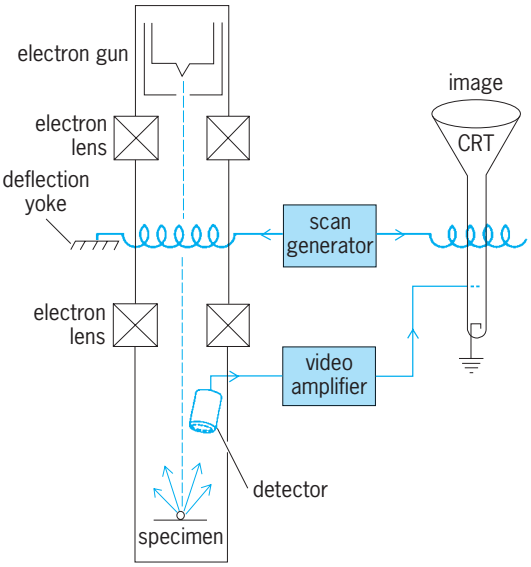


Fig. 1. Schematic diagram of scanning electron microscope showing twin-electron-beam imaging system and some of the radiations used for the information signal.

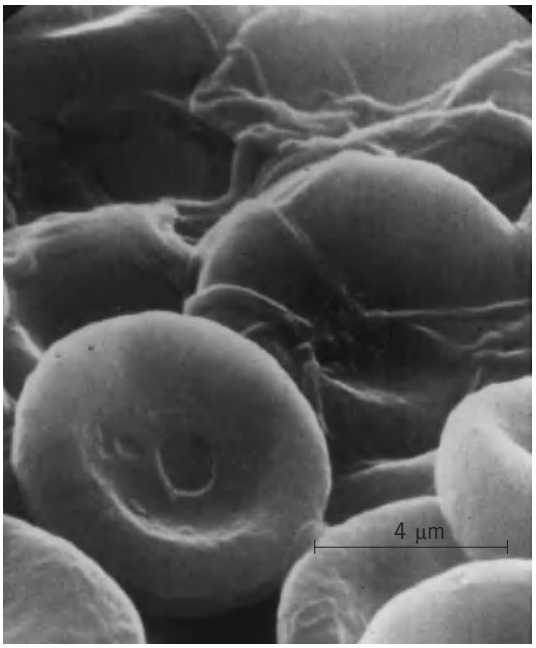


Fig. 2. Human blood clot. Red blood cells are held in fibrin network. (*From L. W. McDonald and T. L. Hayes, Correlation of scanning electron microscope and light microscope images of individual cells in human blood and blood clots, Exp. Molec. Pathol., 10:186–198, 1969*)

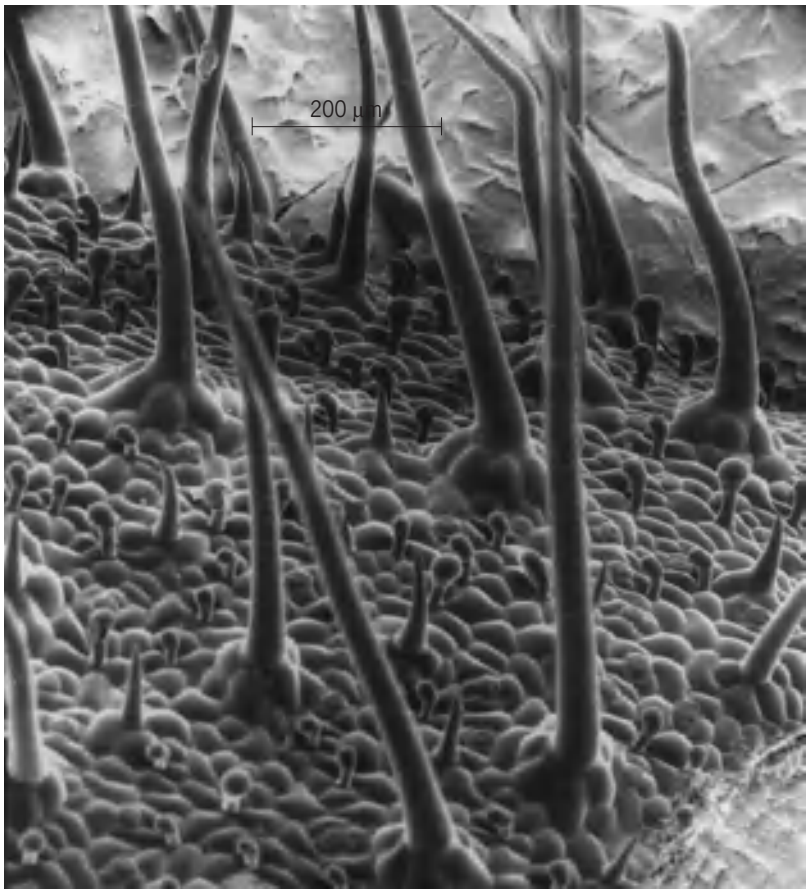


Fig. 3. Leaf surface viewed while held at −238°F (−150°C) in the scanning electron microscope. Specimen is surrounded by ice field and is still hydrated. (*From T. L. Hayes and G. Koch, Some problems associated with low temperature micromanipulation in the SEM, Proceedings of the Scanning Electron Microscopy Symposium, pp. 35–42, 1975*)

tube. In a similar fashion a third and fourth point can be added in sequence, and so on until all parts of the specimen have been covered in a regular array. The one-to-one correspondence that is necessary for the formation of an image occurs because a single position of the probing beam corresponds to a single position of the display cathode-ray-tube beam at each instant in time. In practice, the points follow one another with great rapidity so that the image of each point becomes an image of a line, and the line in turn can move down the screen so rapidly that the human eye sees a complete image as in television. The image can also be recorded in its entirety by allowing the point-by-point information to build up in sequence on a photographic film.

Both high resolution and a variety of kinds of information can be achieved with this system. Albert Crewe developed ultrahigh resolution scanning transmission electron microscopes (STEM) capable of resolving single atoms. Secondary electrons, backscatter electrons, characteristic x-rays, visible light (cathodoluminescence) and specimen-induced current have been used by investigators as SEM signals to provide information-rich images that relate to the size, shape, and chemical and electrical properties of the specimen.

Sensory codes. In addition to a very powerful array of objective analytic information, the SEM offers possibilities for use as an instrument to extend the human senses. Operated in the secondary electron detection mode, the SEM mimics some of the codes that people use to assess depth and shape in the large macroscopic world of everyday life. Since this ability to experience the microsystem through the use of an SEM permits the application of synthetic, intuitive processing, perhaps subjective channels can be added to enhance understanding of the classical disciplines of physics, chemistry, mathematics, or biology. Some of the fascination of the SEM may be related to its ability to provide intuitive contact with the microspecimen (**Fig. 2**).

Gun intensity. Since the resolution of the SEM is controlled in part by the brightness of the electron gun, experimental designs which increase gun intensity have been pursued for some years. Crewe, using a field emission source to increase the gun intensity, has produced instruments with resolutions well below 0.5 nanometer. Bimacromolecular staining as developed by Michael Beer has enhanced the contrast differences available and has resulted in a specific location of heavy-atom components within a macromolecule. In addition to the field emission gun, lanthanum hexaboride guns developed by Alec Broers provide greater brightness than the standard heated filaments.

Specimen preparation. As with all microscopy research, it is very often the preparative methodology that determines the success or failure of the research. A specific requirement of scanning electron microscope preparation is that the material be dried. In conventional electron microscopy or light microscopy, the specimen is carried through a

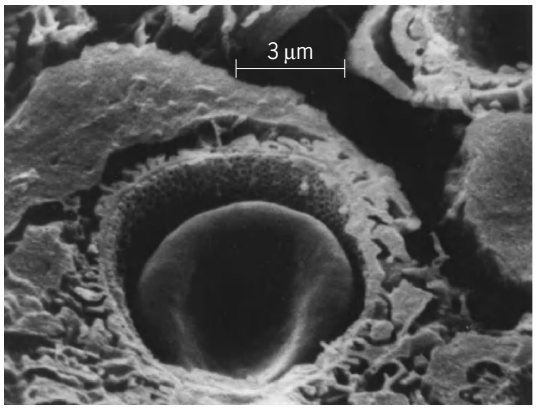


Fig. 4. Red blood cell in a capillary of the kidney. Ethanol cryofracture technique was followed by critical-point drying. (*From W. J. Humphreys, B. O. Spurdock, and J. S. Johnson, Critical point drying of ethanol-infiltrated, cryofractured biological specimens for scanning electron microscopy, Proceedings of the Scanning Electron Microscopy Symposium, pp. 275–282, 1974*)

series of liquid stains, fixatives, dehydrating agents, and embedding materials, but is never dried. In scanning microscopy, on the other hand, the material must be dried, and this preparative step must be done very carefully to minimize surface tension effects. The two methods most often used—critical-point drying and freeze drying—have yielded excellent results even for very fragile biological tissue. Alternatively, the drying step can be avoided completely by observing the specimen while it is still frozen (**Fig. 3**). Interior structure can be revealed by using techniques in which the specimen is broken at a low temperature (**Fig. 4**). *See* MICROTECHNIQUE.

Channels of information transfer. In microscopy there has always been a very useful aspect of image perception associated with the observer's intuitive or subjective contact in "seeing" the specimen. It is this kind of contact that has been particularly valuable when the specimen under study becomes extremely complex and difficult to analyze by strictly objective approaches. The SEM provides both objective and experiential channels of information transfer. In summary, the SEM is an excellent tool for analysis. It also provides valuable channels for human contact with the microworld through extension of the human sense. *See* ELECTRON MICROSCOPE; MICROSCOPE.

Thomas L. Hayes

Bibliography. W. R. Bottoms, *Scanning Electron Microscopy and X-Ray Microanalysis: A Text for Biologists, Materials Scientists, and Geologists*, 2d ed., 1992; P. Echlin and D. E. Newbury, *Scanning Electron Microscopy*, 1977; J. I. Goldstein and H. Yakowitz (eds.), *Practical Scanning Electron Microscopy: Electron and Ion Microprobe Analysis*, 1975; O. Johari and R. P. Becker (eds.), *Scanning Electron Microscopy*, 14th ed., 1981; J. Ohnsorge and R. Holm, *Scanning Electron Microscopy*, 2d ed., 1978; L. Reimer, *Image Formation in Low-Voltage Scanning Electron Microscopy*, 1993.

Scanning tunneling microscope

A device for imaging the surface of conductors and semiconductors with atomic resolution. Due to its relatively simple construction and operation and its ability to achieve atomic resolution with relative ease, the scanning tunneling microscope (STM) has gained worldwide acceptance by scientists studying surface phenomena.

Principles of operation. The operating principles of the scanning tunneling microscope are illustrated in **Fig. 1**. A wire made of tungsten, gold, or some other noble metal is etched electrochemically at one end down to a fine tip whose radius is 0.1 micrometer, or less; it is then attached onto a piezoelectrically scanned (x,y,z) stage. Under a powerful microscope, such as a transmission electron microscope, the tip of the wire typically looks as illustrated in Fig. 1, with one atom always protruding farther from the surface. The piezoelectric tripod can be scanned in x, y, or z over distances of up to 100 μm with a resolution far better than 0.1 nanometer (approximately one-half the diameter of typical atoms) by applying a voltage across its x, y, or z electrode, respectively. A voltage V in the range of 10 mV is applied between the wire tip and the conducting sample to be imaged, and the piezoelectric z-scanner is used to move the tip gently toward the sample. As the end atom on the tip approaches to within a few atomic diameters of a sample atom, electrons can tunnel from the tip to the sample (or vice versa) through the potential barrier imposed by the work functions of the tip and sample, thereby generating a current, which is typically of the order of nanoamperes. *See* ELECTRON MICROSCOPE; WORK FUNCTION (THERMODYNAMICS).

The tunnel current I_t increases exponentially with decreasing spacing z between tip and sample atom.

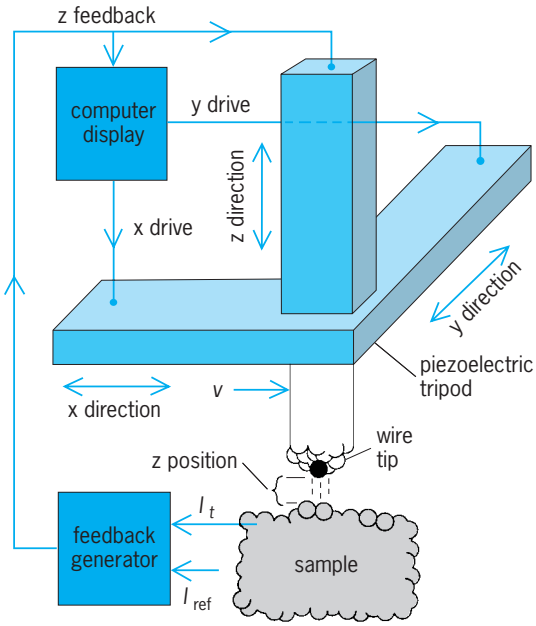


Fig. 1. Schematic diagram of the scanning tunneling microscope showing the configuration of the piezoelectric tripod, wire tip, and sample.

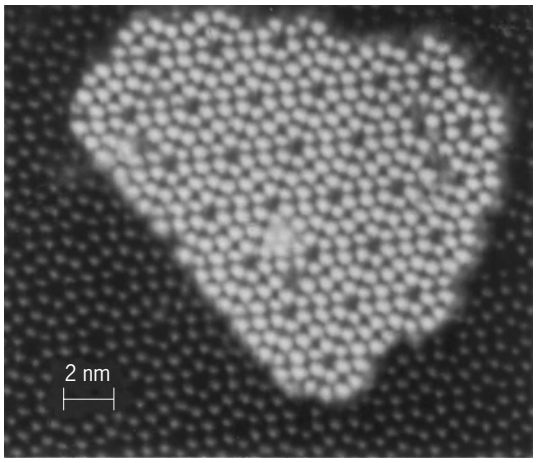


Fig. 2. Silicon ⟨111⟩ surface as imaged with the scanning tunneling microscope. The image extends over a lateral area whose dimensions are 23 × 20 nm. The image shows the (7 × 7) reconstruction of the unit cell on the crystal surface together with the registration of an epitaxially grown single layer. (*Courtesy of U. Kohler, J. E. Demuth, and R. J. Hamers*)

For small applied bias voltages v, the tunnel current is given by the equation below, where ϕ is the mean

$$I_t = f(v) \exp -(A\phi^{1/2}z)$$

of the work functions of the tip and sample, A is a constant equal to $(4\pi/h)(2m)^{1/2}$, m is the electronic mass, h is Planck's constant, and $F(v)$ is the overlap integral of the local density of states of tip and sample taken over the energies between the Fermi levels of the tip and the sample. For a typical tip–sample combination, the tunnel current I_t decreases by a factor of 10 for every 0.1 nm increase in spacing z. For this reason, a negligible contribution to the tunnel current I_t is provided by electron tunneling from other atoms on the tip since the other atoms lie farther from the sample than the atom that protrudes from the tip.

In the scanning tunnel microscope, the tunnel current I_t is compared with a constant reference current (I_{ref}); the error signal so generated is used to move the z-piezoelectric scanner up or down (thereby increasing or decreasing z) in order to maintain a constant tunnel current as the tip is raster-scanned in x and y to record an image. The tunneling image is usually recorded on a computer as variations in the z position (measured through the z-piezoelectric voltage) corresponding to each (x,y) coordinate of the tip. For metallic surfaces such as gold, the local density of states is a weak function of the applied bias voltage v, and the overlap integral $f(v)$ has no structure; it varies linearly with voltage, yielding a voltage-independent tunneling conductance. In these cases the variations in z closely approximate the atomic topography of the surface. However, for more complex surfaces, $f(v)$ is no longer a linear function of v and furthermore can vary from one (x,y) location on the sample to the next, depending on the particular atom being probed. Similarly, the mean work function ϕ could also vary as a function of (x,y) on the sample. It can be shown that local-density-of-states spectroscopy can be performed on the sample by recording $(d \ln I_t)/(d \ln V)$ as a function of V. If local barrier height is desired, $A\phi^{1/2}$ can be measured by recording $(d \ln I_t)/dz$. Such measurements are useful in differentiating atomic features on the surface being imaged. *See* COMPUTER GRAPHICS; PIEZOELECTRICITY; TUNNELING IN SOLIDS.

Applications and extensions. One application of the scanning tunneling microscope is in the mapping of the atomic reconstructions of crystal surfaces. The image in **Fig. 2** demonstrates the potential of the scanning tunneling microscope for studying nucleation and growth of thin films. The brighter region in the center of the image is an epitaxially grown single layer of silicon ⟨111⟩ that shows the near-perfect registration with the substrate.

Other areas of application of the scanning tunneling microscope are in the study of electrochemical processes and biological imaging. It is also being used for surface modifications on an atomic scale.

The success of the scanning tunneling microscope has led to the development of many novel scanned probe microscopes that use the same piezoelectric scanning and feedback principles but rely on different mechanisms of tip–sample interaction to generate the error signal that controls the tip–sample spacing and hence the image. One scanning probe microscope that has been particularly successful is the atomic force microscope. In this microscope (**Fig. 3**) the interaction mechanism sensed is the repulsive force between the charge clouds of the tip atom and the sample atoms. Because this interaction does not rely on a tunneling current, the atomic force microscope can image both conducting and insulating surfaces with atomic resolution. In the early versions of the atomic force microscope, a diamond tip was attached onto a gold foil, and the repulsive force between tip and sample was detected through

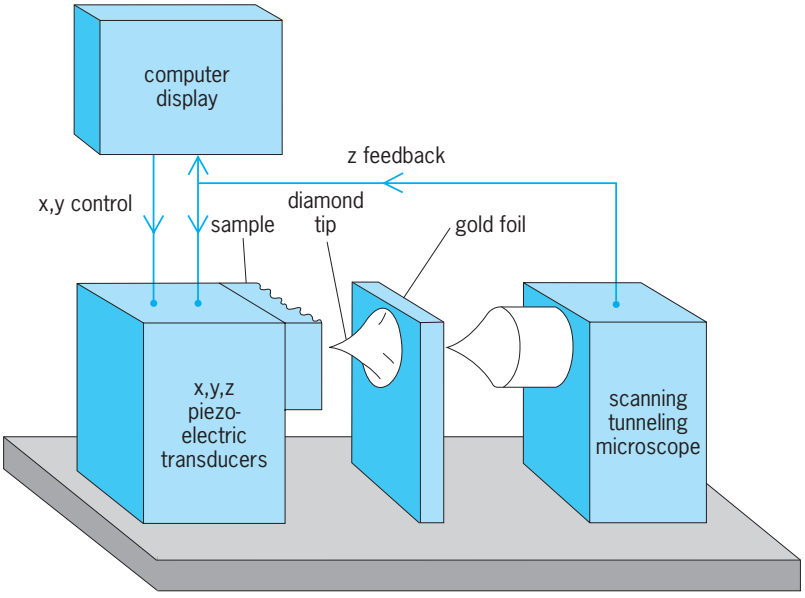


Fig. 3. Schematic diagram of the atomic force microscope.

the deflection of the gold foil, which in turn was sensed by using a scanning tunneling microscope. If the gold foil has a spring constant of 10 newtons/m, a force of 10^{-9} N will result in a deflection of 0.1 nm, which is easily detectable with a tunneling microscope. Later experiments have used silicon micromechanical cantilevers with integrated silicon tips and laser probes for sensing the deflection. Typical tracking forces are in the range of 10^{-7} N for instruments operating in air and 10^{-9} N under water. Applications include the imaging of the atomic arrangement of insulating crystal surfaces and the high-resolution imaging of biological surfaces and processes. Extensions of the atomic force microscope are the magnetic force microscope, which uses a magnetic tip to image magnetic surfaces, and the electrostatic force microscope, which measures the local dielectric polarizability of a sample through the attractive force caused by applying a voltage between the tip and the sample. *See* MICROSCOPE. H. K. Wickramasinghe

Bibliography. G. Binnig and H. Rohrer, The scanning tunneling microscope, *Sci. Amer.*, 253(8):50–56, 1985; G. Binnig and H. Rohrer, Scanning tunneling microscopy: From birth to adolescence, *Rev. Mod. Phys.*, 59(3):651–625, 1987; C. J. Chen, *Introduction to Scanning Tunneling Microscopy*, 1993; P. K. Hansma et al., Scanning tunneling microscopy and atomic force microscopy: Applications to biology and technology, *Science*, 242(4876):209–216, 1988; U. Kohler, J. E. Demuth, and R. J. Hamers, Scanning tunneling microscopy study of low-temperature epitaxial growth of silicon on Si(111)-(7 × 7), *J. Vac. Sci. Technol.*, A7(4):2860–2867, 1989; C. F. Quate, Vacuum tunneling: A new technique for microscopy, *Phys. Today*, 39(8):26–33, 1986; H. K. Wickramasinghe, Scanned probe microscopes, *Sci. Amer.*, 260(10):98–105, 1989.

Scaphopoda

A class in the phylum Mollusca, comprising two orders, Gadilida and Dentaliida. The class is wholly marine; it is common and probably most diverse in the deep sea, and poorly represented in brackish, littoral, or estuarine habitats. Scaphopods are characterized by a tapering tubular shell with two apertures: the large ventral aperture that the foot extends out of and a small dorsal aperture. The tapering shells found in the Dentaliida give the class its common names tusk or tooth shells. Adult scaphopods range in length from about 3 to 150 mm. The shell consists of aragonite arranged in three or four layers.

Classification. Systematically the class is related to bivalves, and probably arose from a member of the extinct molluscan class, the Rostroconchia. The two orders can be separated on the basis of shell structure. Typically gadilids have the widest part of the shell more toward the middle of the animal, while in the dentaliids the widest part of the shell is the ventral aperture. Gadilid shells are typically highly polished, while dentaliid shells lack this luster. In addition, there are significant internal morphological differences between the two orders.

Anatomical features. Scaphopods are bilaterally symmetrical and surrounded by the mantle, which forms a tube. The body is suspended in the tube from the dorsal-anterior part of the shell. The largest part of the mantle cavity is adjacent to the ventral aperture and contains the foot, head, and feeding tentacles, called captacula. Each captaculum consist of a long extendable stalk and a terminal bulb. The captacular bulb is covered with cilia, and the stalk may have a ciliated band or tufts of cilia along it. Captacula are hollow and have longitudinal muscles running from their base to the bulb.

The mantle cavity extends to the dorsal shell aperture along the ventral part of the body. Scaphopods lack the typical molluscan gill, the ctenidium. Respiration occurs through the mantle wall and is probably facilitated by the presence of large ciliated ridges found in the middle of the mantle cavity.

The foot is large and robust and capable of significant extension from the shell. In gadilids, where the foot is extended hydrostatically and retracted by inversion, the animals are often capable of rapid prolonged burrowing. The gadilid foot ends in a terminal disk. In dentaliids, where the foot is extended and retracted by the use of a muscular hydrostat and flexion, the animals move much more slowly. The dentaliid foot has a conical tip, and lateral lobes a bit proximal to the end.

The head, which is reduced and incapable of extension from the shell, bears a pair of lateral ridges and a proboscis which is tipped by the mouth surrounded by lips. The captacula arise from the lateral ridges; generally several hundred captacula are present. Scaphopods lack eyes and the specialized molluscan sensory organ, the osphradium.

The body is elongated dorsally and terminates internally near the dorsal shell aperture; however, the gut curves back on itself, and the anus is normally found at about the midpoint of the body and empties into the mantle cavity. The radula is mineralized and functions either to crush prey or to ratchet them into the rest of the gut through the esophagus. A single unmineralized jaw is anterior to the radula in the buccal region. There are five radular teeth per row. The stomach is capacious and is connected to one (in gadilids) or two (in dentaliids) digestive glands. The intestine is coiled, and the rectum is just proximal to the anus. There are rectal glands of unknown function near the anus.

The large dorsal retractor muscles pass laterally and ventrally to the gut. These originate dorsally near the aperture and extend to the base of the muscular dentaliid foot or to the tip of the hollow gadilid foot. They retract the foot or, when the foot is extended into sediments, pull the animal into the sediments.

The nervous system is relatively large and is built on the basic molluscan pattern. The largest ganglia are the paired cerebral and pedal ganglia; the latter bear statocysts. Other small ganglia are present, as are major nerves to the main body regions. Each captaculum also contains a small sensory ganglion.

The circulatory system consists of a series of blood

sinuses and channels. There is no heart that is homologous to other molluscan hearts, although contractile tissues are found near where a heart should be. These contractile elements cause fluid movement in the central blood sinus but are incapable of moving the colorless blood to the periphery of the animal. Most blood movement is due to the movement of the musculature of the foot. There are paired excretory organs, but no evidence of the urine formation by ultrafiltration that is typical of other mollusks.

The gonads are positioned dorsally and empty into the mantle cavity; sexes are separate. Early embryonic development and complete larval development have been observed in the dentaliids, but little is known about any aspect of embryonic or larval development in the gadilids.

Feeding mechanism. Scaphopods eat foraminiferans, and other small shelled organisms such as bivalves or shelled eggs. Dentaliids can also ingest sediment. Prey are captured by the use of the captacula, which have a chemical adhesive system to attach to prey. The foot is used to burrow and construct a feeding cavity. The captacula move out into the feeding cavity through the ventral aperture by the motion of the cilia covering the bulb. The cilia pull the captacular stalk along behind.

When a captaculum encounters an acceptable food item, it attaches to that item with a duo-gland adhesive system. After several other captacula attach, the longitudinal muscles within them contract, pulling the food item out of the feeding cavity and into the mantle cavity. Food is manipulated internally by the captacula and the lips. If the food item is acceptable, it is ingested and held in the proboscis prior to radular maceration. Up to 300 foraminiferan prey have been counted in the proboscis of one species of scaphopods.

Habitat. Scaphopods live in unconsolidated sediments, and most members of any population are typically completely buried at any one time. However, scaphopods back up to the surface and extend the dorsal shell apex to release eggs or sperm, and presumably to flush the mantle cavity with fresh water. They are vulnerable to predation during these periods, which are of limited duration.

In some localities along the Canadian Pacific coast, scaphopods are abundant and possibly constitute the major predators on organisms living in the sediments of the fiord bottoms. Scaphopods, in turn, are eaten by fishes, particularly chimerids, and crabs. *See* MOLLUSCA.

Ronald Shimek

Bibliography. F. Harrison and A. J. Kohn (eds.), *Mollusca II: Microscopic Anatomy of the Invertebrates*, vol. 6B, 1996; E. N. Kozloff, *Invertebrates*, 1990, reprint 1997; R. L. Shimek, Diet and habitat utilization in a Northeastern Pacific Ocean scaphopod assemblage, *Amer. Malacol. Bull.*, 7:147–169, 1990.

Scapolite

An aluminosilicate mineral containing sodium (Na), calcium (Ca), chlorine (Cl), and carbonate (CO_3). It is commonly found as light-colored, translucent tetragonal prisms (**Fig. 1**). Scapolite is normally white, but many other colors are known, including some used as semiprecious gems resembling amethyst and citrine. It has good 90° cleavage parallel to its

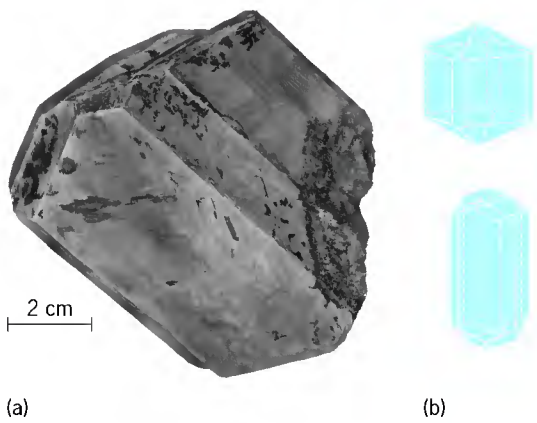


Fig. 1. Scapolite. (a) Crystal taken from Pierrepont, New York (*specimen from Department of Geology, Bryn Mawr College*). (*b*) Crystal habits (*after C. S. Hurlbut, Jr., Dana's Manual of Mineralogy, 17th ed., John Wiley and Sons, 1959*)

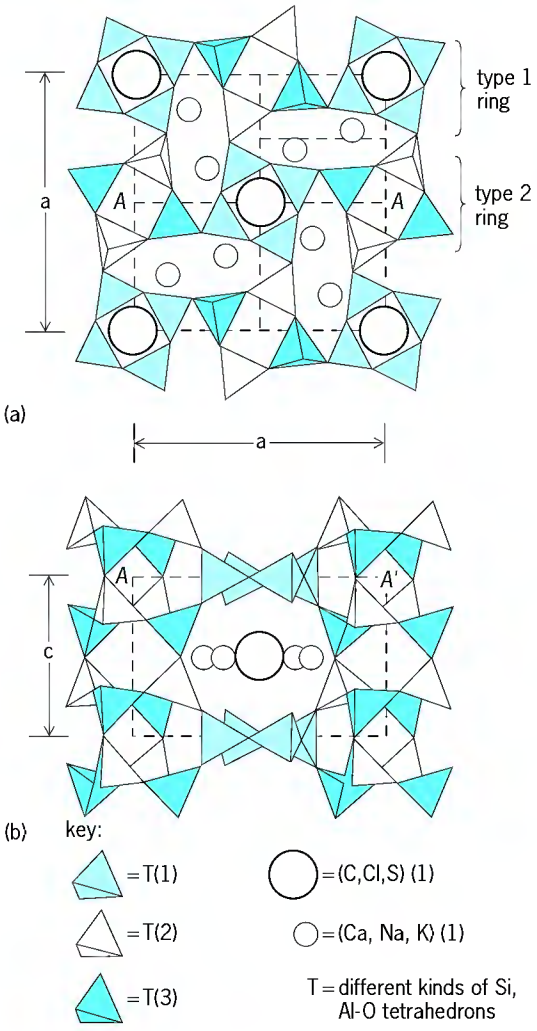


Fig. 2. Crystal structure of scapolite, (a) perpendicular to and (*b*) parallel to the tetragonal *c* axis. The dimensions designated a and c are lengths of the unit cell edges. (*After S. B. Lin and B. J. Burley, The crystal structure of an intermediate scapolite-wernerite, Tschermaks Min. Petr. Mitt., 21:196–215, 1974*)

c axis, and has a Mohs hardness of 5–6.

The formula of scapolite is $(Na,Ca)_4(Al,Si)_6Si_6O_{24}(Cl,CO_3,SO_4)$. Natural scapolites have compositions which vary from $Ca/(Ca + Na) = 0.1–0.9$ and $Cl/(Cl + CO_3 + SO_4) = 0.0–0.7$. In nature significant amounts of K and SO_4 substitute for Na and CO_3 (up to 2% K and 4% SO_4). Crystal-structure studies indicate that scapolite crystallizes in the space group $P4_2/n$. The structure (**Fig. 2**) consists of a framework of Al-O and Si-O tetrahedra, with oval cavities which hold Na and Ca and large cavities in which the Cl, CO_3 and SO_4 occur. Optically scapolite is uniaxial negative, with moderate to high birefringence. Unit cell dimensions, density, and refractive indices all increase with Ca content.

Scapolite is a common mineral in metamorphic rocks, particularly in those which contain calcite. It is found in marbles, gneisses, skarns, and schists. Feldspar, quartz, calcite, pyroxene, amphibole, and other Na-Ca minerals are common associates. Scapolite probably forms about 0.1% of the Earth's upper crust. It is commonly found as inclusions in igneous rocks derived from deep within the Earth's crust, and probably makes up several percent of the lower crust. Experiments have shown that scapolite is stable up to 2700°F (1500°C) at 20 kbar (2 gigapascals) pressure, and that the range of compositions can be much greater at those pressures. Assemblages of scapolite, plagioclase, calcite, and halite simulating those found in nature have been synthesized at 750°C (1400°F) and 4 kbar (400 MPa). Many scapolitic rocks undoubtedly formed at much lower temperatures and pressures, under conditions not yet experimentally reproduced. *See* SILICATE MINERALS.

David E. Ellis

Bibliography. D. E. Ellis, Stability and phase equilibria of chloride and carbonate bearing scapolites at 750°C and 4000 bar, *Geochim. Cosmochim. Acta*, 42:1271–1281, 1978; J. R. Goldsmith, Scapolites, granulites, and volatiles in the lower crust, *Geol. Soc. Amer. Bull.*, 87:161–168, 1976; S. B. Lin and B. J. Burley, The crystal structure of an intermediate scapolite-wernerite, *Tschermaks Min. Petr. Mitt.*, 21:196–215, 1974; D. M. Shaw, The geochemistry of scapolite, *Geochim. Cosmochim. Acta*, 1:218–285, 1960.

Scarlet fever

An acute contagious disease that is a consequence of infection with *Streptococcus pyogenes* (group A streptococci). It most often accompanies pharyngeal (throat) infections with this organism but is occasionally associated with wound infection or septicemia. Scarlet fever is characterized by the appearance, about 2 days after development of pharyngitis, of a red rash that blanches under pressure and has a sandpaper texture. Usually the rash appears first on the trunk and neck and spreads to the extremities. The rash fades after a week, with desquamation, or peeling, generally occurring during convalescence. Like streptococcal pharyngitis without such a rash, the disease is usually self-limiting, although severe forms are occasionally seen with high fever and systemic toxicity. Appropriate antibiotic therapy is recommended to prevent the onset in susceptible individuals of rheumatic fever and rheumatic heart disease.

Scarlet fever occurs after infection with streptococci that produce the extracellular products, erythrogenic toxins A, B, and C (also known as streptococcal pyrogenic exotoxins). The genes for the toxins are carried by lysogenic bacteriophages that, on infection of streptococci, can convert them to toxin-producing strains. The toxins are believed to be responsible for the rash of scarlet fever through interaction with the host defense system. Formerly, scarlet fever was associated with toxin A–producing streptococcal strains. Then there was a shift to a prevalence of toxin B and toxin C producers. A resurgence of toxin A–producing streptococci, however, is again being noted. The incidence of scarlet fever is declining, but this decline is apparently not a consequence of a reduction in streptococcal infections. The decline remains unexplained. *See* MEDICAL BACTERIOLOGY; RHEUMATIC FEVER; STREPTOCOCCUS.

Ernest D. Gray

Bibliography. G. L. Mandell, G. Douglas, Jr., and J. E. Bennett (eds.), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 5th ed., 2000; L. W. Wannamaker, Streptococcal toxins, *Rev. Infect. Dis.*, 5(suppl. 14):S723–S732, 1983; G. P. Youmans, P. Y. Paterson, and H. M. Sommers (eds.), *Biologic and Clinical Basis of Infectious Diseases*, 1985.

Scattering experiments (atoms and molecules)

Experiments in which a beam of incident electrons, atoms, or molecules is deflected by collisions with an atom or molecule. Such experiments provide tests of the theory of scattering as well as information about atomic and molecular forces. Scattering experiments can be designed to simulate conditions in planetary atmospheres, electrical discharges, gas lasers, fusion reactors, stars, and planetary nebulae. *See* ELECTRICAL CONDUCTION IN GASES; GAS DISCHARGE; LASER; NUCLEAR FUSION; PLANET; PLANETARY NEBULA; STAR.

Classification of collisions. A collision between two atomic systems is termed elastic if there is no transfer of energy between the internal motions of the two systems and the kinetic energy of their relative motion. A collision is termed inelastic or superelastic, depending on whether kinetic energy is given to or taken from the internal motion.

If electromagnetic radiation is emitted during the impact, the collision is termed radiative. In an ionizing collision, one or more electrons are set free by the colliding partners. Free electrons may also become bound to other particles in collisions by a process called recombination (with positive ions to form a neutral particle) or attachment (with atoms

or molecules to form a negatively charged ion). In a rearrangement collision, there is a redistribution of particles between the colliding systems during the impact. This rearrangement may involve the transfer of atoms or electrons from one system to the other. The latter process is called electron transfer or charge exchange.

In general, in any type of collision, scattering occurs, which causes the direction of relative motion of the two systems to be rotated to a new direction after the collision. More than two systems may also result from such an impact. A complete description of a collision event requires measurement of the directions, speeds and internal states of all the products. *See* COLLISION (PHYSICS).

Types of scattering experiments. There are two basic types of scattering experiments. The simpler (**Fig. 1***a*) involves passing a collimated beam of particles (electrons, atoms, molecules, or ions) through a dilute target gas (in a cell or a jet) and measuring the fraction of incident particles that are deflected into a certain angle relative to the incident beam direction. In the second method (Fig. 1*b*), a collimated beam of particles intersects a second beam. The scattering events are usually registered by measuring the deflection or internal-state change of the beam particles. *See* MOLECULAR BEAMS.

Collision cross sections. A scattering event is usually quantified in terms of a collision cross section, which may be simply thought of as an effective area surrounding one system into which the second system must pass in order for a specific process to occur. It is helpful to consider the simple case of a beam of electrons passing through a dilute target gas. An electron may be considered lost from the beam because of a change of either its direction or speed (energy). Through these processes, the beam current will be reduced by a fraction αx of the total beam in passing an infinitesimal distance x through the gas. Here, the quantity α is the product of the number of gas atoms or molecules per unit volume and $\sigma(E)$, the total effective cross section for collisions between electrons of energy E and the gas atoms or molecules.

Effective cross sections for different classes of collisions follow by introducing probabilities for a particular type of collision to occur. The product of this probability and the total effective cross section is

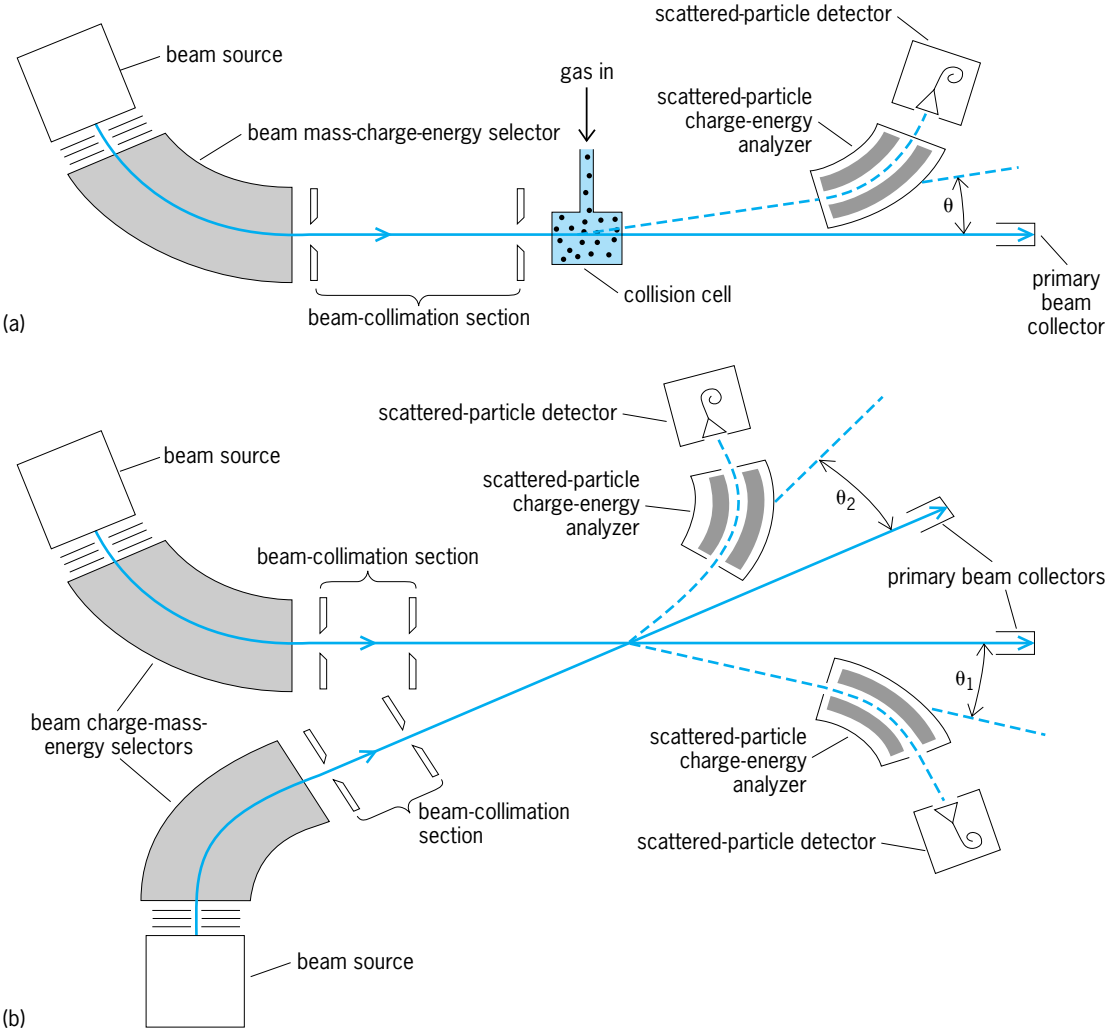


Fig. 1. Two different types of scattering experiments. (*a*) Beam-gas-target experiment. θ = scattering angle. (*b*) Colliding-beams experiment. θ_1, θ_2 = scattering angles.

then the effective total cross section for the particular type of collision.

Scattering in a particular type of collision is specified in terms of a differential cross section. The probability that, in a particular type of collision, the direction of motion of the electron is turned through a specified angle θ (called the scattering angle) into a specified solid angle is proportional to the corresponding differential scattering cross section.

Collision cross sections are most often expressed in square centimeters, and can be measured with appropriately designed experimental apparatus (Fig. 1). Depending on the type of collision process, that apparatus may measure the scattering angle, energy, charge, or mass of the scattered systems.

Scattering theory. For the simplest case, the scattering of a beam of structureless particles of specified mass and speed by a structureless scattering center, the differential cross section may be calculated exactly by using the quantum theory. In the special case where the Coulomb force fully describes the interaction, both the quantum and classical theory give the same exact value for the differential cross section at all values of the scattering angle. *See* NONRELATIVISTIC QUANTUM THEORY.

For scattering of systems with internal structure (for example, molecules, and their ions), no exact theoretical calculation of the cross section is possible. Methods of approximation specific to different types of collisions have been developed. The power of modern high-speed computers has greatly increased their scope and effectiveness, with scattering experiments serving as benchmarks. *See* ATOMIC STRUCTURE AND SPECTRA; SUPERCOMPUTER.

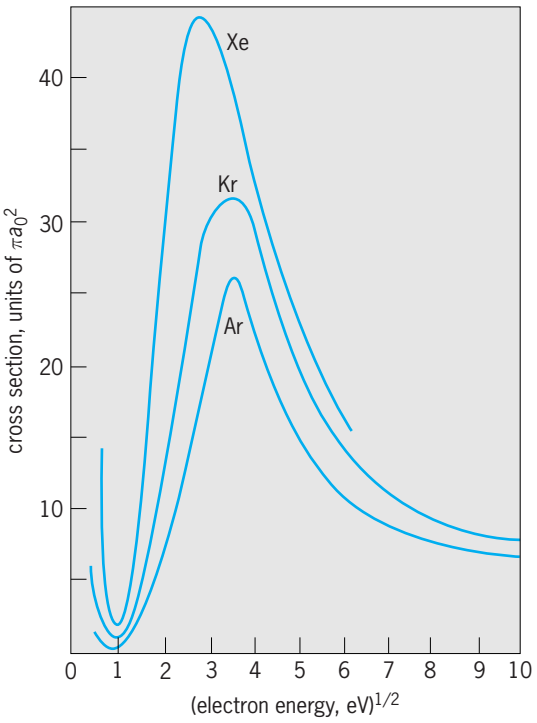


Fig. 2. Measured variation of total cross sections for collisions of slow electrons in argon (Ar), krypton (Kr), and xenon (Xe), as a function of the square root of the electron energy in electronvolts, which is proportional to electron velocity; $a_0 = 5.3 \times 10^{-9}$ cm.

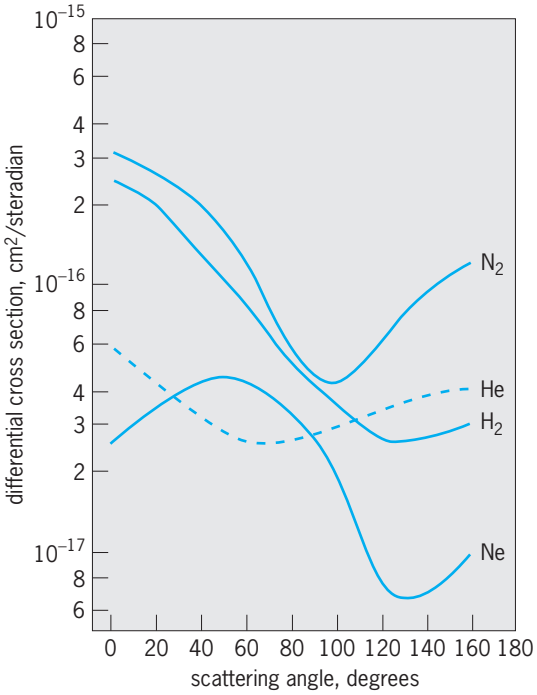


Fig. 3. Measured differential cross sections as a function of scattering angle for the elastic scattering of electrons from neon atoms (Ne), hydrogen molecules (H_2), helium atoms (He), and nitrogen molecules (N_2) at an energy of 10 eV.

Electron-atom collisions. The different types of collisions that can occur are elastic, inelastic involving excitation of discrete atomic states, ionization, radiative electron capture, and radiative collisions in which the electron is not captured. Radiative collisions are much less probable than nonradiative collisions, whereas cross sections for inelastic collisions are comparable to those for elastic collisions at electron energies that are large compared to the minimum (threshold) energy necessary to produce excitation or ionization.

Elastic scattering. The cross section for electron scattering from atoms varies with electron energy (**Fig. 2**). The measured differential cross sections in this energy range exhibit maxima and minima as a function of the scattering angle θ (**Fig. 3**). These effects are due to diffraction of the electrons by the scattering atoms. *See* ELECTRON DIFFRACTION.

If the energy homogeneity of the electron beam is sufficient, fine structure is observed in the variation of the total cross section with electron energy. For example, if the fraction of electrons in a beam that pass undeviated through an argon gas target is plotted against electron energy (**Fig. 4**), the transmitted fraction fluctuates rapidly over a narrow 0.5-eV energy range. This fine structure is due the existence of unstable atomic states called resonances, in which the electron is temporarily bound to an excited atom. After a characteristic lifetime, τ, of typically 10^{-14} s, the captured electron will be reemitted in some direction that is uncorrelated with its incident direction; that is, it will be scattered. Such scattering is

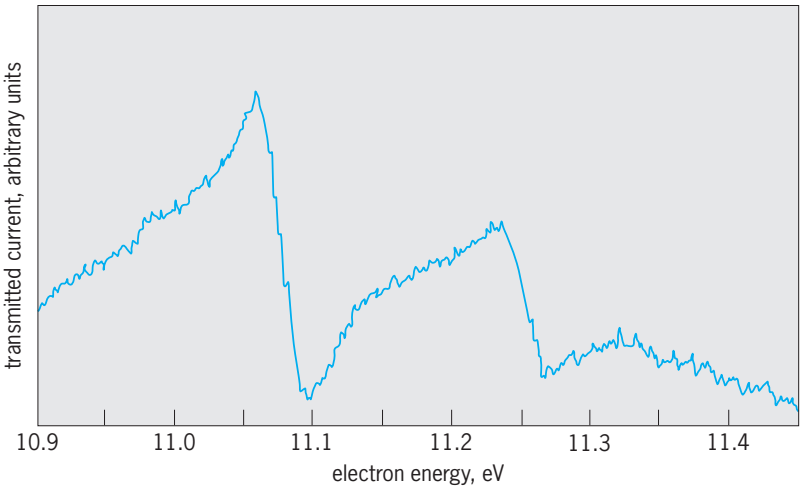


Fig. 4. Observed variation with electron energy of the intensity of an electron beam after transmission through argon gas; the greater the transmitted intensity, the smaller the scattering cross section.

probable only if the incident electron has an energy within ΔE, given by Eq. (1), of an unstable

$$\Delta E = \frac{h}{2\pi\tau} \tag{1}$$

excited state, where h is Planck's constant (4.14×10^{-15} eV-s). *See* RESONANCE (QUANTUM MECHANICS); UNCERTAINTY PRINCIPLE.

Inelastic scattering. A variety of experimental techniques have been used to study inelastic collisions. The excitation of atomic states has been investigated by both optical and energy-loss methods. In the optical method, measurement is made of the intensity of radiation emitted at a particular wavelength by gas atoms excited along the electron beam path. More detailed information about the scattering process has been obtained by detecting the scattered electron and emitted photon in coincidence. Simultaneous measurement of the electron scattering angle and light polarization in such an experiment allows the relative population of angular momentum states of the excited atom to be determined. *See* ANGULAR MOMENTUM.

The energy-loss method involves measurement of the fraction of electrons that have lost specific amounts of energy after passing through a gas target. The energy lost is the energy difference between the ground and excited atomic states. For energies not much greater than the excitation threshold, the scattered electron current exhibits fluctuations (**Fig. 5**) similar to that described above for resonances in elastic scattering, except that in this case the unstable excited state breaks up to leave the atom still excited, so that the ejected electron has a reduced energy.

Ionizing collisions have been studied primarily by measuring the number of positive ions produced by an electron beam of a definite energy passing through an atomic gas target. The relative cross sections for collisions in which one, two, or more electrons are removed from the atom have been determined by performing a charge analysis of the product ions. Fluctuations in the ionization cross section also arise from the excitation of inner-shell electrons, which produces unstable doubly excited states of the atom. Such states usually decay with a characteristic time of about 10^{-14} s by a radiationless process called Auger decay or autoionization, which involves two outer-shell electrons. One electron drops down to fill the inner-shell vacancy, and the other is ejected from the atom with a well-defined energy. It is also possible for such a doubly excited state to decay radiatively, and many experimental measurements have been made of x-ray emission resulting from atomic collisions. *See* AUGER EFFECT; IONIZATION.

There are more sophisticated ionization experiments in which both the scattered and ejected electrons are detected in coincidence. With a suitable choice of incident-electron energy and scattering geometry, the momentum distributions of electrons in atoms and molecules can be probed.

The cross section for radiative capture of an electron by an atom to form a negative ion is quite small (of the other of 10^{-21} cm^2 or less), and has been studied in shock-wave excited discharges and high-pressure arcs. The inverse process, called photodetachment of an electron from a negative ion, has been studied extensively by using lasers, and is of major importance in the solar atmosphere. The H^- ions determine the wavelength distribution of the continuous emission from the Sun in the visible region of the spectrum. *See* ARC DISCHARGE; SHOCK WAVE; SUN.

The experimental study of the scattering of electrons by atoms in short-lived excited states has been made possible by the use of lasers to prepare the atoms in specific states. Lasers also permit the study of collisions in the presence of strong electromagnetic fields.

Electron-ion collisions. Cross sections for collisions between electrons and positive or negative ions have been measured primarily by the colliding-beams method. Such measurements are often more

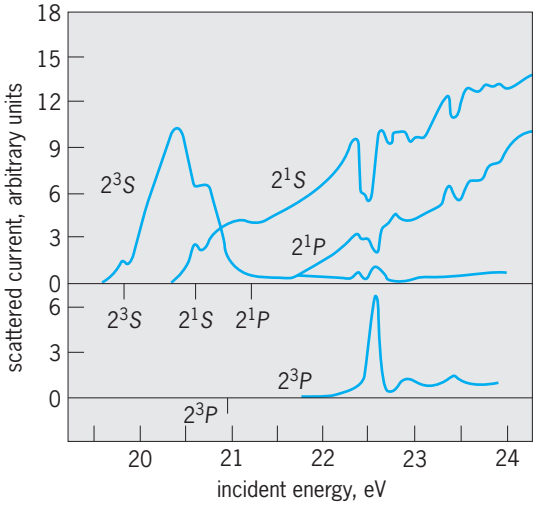


Fig. 5. Observed variation with incident electron energy of the current of electrons scattered without deviation after exciting the 2^1S, 2^1P, 2^3S, and 2^3P electronic states of helium. Excitation threshold energies are indicated. The values for the 2^3S curve are multiplied by 0.3.

accurate than those for atoms because the ions carry a charge and are more easily registered.

Excitation. The behavior with electron energy of cross sections for excitation of ions by electron impact are similar to those for atoms, except that for ions the cross section is finite and often largest at the threshold energy. This phenomenon is a consequence of the Coulomb attraction between the electron and ion, and is predicted by quantum theory. Total excitation cross sections have been measured for approximately a dozen singly and multiply charged ions, and differential cross sections have been measured for a few.

Ionization. Total cross sections for ionization of ions by electron impact have been measured for a large number of singly and multiply charged positive ions by using the crossed-beams method, in which the further-ionized ions are separated and detected after the collision. The ionization of singly charged ions is dominated by the direct "knock-on" ejection of outer-shell electrons, whereas an indirect mechanism involving excitation of inner-shell electrons can be important for multiply charged ions. In the latter case, doubly excited levels are formed that decay rapidly by autoionization, resulting in the net loss of one or more electrons from the ion. Cross sections have been measured for electron-impact single ionization of iron atoms and ions in charge states as high as +15. Iron ions play an important role in the solar corona and in high-temperature reactors used in fusion energy research. Cross sections for electron detachment from negative ions such as H^- have also been measured by the crossed-beams method.

Electron-ion recombination. An electron and a positive ion may recombine to form a neutral system or systems, if there is a mechanism available by which the recombination can be stabilized and the total energy conserved. This may be accomplished via the processes given in reactions (2)–(5).

Radiative recombination:

$$e^- + A^+ \rightarrow A + h\nu \qquad (2)$$

Dielectronic recombination:

$$e^- + A^+ \rightarrow A^{**} \rightarrow A^* + h\nu \qquad (3)$$

Dissociative recombination:

$$e^- + AB^+ \rightarrow A + B \qquad (4)$$

Three-body recombination:

$$e^- + A^+ + B \rightarrow A + B \qquad (5a)$$

$$e^- + e^- + A^+ \rightarrow A + e^- \qquad (5b)$$

Radiative recombination, given in reaction (2), where $h\nu$ represents an emitted photon, is important only at very low electron energies, and cross sections have been measured with a merged-beams technique. It has been predicted theoretically that this process, which is the inverse of photoionization, can be enhanced by an external radiation field. The stimulated radiative recombination process has been observed experimentally with a collinear laser beam. *See* PHOTOIONIZATION; PHOTON.

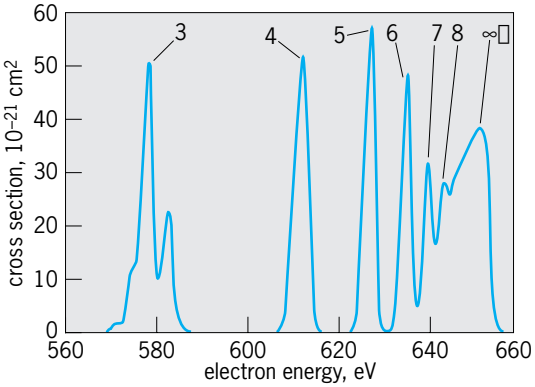


Fig. 6. Theoretical cross section for dielectronic recombination of electrons and hydrogenlike O^{7+} ions. The number above each resonance is the principal quantum number of the captured electron. Measurements made in a heavy-ion storage ring were generally in agreement with these theoretical values.

Dielectronic recombination, given in reaction (3), where A^* and A^{**} refer to a singly excited and doubly excited atom, respectively, is important in stellar atmospheres and in fusion-reactor plasmas. It is a resonant process resulting in the formation of an unstable, doubly excited state, A^{**}, of the electron-ion system. If the unstable state decays by a radiative transition rather than by autoionization, the recombination is stabilized. The branching ratio for these two pathways is critical in determining the cross section for dielectronic recombination. Intense electron beams have been merged with ion beams in heavy-ion synchrotron storage rings in order to measure cross sections for this process with extremely high energy resolution and precision (**Fig. 6**). Laser-stimulated dielectronic recombination has also been observed. *See* PARTICLE ACCELERATOR; PLASMA (PHYSICS).

Dissociative recombination, reaction (4) has been studied experimentally for many molecular ions by using crossed-beams, merged-beams, and flowing-afterglow techniques. Cross sections for this process are largest at near-thermal electron energies and, at their maximum, range from 10^{-15} cm^2 for simple diatomic ions to 10^{-12} cm^2 for complex polyatomic ions. This process is believed to play an important role in planetary ionospheres and in the formation of molecules in interstellar space. *See* INTERSTELLAR MATTER; IONOSPHERE.

Three-body recombination processes, reactions (5*a*) and (5*b*), have received less experimental study, and are important in high-density environments such as laser-produced plasmas.

Electron-molecule collisions. If an electron collides with a molecule rather than an atom, additional possibilities arise. Molecular vibration or rotation may be excited, and the molecule may dissociate into two or more neutral or ionized fragments. More than one of these processes may occur in the same collision event. It has been found experimentally that vibrational excitation of some molecules occurs most efficiently via the formation of a temporary unstable negative molecular ion. The excitation of rotation

is much more likely for molecules with permanent electric dipole or quadrupole moments. *See* MOLECULAR STRUCTURE AND SPECTRA.

Extensive studies have been made of the production of different ion fragments due to collisions of 100-eV electrons with various molecules. Such analysis using mass spectrometry is useful for industrial gases and vapors, especially hydrocarbons. Negative ion formation is most likely to occur via the process of dissociative electron attachment, given in reaction (6).

$$e^- + AB \rightarrow A + B^- \tag{6}$$

Collisions of atomic systems. If both the colliding systems are structured (for example, atoms, molecules, orions), elastic, excitation, or ionization processes may occur. In general, electronic excitation and ionization will be likely only if the relative speed of the colliding systems exceeds that of the bound electrons, and the cross sections will be similar to those for impact of electrons of the same speed. Either or both of the colliding systems may be excited in the collision.

Transfer collisions. There are additional possibilities called transfer collisions. They may involve the transfer of excitation, as in reactions (7) and (8). Reaction

$$A^* + B \rightarrow A + B^* \tag{7}$$

$$A^* + B \rightarrow A + B^+ + e^- \tag{8}$$

(8) is called Penning ionization.

Transfer of electrons from a neutral atom or molecule to an ion is most often studied by passing an ion beam through a gas target and measuring the fraction of ions whose charge has been changed. Slow ions produced in the gas may also be collected. At very low collision energies, the process has been studied by using merged beams. The electron transfer process may be represented in general by reaction (9), where q is the initial ion charge, k elec-

$$A^{q+} + B \rightarrow A^{(q-k)+} + B^{k+} + \Delta E \tag{9}$$

trons are transferred, and ΔE is the potential energy difference between the initial and final states of the system. If $\Delta E > 0$, the process is exoergic. For an endoergic process, $\Delta E < 0$, and this energy must be supplied by the initial kinetic energy of the reactants. The process is resonant if $\Delta E = 0$. Cross sections are large (10^{-14} cm^2) at low energies if ΔE is zero or small. For multiply-charged ions, the process is often exoergic ($\Delta E > 0$). Such a collision is characterized as high- or low-energy, depending on whether the relative velocity is large or small compared to the orbital velocity of the transferred electron (2.2×10^8 cm/s for hydrogen). The initial ionic charge determines the cross section at intermediate and high collision energies. At low energies the cross section depends on the structure of the transient quasimolecule formed during the collision.

Multiple electron transfer from multielectron atoms and molecules occurs predominantly into multiply excited levels of the ion which stabilize either radiatively, leading to true electron transfer, or via autoionization, leading to transfer ionization. Electron transfer processes are important in determining the ionization and energy balance in fusion plasmas and in the interstellar medium.

Radiative collisions. Extensive studies have been made of the emission of radiation by complex atoms and ions in collision at energies in the kiloelectronvolt range. This radiation is characteristic of transitions between states of the quasimolecule formed upon close approach of the colliding systems.

Ronald A. Phaneuf

Bibliography. E. W. McDaniel, *Atomic Collisions: Electron and Photon Projectiles*, 1989; H. S. W. Massey, E. H. S. Burhop, and H. B. Gilbody, *Electronic and Ionic Impact Phenomena*, 2d ed., vols. 1–5, 1969–1974; N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions*, 3d ed., 1965, paper 1987.

Scattering experiments (nuclei)

Experiments in which beams of particles such as electrons, nucleons, alpha particles and other atomic nuclei, and mesons are deflected by elastic collisions with atomic nuclei. Much is learned from such experiments about the nature of the scattered particle, the scattering center, and the forces acting between them. Scattering experiments, made possible by the construction of high-energy particle accelerators and the development of specialized techniques for detecting the scattered particles, are one of the main sources of information regarding the structure of matter. *See* PARTICLE ACCELERATOR; PARTICLE DETECTOR.

In the broad sense, any nuclear reaction is an example of scattering. However, this article treats elastic scattering only in the more restricted sense given below, and only insofar as it involves atomic nuclei. *See* COULOMB EXCITATION; GIANT NUCLEAR RESONANCES; NEUTRON SPECTROMETRY; NUCLEAR REACTION; NUCLEAR SPECTRA; NUCLEAR STRUCTURE.

Definitions of elastic scattering. The word "elastic" is used to indicate the absence of energy loss. If particle A collides with particle B of finite mass, there is a loss in the energy of A even if no energy has been transferred to the internal degrees of freedom of either A or B. Sometimes such a collision is referred to as inelastic, in order to distinguish its character from that of a collision with a particle having an infinite mass or its idealization, a fixed center of force. This terminology is not useful in the present context, because in the center-of-mass system of the two particles the sum of kinetic energies after the collision is the same as before. The distinction between elastic and inelastic scattering is made therefore on the basis of whether there are internal energy changes in the colliding particles. The collision is said to be inelastic even if the energy changes of the two particles compensate so as to leave the sum of the kinetic energies in the center-of-mass system unaltered. The treatment of inelastic scattering involves nuclear reaction theory because nuclear reactions markedly

influence the scattering. *See* COLLISION (PHYSICS).

Cross sections. The results of scattering measurements and calculations are generally expressed in terms of differential cross sections, which furnish a quantitative measure of the probability that the incident particle is scattered through an angle, θ. Cross sections have units of area, and in nuclear physics the convenient measure of area is the barn and its subunits, the millibarn and microbarn (1 barn = 10^{-28} m^2).

Coulomb scattering by nuclei. The simplest type of elastic scattering experiment relevant to the study of nuclei involves the deflection of incident electrically charged particles by the Coulomb field of the target nucleus. Such experiments provided the first evidence for atomic nuclei. H. Geiger and E. Marsden observed in 1909 that low-energy alpha particles could be scattered through large angles in collisions with gold and silver targets, and Ernest Rutherford showed in 1911 that these results could be understood if the scattering center consisted of a positively charged region (the nucleus) considerably smaller in size than that occupied by an individual target atom. The yield of scattered alpha particles as a function of angle, called the angular distribution, was measured in detail in a subsequent (1913) experiment by Geiger and Marsden, and the results were entirely consistent with the Rutherford expression for the differential cross section given by Eq. (1), where Z_1 and

$$\frac{d\sigma}{d\Omega_{\text{point}}}(\theta) = \left(\frac{Z_1 Z_2 e^2}{16\pi\epsilon_0 E}\right)^2 \sin^{-4}\frac{\theta}{2} \qquad (1)$$

Z_2 are the atomic numbers of the target and projectile, e is the electronic charge, ϵ_0 is the permittivity of free space, E is the center-of-mass energy of the alpha particle projectile, and θ is the center-of-mass scattering, or observation angle.

Equation (1) is appropriate only when the classical distance of closest approach, $d = Z_1 Z_2 e^2/4\pi\epsilon_0 E$, is greater than the combined size of the colliding pair; that is, only if the two charge distributions never overlap during the course of their interaction. The 1913 experiments involved collisions of 7.68-MeV alpha particles (1 MeV = 1.6×10^{-13} joule) with gold targets, and for this situation $d = 2.96 \times 10^{-14}$ m (29.6 fermis or femtometers in conventional nuclear physics notation, with 1 fm = 10^{-15} m). The detailed agreement between the Rutherford prediction of Eq. (1) and the 7.68-MeV data thus implied that the nuclear charge is contained within a region smaller than 30 fm (1.2×10^{-12} in.) in radius. Subsequent experiments involving higher-energy projectiles have determined the actual half-density radius of the gold nucleus to be approximately 7 fm (3×10^{-13} in.).

Although nuclear charge radii can be measured by systematically increasing the energy of the incident alpha particle until Eq. (1) no longer describes the angular distributions, specifically nuclear interactions come into play when the colliding nuclei overlap. In that situation, nuclear size is no longer the only factor being probed. This complexity is avoided when high-energy electrons are used as projectiles, since electrons are not subject to the nuclear forces. At very low energies the elastic scattering of electrons by nuclei conforms to the predictions of Eq. (1), but much more information is revealed at higher energies where the electrons can penetrate into the target nucleus. In the latter case and neglecting magnetic effects for the moment, the finite extent of the nucleus requires that Eq. (1) be modified by the presence of a nuclear form factor, $F(q^2)$, as in Eq. (2), in

$$\frac{d\sigma}{d\Omega}(\theta) = \frac{d\sigma}{d\Omega_{\text{point}}}(\theta)[F(\vec{q}^2)]^2 \qquad (2)$$

which q^2 is the momentum transfer in the collision. The significance of $F(q^2)$ can be appreciated by noting that at nonrelativistic energies, and assuming the Born approximation to be valid, $F(q^2)$ is the Fourier transform of the nuclear charge distribution.

If sufficiently complete measurements over a wide range of momentum transfer are available, the form factor can be determined in detail, and the nuclear charge density, $\rho_{\text{ch}}(r)$, obtained directly through the inverse transform. In practice, a functional form is usually assumed for $\rho_{\text{ch}}(r)$, and its parameters are adjusted to fit the electron scattering data. In this manner it is found that the experiments are consistent with nuclear charge distributions specified by Eq. (3). The parameter c is the radius at which the

$$\rho_{\text{ch}}(r) = \rho_0[1 + e^{(r-c)/a}]^{-1} \qquad (3)$$

density falls to one-half its central value. It assumes the typical value $c \approx 1.1A^{1/3}$ fm, where A is the atomic number of the target nucleus. The quantity a (typically ≈ 0.55 fm) reflects the fact that nuclear surfaces are not sharply defined, so that the density falls off gradually with increasing radius. The so-called central charge density, ρ_0, is found to be approximately 1.1×10^{25} coulombs/m^3, which corresponds to 0.07 proton per cubic femtometer (1.1×10^{39} protons per cubic inch). That ρ_0 is approximately the same for all but the lightest nuclei and $c \propto A^{1/3}$ implies that nuclear matter is nearly incompressible. The charge distributions of a variety of nuclei, as determined by electron scattering measurements, are shown in **Fig. 1**.

Electron scattering measurements involving large angular momentum transfer are sensitive to the distribution of magnetism as well as charge, and Eq. (2) then must be generalized to reflect the contributions from the two distributions. In addition, very precise measurements show that Eq. (3) is only approximately correct and the form of $\rho_{\text{ch}}(r)$ varies slightly from nucleus to nucleus as a consequence of variations in the structure of the nuclei in question.

Electron-nucleon scattering. Scattering of electrons by hydrogen gives information regarding the electron-proton interaction. From measurements of the variation of the differential cross section, it has been found necessary to postulate that both the proton charge and its intrinsic magnetic moment are distributed through a finite volume. Existing work favors the assumption of similarity of shape of these

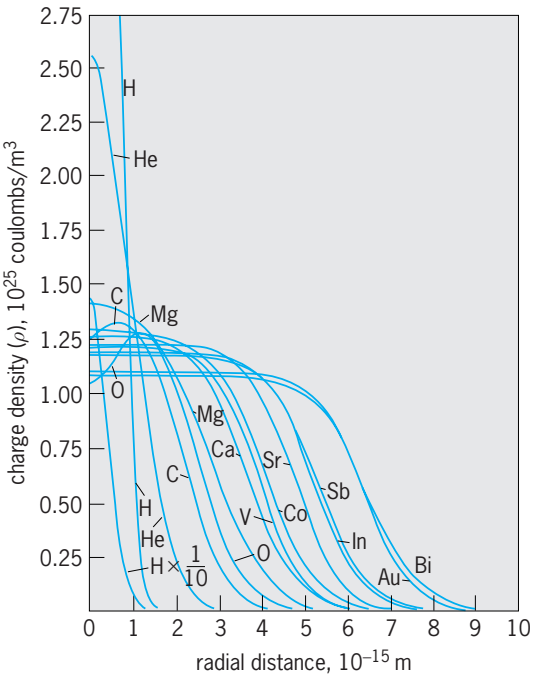


Fig. 1. Summary of the nuclear charge distributions found for various nuclei by electron scattering methods. (*After R. Hofstadter, Nuclear and nucleon scattering of high-energy electrons, Annu. Rev. Nucl. Sci., 7:231–316, 1957*)

distributions. Energies in excess of 200 MeV are required for the detection of these effects, and the anomalous magnetic moment of the proton plays an important part at high energies. The experiments make it probable that the charge density has a root-mean-square radius of approximately 8×10^{-14} cm (3×10^{-14} in.). Measurements on the scattering of electrons by deuterium at large angles and higher energies lead to the conclusion that the magnetic moment of the neutron is not concentrated at this point. If it is assumed that neutron and proton magnetic moments are distributed through nearly the same volumes, a good representation of the scattering measurements is obtained.

Analyses of scattering data at various angles and energies indicate that there is no net charge density within the volume occupied by the neutron. This result is in agreement with measurements of the neutron-electron interaction made by scattering very slow neutrons from atomic electrons and atomic nuclei. While there is an interaction equivalent to a potential energy of approximately −3900 eV through a distance of e^2/mc^2, where m is the electron mass and c is the speed of light, amounting to approximately 2.8×10^{-13} cm (1.1×10^{-13} in.), it is accounted for qualitatively as a consequence of what E. Schrödinger called the *Zitterbewegung* (tremblatory motion) expected for the neutron.

Electron scattering by nucleons has also provided valuable information concerning subnuclear processes. *See* ELEMENTARY PARTICLE; QUARKS.

Low-energy np scattering. Since the neutron is electrically neutral, an understanding of how it scatters in an interaction with a proton requires, even for low collision energies, a knowledge of the properties of the nuclear forces. Conversely, the scattering experiments probe these properties. The *np* force is responsible for the binding together of a neutron, n, a proton, p, to form a deuteron, and the detailed connection between *np* scattering cross sections and the properties of the deuteron was perceived in the early 1930s. For example, the assumption that the spatial extension, or range, of the *np* force is relatively small makes it possible to estimate the magnitude of the force, given the measured binding energy of the deuteron, and Eugene Wigner used this information in 1933 to calculate *np* scattering cross sections. The measured yield was found to be larger than would be expected if the forces between free neutrons and free protons were the same as those in the deuteron. To explain this difference, it was postulated by Wigner in 1935 that the *np* interaction is spin-dependent; that is, it depends on the relative orientation of the spins of the interacting particles. The proton and neutron are known to have a spin of $^1/_2$; that is, their intrinsic angular momenta are known to be $(h/2\pi)/2 = \hbar/2$, where h is Planck's constant. According to quantum mechanics, when two spins s_1, s_2 combine vectorially, only the values given by Eq. (4) are possible for the resultant s. For the *np* system, therefore, the resultant spins are 0 or 1. In the first case one speaks of a singlet, and in the second of a triplet. *See* ANGULAR MOMENTUM; SPIN (QUANTUM MECHANICS).

$$s = s_1 + s_2, s_1 + s_2 - 1, \ldots, |s_1 - s_2| \qquad (4)$$

The singlet state behaves much like a round and perfectly smooth object which has the same appearance no matter how it is viewed, corresponding to only one possibility of forming a state with $s = 0$. The state with $s = 1$, on the other hand, can have three distinct spin orientations. Measurement of the projection of s on an axis fixed in space can give only the three values (1, 0,−1), again in units $\hbar$. When protons with random spin directions collide with neutrons also having random spin directions, the triplet state is formed three times as often as the singlet. The deuteron, however, is in a triplet state. Thus the hypothesis of spin dependence can account for the difference between the forces in the deuteron and those in *np* scattering.

The neutron-hydrogen scattering experiments on which these conclusions were based were performed with slow neutrons having energies of a few electronvolts or less. The general quantum-mechanical theory of scattering is much simplified in this case. Because of the small range of nuclear forces, only the collisions with zero orbital angular momentum ($L\hbar = 0$) play a role, collisions with higher orbital angular momenta missing the region within which nuclear interactions take place. States with $L = 0$ (called S states) have the property of spherical symmetry, and nuclear forces matter in this case only inasmuch as they modify the spherically symmetric part of the wave functions.

The long-wavelength or low-energy scattering cross section can be described completely and sim-

ply by a quantity called the scattering length. For interparticle distances r greater than the range of nuclear forces, the spherically symmetric part of the wave function has the form given by expression (5),

$$C\left(\frac{1+a}{r}\right) \tag{5}$$

where a and C are constants. The constant a is called the scattering length. It has the following meaning. If $R(r)$ denotes the wave function describing the np relative motion, the product $rR(r)$, when plotted against r, is represented by a straight line which cuts the axis of r at a distance a from the origin of coordinates. If the intersection is to the left (right) of the origin, a is counted as positive (negative). The two conditions are illustrated in **Figs. 2** and **3**. The possibilities $a > 0$ and $a < 0$ are sometimes referred to as those of the virtual and real level, respectively. Sometimes the opposite convention regarding the sign of a is used. The convention adhered to here provides the simplest connection with phase shifts.

The scattering cross section for a state with well-defined a is $4\pi a^2$, and thus low-energy np scattering is described by Eq. (6), where σ_{np} is the total

$$\sigma_{np} = 4\pi({}^1/_4 a_s^2 + {}^3/_4 a_t^2) \tag{6}$$

cross section and a_s and a_t are the singlet and triplet scattering lengths, respectively. Comparison of the scattering yields with the deuteron data provides the values $a_s = 24$ fm and $a_t = 5.4$ fm.

There are several additional means by which np scattering can be calculated. For example, with the techniques outlined below ("Potential scattering"), effective interaction potentials can be found which simulate the np nuclear forces to the extent that calculations using the potentials reproduce the measured cross sections. Two potentials corresponding

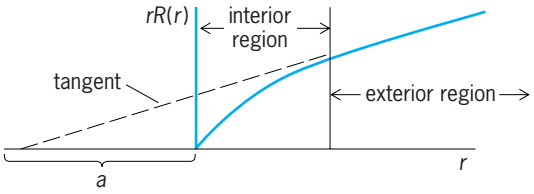


Fig. 2. Scattering length in the case of a virtual level.

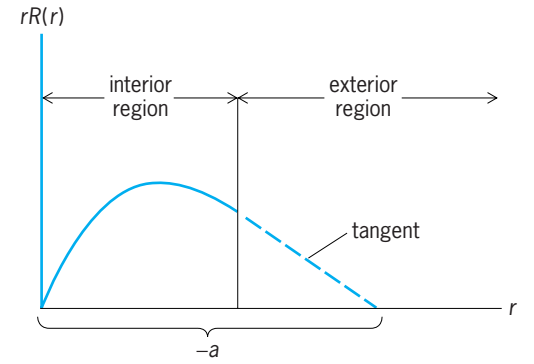


Fig. 3. Scattering length in the case of a real level.

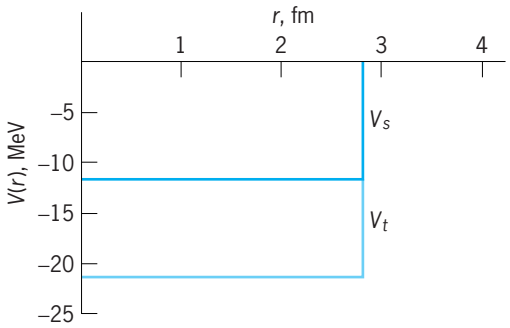


Fig. 4. Examples of singlet (V_s) and triplet (V_t) potential wells suitable for low-energy *np* scattering. The potentials are not uniquely determined; many other equally satisfactory possibilities exist.

to singlet and triplet $L = 0$ scattering are required for a description of low-energy np data, as is illustrated in **Fig. 4**. The term "potential well" is often used to describe these potential energy surfaces, because the system can be trapped in the region of space occupied by the potential somewhat similarly to the way in which water is trapped in a well. For nucleon-nucleon scattering, it is frequently useful to express the potential energy in the form $V(r) = V_0 f(r/b)$, where f is a function which determines the shape of the well. The constants V_0 and b are usually referred to as the depth and range parameters. The scattering length determines approximately the product $V_0 b^2$ for a potential well of assigned shape. The variation of the cross section with energy E through an energy range of a few MeV can be used for the determination of b. *See* NONRELATIVISTIC QUANTUM THEORY; QUANTUM MECHANICS.

Low-energy pp scattering. The pp and pn interactions are believed to be closely equal. Gregory Breit proposed in 1936 the hypothesis of charge independence of nuclear forces, which supposes that nuclear forces, acting in addition to the electrostatic (Coulomb) repulsion, in the pp, pn, and nn cases are equal to each other. Because of the limits of experimental accuracy and uncertainties in the theoretical interpretation, the hypothesis is not established with perfect accuracy, but it is believed to hold within a few percent for the depth parameter V_0 if the range parameter b is specified as the same in the three cases. The concept of charge independence also has important implications for the structure of nuclei. *See* I-SPIN.

Phase shifts. Scattering can be treated by means of phase shifts, which will be illustrated for two spinless particles. The wave function of relative motion will be considered first for a state of definite orbital angular momentum $L\hbar$, with L representing an integer. Outside the range R of nuclear forces, the wave function may be represented by Eq. (7), where k is

$$\psi_L = \frac{Y_{LM}(\theta, \phi)\mathscr{F}_L(kr)}{(kr)} \tag{7}$$

$$(r > R)$$

2π times the reciprocal of the wavelength, the so-called wave number, θ and σ are the colatitude

and azimuthal angles of a polar coordinate system, and Y_{LM} is the spherical harmonic of order L and azimuthal quantum number M. The form of ψ_L is determined by the Schrödinger wave equation which restricts $\mathscr{F}_L$ by the differential equation (8),

$$\left\{\frac{d^2}{dr^2} + \frac{k^2 - L(L+1)}{r^2}\right\}\mathscr{F}_L = 0 \qquad (r > R) \tag{8}$$

it being supposed that there is no Coulomb field. In the absence of nuclear forces, $\mathscr{F}_L$ satisfies the same equation at all distances and, aside from a constant factor, has its asymptotic form determined by the boundary conditions at $r = 0$ as in notation (9). In

$$\mathscr{F}_L \sim \sin\left(kr - \frac{L\pi}{2}\right) \text{ as } r \to \infty \tag{9}$$

the presence of nuclear interactions, the asymptotic form is given by notation (10), where δ_L is a constant, called the phase shift, which determines the scattering. *See* SPHERICAL HARMONICS.

$$\mathscr{F}_L \sim \sin\left(kr - \frac{L\pi}{2} + \delta_L\right) \tag{10}$$

When elastic scattering is the only process that can take place, as for low- and medium-energy nucleon-nucleon scattering, the phase shifts are real numbers. Above a few hundred megaelectronvolts, nucleon-nucleon scattering becomes strongly inelastic because of pion production, however, and the phase parameters become complex numbers reflecting the loss of flux from the elastic channel. Nucleon-nucleus and nucleus-nucleus interactions are characterized by the presence of many nonelastic scattering channels, except at very low energies.

The wave packet representing the scattered particle consists of a superposition of all possible partial waves, $\mathscr{F}_L$, and calculation of the scattering cross section requires a knowledge of all phase shifts. In practice, however, only partial waves with orbital angular momenta corresponding to impact parameters within the range of the nuclear forces suffer nuclear phase shifts different from zero. The direct method of varying phase shifts to reproduce experimental data is most useful, therefore, when only a few L values satisfy this condition, such as for low- and intermediate-energy nucleon-nucleon scattering and for low-energy nucleon-nucleus and nucleus-nucleus collisions.

For charged spinless particles, such as two alpha particles, the phase shifts caused by specifically nuclear forces add to the asymptotic phase of the functions $\mathscr{F}_L$ for the Coulomb case, which differs from the non-Coulomb case only through the replacement of $kr - L\pi/2$ by $kr - L\pi/2 - \eta \ln(2kr) + \arg\Gamma(L + 1 + i\eta)$, where $\eta = Z_1Z_2e^2/4\pi\epsilon_0\hbar v$, Z_1e and Z_2e represent the charges on the colliding particles, e represents the electronic charge, and v represents the relative velocity. *See* GAMMA FUNCTION.

As in the *np* scattering case discussed above, the presence of spin adds additional complications. For example, if two particles with spin $^1/_2$ collide, it is necessary, in general, to introduce phase shifts for each state with definite total angular momentum $J\hbar$.

Potential scattering. The direct determination of phase shifts through comparison with scattering data becomes difficult in situations where large numbers of partial waves, $\mathscr{F}_L$, are modified by the nuclear scattering center. The usual procedure in this case is to simulate the influence of the nuclear interactions by means of potentials, $V(r)$. For the nonrelativistic scattering of two spinless, uncharged particles, Eq. (8) is supplemented by the Schrödinger equation for $r < R$, Eq. (11), in which m is the reduced mass of the colliding pair. Equations (8) and (11) are then solved subject to appropriate boundary conditions to determine the $\mathscr{F}_L(r)$ and hence δ_L for each L. A functional form is generally assumed for $V(r)$, and its parameters are varied and Eqs. (8) and (11) solved iteratively until the measured cross sections, polarizations, and so forth are reproduced.

$$\left\{\frac{d^2}{dr^2} + \left[k^2 - \frac{L(L+1)}{r^2} - \frac{2mV(r)}{\hbar^2}\right]\right\}\mathscr{F}_L(r) = 0 \qquad (r > R) \tag{11}$$

From the viewpoint of microscopic models of nucleon-nucleon interactions, it appears highly improbable that a description of nucleon-nucleon scattering in terms of two-body energy-independent local potentials can have fundamental significance. Nevertheless, in a limited energy range, it is practical and customary to represent scattering by means of such a potential. As noted above, the simple potentials illustrated in Fig. 4 suffice for the description of low-energy *np* scattering. At somewhat higher energies, but below the threshold of meson production, a real potential may still be used, but different potentials are required for triplet-even, triplet-odd, singlet-even, and singlet-odd states to account for the (even, odd) parity dependence of the interactions. *See* PARITY (QUANTUM MECHANICS).

Intermediate energy np and pp scattering. In the intermediate-energy region (10–440 MeV) the analysis of experimental material is more difficult than at low energies because of the necessity of employing many phase shifts and coupling constants. Analysis in terms of phase parameters involves fewer assumptions than that in terms of potentials. Except for approximations connected with the infrared catastrophe and related small inaccuracies in relativistic treatment of Coulomb scattering, it is based on very generally accepted assumptions, such as the validity of time reversal and parity symmetries for strong interactions. With infinite experimental accuracy, it should be possible to extract all the phase parameters from measurements of the differential cross section, the polarization spin correlation coefficients, and "triple scattering" quantities describing spin orientation which, for unpolarized incident beams and unpolarized targets, require three successive scatterings.

The analysis is usually carried out by assuming that for sufficiently high L and J the phase parameters may

be represented by means of the one-pion exchange approximation. The value of the pion-nucleon coupling constant g^2 is often varied in an attempt to improve the fit to experimental data, and values for best fits are compared with those from pion physics. Reasonable agreement usually results.

The consistency of values of g^2 from *pp* data with those from *np* measurements indicates approximate validity of charge independence at the larger distances.

Potentials to be used in a nonrelativistic Schrödinger equation and capable of representing *pp* and *np* scattering have been devised either on a purely phenomenological or semiphenomenological basis. The former way provides a more accurate representation of the data. Nonrelativistic local potentials required from 0 to 310 MeV are different according to whether the state is even or odd, singlet or triplet. It is necessary to use central, tensor, spin-orbit, and quadratic spin-orbit parts of the potential. Most of the accurately adjusted potentials employ hard cores within which the potential is infinite. The spin-orbit potential suggested by *pp* scattering data indicated the probable participation of vector-meson exchange in nucleon-nucleon scattering, anticipating the discoveries of the ω- and ρ-mesons, as well as fictitious mesons, the latter partly intended as a representation of simultaneous two-pion exchange. At short distances, the potentials are often modified in order to improve agreement with experiment. Superposition of the single-boson potentials combined with the short-range modifications of the resulting potential gives the so-called one-boson exchange (OBE) potentials. These provide a fair but not excellent reproduction of phenomenological fit phase shifts. *See* QUANTUM FIELD THEORY.

Nucleon-nucleus scattering. Nucleon-nucleus scattering at low and intermediate energies ($E <$ 300 MeV) is of interest primarily in connection with nuclear structure studies. The scattering wave functions, $\mathscr{F}_L(r)$, are required for quantitative interpretation of the inelastic scattering and reaction experiments which have been the main source of information regarding the properties of nuclear excited states. These scattering functions are obtained as the solutions of Eq. (11) when elastic data are reproduced by using potential models.

Nucleon-nucleus scattering experiments can be accounted for by a potential-well model with a complex potential (optical model) making use of a spin-orbit interaction term. The details of angular distributions of the cross section and polarization are reproduced remarkably well, and experiments favor some potential well shapes over others. Wells thus determined are wider than those obtained from electron-nucleus scattering experiments, and similar functional shapes work in both cases. The potential energy represented by the wells is added to the electrostatic potential energy in the calculations. The electrostatic potential energy is approximated by a central potential corresponding to the average distribution of nuclear charge. The quantitative success of the potential-well approach to the scattering of 30-MeV protons by a variety of nuclei is illustrated in **Fig. 5**.

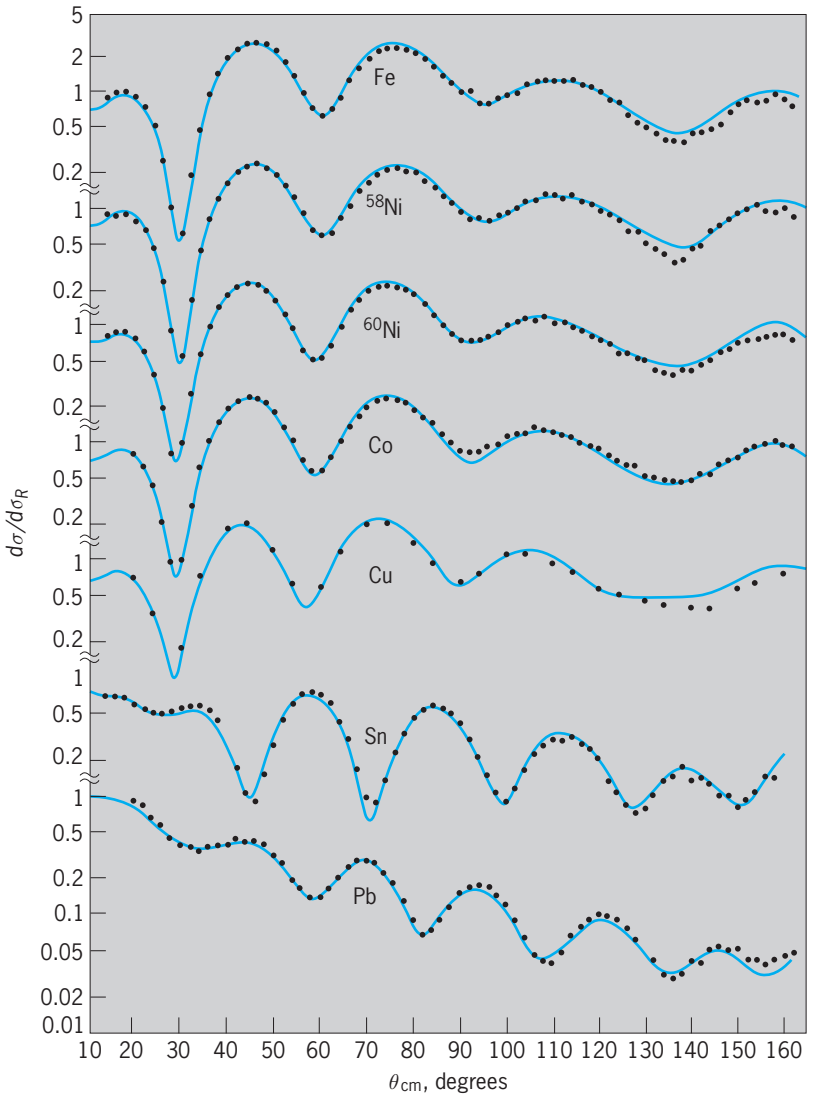


Fig. 5. Measured (points) and calculated (curves) differential cross sections for the scattering of 30-MeV protons from a variety of target nuclei, as a function of scattering angle in the center-of-mass system θ_{cm}. The theoretical calculations made use of optical potential wells. The vertical scale represents the differential cross section $d\sigma$ divided by the Rutherford differential cross section $d\sigma_R$. (*After G. R. Satchler, Optical model for 30 MeV proton scattering, Nucl. Phys., A92:273–305, 1966*)

Optical model potential fits to nucleon-nucleus data are in general agreement with the data in a wide energy range from several MeV to about 300 MeV, but the parameters of the potential have to be varied progressively. The spin-orbit potential found to represent the scattering data at the lower energies has the same value of its ratio to $(dV_{ct})/(rdr)$, where V_{ct} is the central potential, as it has in the shell theory of nuclear structure. At higher energies (300 MeV), it has been found that the spin-orbit potential has to be used with a smaller strength than in shell theory. The real part of the central potential has to be used with a smaller strength than in shell theory. The real part of the central potential decreases with energy and becomes almost zero at 300 MeV. Data at 1 GeV on proton scattering from carbon can be accounted

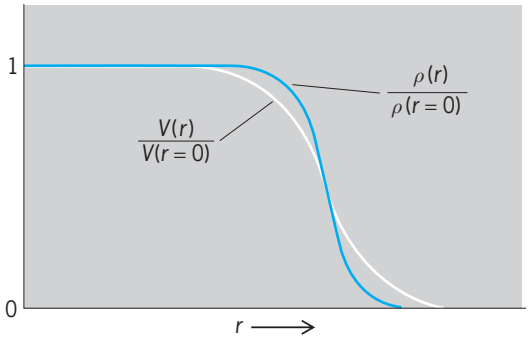


Fig. 6. Comparison between the shape of a nuclear density distribution, $\rho(r)$, and a potential, $V(r)$, obtained from it by folding in an interaction potential, $v(r)$, with a finite but short range as in Eq. (12). (*After G. R. Satchler, Introduction to Nuclear Reactions, John Wiley and Sons, 1980*)

for on the optical model by means of an imaginary central and real spin-orbit potential.

The real part of the volume potential $V(\vec{r}_p)$ can be calculated by folding the interaction potential between the projectile and a target nucleon, $v_{tp}(\vec{r})$, with the nuclear density of the target as in Eq. (12).

$$V(\vec{r}_p) = \int \rho(\vec{r}_t) v(\vec{r}_{tp}) d\vec{r}_t \qquad (12)$$

Effective, rather than free, nucleon-nucleon interactions are used for $v(\vec{r}_{tp})$ because the target nucleon is embedded in a nucleus. The short range of $v(\vec{r}_{tp})$ implies that the resulting central potential has a spatial distribution closely similar to, but somewhat more rounded than, that of the target nuclear density, as is illustrated in **Fig. 6**. The imaginary part of the optical potential is phenomenologically determined.

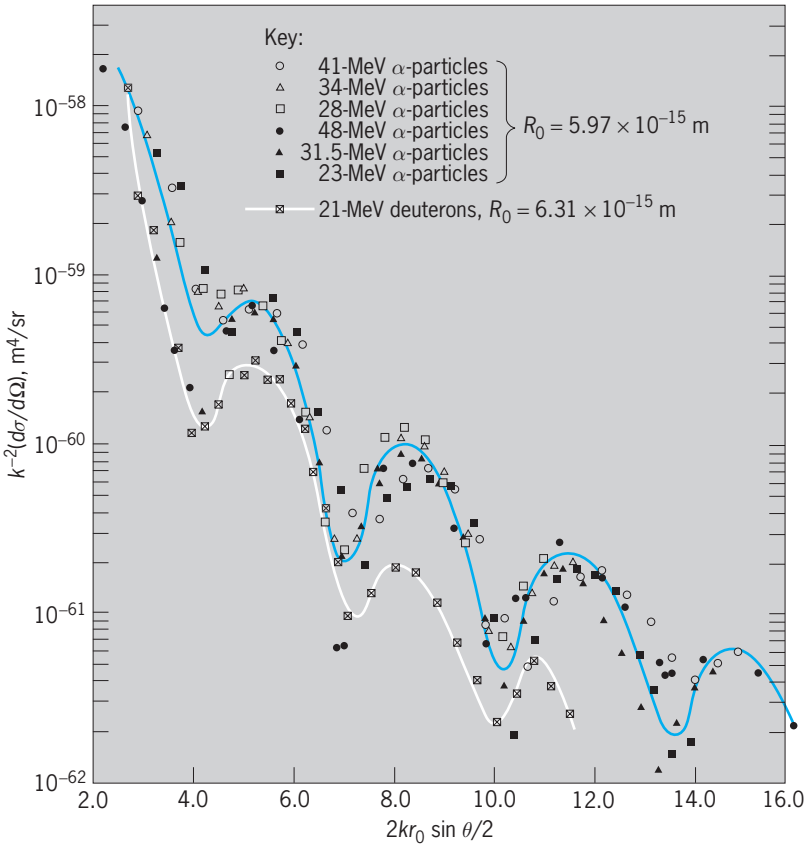


Fig. 7. Angular distribution for the elastic scattering of deuterons and alpha particles from magnesium, showing the typical Fraunhöfer-like diffraction pattern. (*After J. S. Blair et al., Diffraction analysis of elastic and inelastic scattering by magnesium, Nucl. Phys., 17:641–654, 1960*)

Nucleus-nucleus scattering. Composite nuclei ranging from deuterium through uranium have been accelerated to energies as high as several GeV per constituent nucleon. Elastic scattering is an important process only toward the lower end of this energy range, perhaps only below a few tens of MeV per nucleon; at the higher energies, the principal interest is in nonelastic collisions, which are expected to provide information concerning properties of nuclear matter under unusual conditions (like high temperature and high density).

Nucleus-nucleus scattering at low and intermediate energies is characterized by strong absorption associated with the relatively short mean free path for nuclei in nuclear matter. Only peripheral or glancing collisions are likely to lead to direct elastic scattering; more head-on collisions lead to more complicated processes such as compound nucleus formation which absorb flux from the incident beam. Since the de Broglie wavelengths of the incident nuclei are typically comparable to or smaller than the size of the target nuclei, the strong absorption for head-on collisions frequently causes the scattering to be diffractive in nature, and the elastic scattering angular distributions to resemble those observed in the scattering of light by small opaque obstacles. A typical example, corresponding to optical Fraunhöfer diffraction, is illustrated in **Fig. 7**.

Good fits to nucleus-nucleus scattering data are obtained by using potential wells. For collisions involving relatively light nuclei, the real potentials may be derived by folding an effective nucleon-nucleon interaction with the densities of the two colliding nuclei, in a slight generalization of Eq. (12). For more massive nuclei, analogies with the collisions of liquid droplets, provide guidance in determining the potentials. In many cases, purely phenomenological potentials provide the best fits to the data. There is considerable ambiguity in the potential parameters, because the scattering depends primarily on the values of the potentials at distances corresponding to glancing collisions, and is relatively insensitive to the values at smaller distances. The imaginary potentials are deeper than those used for nucleon-nucleus scattering because of the strong absorption property. Except for collisions involving light nuclei, no very strong evidence for the influence of spin-orbit potentials on elastic scattering yields has been found, but there is evidence suggesting that these may become important for more massive nuclei at energies in excess of 10 MeV per nucleon.

The optical model potential is not a potential in the ordinary sense. When inserted in the wave equation (11), it gives agreement with experiment by simulating the complicated interactions between the two colliding many-body systems in terms of a prescribed

functional form. Possible variations in this form, associated with the varying proximity of the two nuclei during the collision, are generally neglected. Alternative methods, such as the time-dependent Hartree-Fock approach, are being developed to treat nucleus-nucleus interactions from a more fundamental perspective.

Nucleus-nucleus scattering sometimes exhibits anomalous behavior which cannot be reproduced with potential models and which is believed to reflect the transient formation of moleculelike configurations. These configurations, or resonances, signal their presence by producing a rapid variation in the energy dependence of the scattering yields and by modifying the behavior of the angular distribution data. *See* NUCLEAR MOLECULE.

Meson scattering. The scattering of π mesons (pions) by nucleons has been studied intensively as an example of a strong interaction between a boson and a baryon, and particularly because of the connection with nucleon-nucleon interactions. Prominent resonances occur in pion-nucleon scattering yields, the best known of which appears for pion bombarding energies of about 200 MeV. Charge independence in pion-nucleon interactions is confirmed by the scattering measurements and their phase-shift analyses.

The scattering of pions by nuclei has been studied over the energy range extending from below 20 MeV to beyond 60 GeV. At low energies the experiments probe nucleon-nucleon correlations within nuclei. Attempts to relate the scattering behavior to the properties of the nuclear wave functions have been partially successful. Comparison of π^+ with π^- scattering yields provides information concerning differences between proton and neutron distributions within nuclei. Future study of pion-nucleus interactions at high energies is expected to reveal properties of the strong interactions, modified by the nuclear movement, at short distances.

The study of kaon-nucleus interactions, still in its infancy, is currently of interest primarily in connection with the production of nuclei in which, for example, a neutron is replaced by a lambda particle (Λ), and the investigation of Λ-nucleon and related interactions. *See* MESON. K. A. Erb

Bibliography. R. C. Barrett and D. F. Jackson, *Nuclear Sizes and Structure*, 1977; G. Breit and R. D. Haracz, Nucleon-nucleon scattering, in E. H. S. Burhop (ed.), *High-Energy Physics*, vol. 1, 1967; D. M. Brink, *Semi-Classical Methods for Nucleus-Nucleus Scattering*, 1986; M. Goldberger and K. M. Watson, *Collision Theory*, 1964, reprint 1975; G. R. Satchler, *Introduction to Nuclear Reactions*, 2d ed., 1990.

Scattering layer

A layer of organisms in the sea which causes sound to scatter and returns echoes. Recordings by sonic devices of echoes from sound scatterers indicate that the scattering organisms are arranged in approximately horizontal layers in the water, usually well above the bottom. The layers are found in both shallow and deep water.

Shallow water. In the shallow water of the continental shelves (less than 600 ft or 200 m deep), scattering layers and echoes from individuals or compact groups are very irregularly distributed and are probably made up of a variety of sea animals and possibly plants. While some fishes, notably herring, and some zooplankton have been identified by fishing, many others have not been identified.

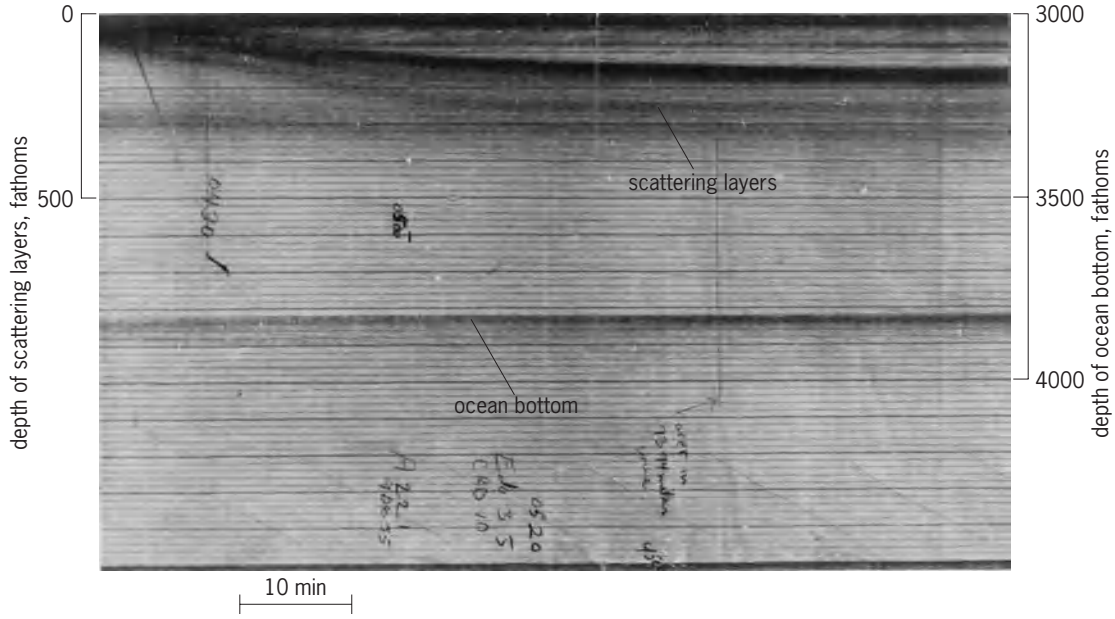


Fig. 1. Record by a 12-kHz echo sounder of the sunrise descent of deep scattering layers in eastern Pacific off northern Chile. A layer, which appears to have remained at depth throughout the night, is shown near 300 fathoms (540 m). 1 fathom = 1.8 m. (*From M. N. Hill, ed., The Sea, vol. 1, Interscience, 1962*)

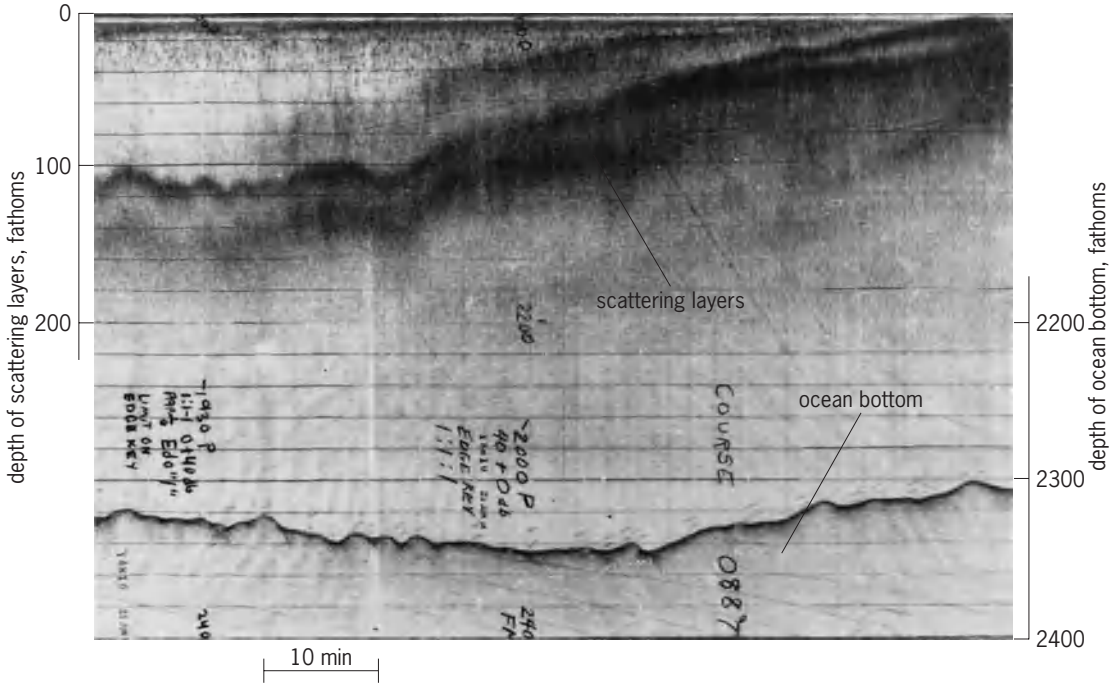


Fig. 2. Sunset ascent of deep scattering layers in western North Atlantic near 40°30′N, 50°W. 1 fathom = 1.8 m. (*From M. N. Hill, ed., The Sea, vol. 1, Interscience, 1962*)

Deep scattering layers. In deep water (greater than 600 ft or 200 m) one or more well-defined layers generally are present. Though commonly variable in detail, they are found to be very widely distributed. They are readily detected by echo-sounding equipment operating in the frequency range 3–60 kHz, the sound spectrum of each layer generally having maximum scattering at a somewhat different frequency than others found at the same place. Commonly, but not universally, the deep-water layers migrate vertically in apparent response to changes in natural illumination. The most pronounced migration follows a diurnal cycle, the layers rising at night, sometimes to the surface, and descending to greater depths during the day (**Figs. 1** and **2**). The common range of daytime depths is 600–2400 ft (200–800 m). The migration is modified by moonlight and has been observed to be modified during the day by heavy local cloud cover, for example, a squall. Occurrence of the layers in deep water was first demonstrated by C. Eyring, R. Christiansen, and R. Raitt. *See* ECHO SOUNDER; UNDERWATER SOUND.

Deep scattering organisms. All animals and plants, as well as nonliving detritus, contrast acoustically with seawater and, hence, any may be responsible for observed scattering in a particular instance. Many animals and plants have as part of their natural structure a gas-filled flotation organ which scatters sound many times more strongly than would be inferred from the sound energy they intercept. These are the strongest scatterers of their size. M. W. Johnson pointed out in 1946 that the layers which migrate diurnally must be animals that are capable of swimming to change their depth, rather than plant life or some physical boundary such as an abrupt

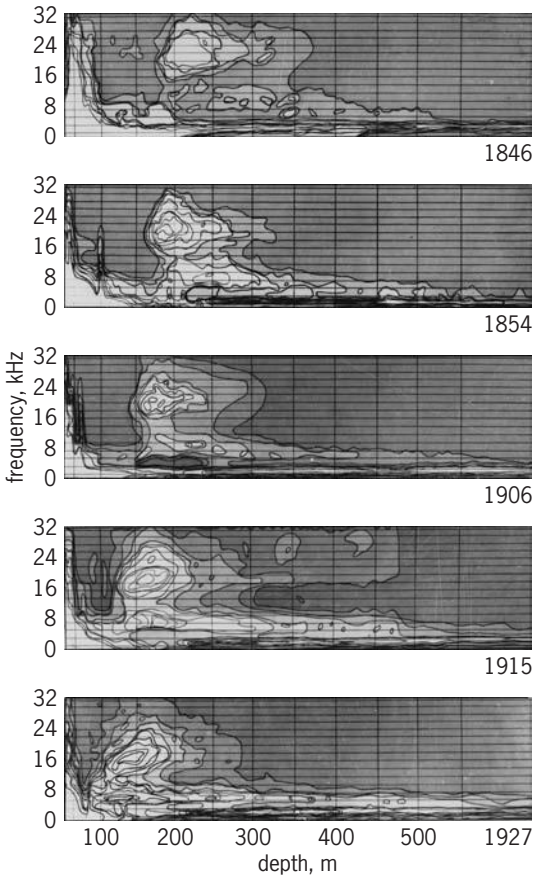


Fig. 3. Sequence of sunset observations showing scattering as a function of depth and frequency. Contours of equal sound level are 2 dB apart, with lightest areas denoting highest levels. Time of day of each observation is indicated by the number at the lower right-hand corner of each record. 1 m = 3.3 ft. (*From M. R. Hill, ed., The Sea, vol. 1, Interscience, 1962*)

temperature change in the water. In 1953 V. C. Anderson demonstrated that some of the deep-water scatterers have a much smaller acoustical impedance than seawater. This fact fits the suggestion by N. B. Marshall in 1951 that the scatterers may be small fishes with gas-filled swim bladders, many of which are known to be geographically distributed much as the layers are. In 1954 J. B. Hersey and R. H. Backus found that the principal layers in several localities migrate in frequency of peak response while migrating in depth, thus indicating that the majority of scatterers fit Marshall's suggestion (**Fig. 3**). In deep water, corroboration of Marshall's suggestion has come from several independent combined acoustical and visual observations made during dives of the deep submersibles *Trieste*, *Alvin*, *Soucoupe*, *Deep Star*, and others. Nearly all these dives were made within a few hundred miles of the east or west coasts of the continental United States. The fishes observed most commonly to form scattering layers are the lantern fishes, or myctophids, planktonic fishes that are a few inches long and possess a small swim bladder. The siphonophores (jellyfish) also have gas-filled floats and have been correlated with scattering layers by observation from submersibles. It is not clear whether fishes with swim bladders and siphonophores make up the principal constituents of the deep-water scattering layers in deep-ocean areas, but it is nearly certain that they are not exclusively responsible for all the widely observed scattering. *See* DEEP-SEA FAUNA.

John B. Hersey

Bibliography. M. N. Hill (ed.), *The Sea*, vol. 1, 1962, reprint 1982; J. S. Levinton, *Marine Biology: Function, Biodiversity, Ecology*, 2d ed., 2001; G. L. Pickard and W. J. Emery, *Descriptive Physical Oceanography*, 5th ed., 1990.

Scattering matrix

An infinite-dimensional matrix or operator that expresses the state of a scattering system consisting of waves or particles or both in the far future in terms of its state in the remote past; also called the S matrix. In the case of electromagnetic (or acoustic) waves, it connects the intensity, phase, and polarization of the outgoing waves in the far field at various angles to the direction and polarization of the beam pointed toward an obstacle. It is used most prominently in the quantum-mechanical description of particle scattering, in which context it was invented in 1937 by J. A. Wheeler to describe nuclear reactions. Because an analog of the Schrödinger equation for the description of particle dynamics is lacking in the relativistic domain, W. Heisenberg proposed in 1943 that the S matrix rather than the hamiltonian or the lagrangian be regarded as the fundamental dynamical entity of quantum mechanics. This program played an important role in high-energy physics during the 1960s but is now largely abandoned. The physics of fundamental particles is now described primarily in terms of quantum gauge fields, and these are used to determine the S matrix and its elements for the collision and reaction processes observed in the laboratory. *See* ELEMENTARY PARTICLE; GAUGE THEORY; NONRELATIVISTIC QUANTUM THEORY; NUCLEAR REACTION; QUANTUM MECHANICS; RELATIVISTIC QUANTUM THEORY; SCATTERING EXPERIMENTS (ATOMS AND MOLECULES); SCATTERING EXPERIMENTS (NUCLEI).

Definition and properties. A normalized quantum wave packet $\Psi_{in}(t)$ may be defined that describes a free particle, with specified quantum numbers (such as spin and isospin) and almost fixed energy and momentum (almost, because fixing the energy or momentum would preclude normalization); its development in time is governed by the free Schrödinger equation. In the presence of an obstacle, however, the actual time development of a physical wave packet is governed by a Schrödinger equation that includes the interaction force. Thus, a physical wave packet $\Psi_+(t)$ is uniquely determined by the full hamiltonian that governs its time development and by the freely developing packet $\Psi_{in}(t)$ from which it asymptotically originated. A physical wave packet $\Psi_-(t)$ can equally well be defined that develops according to the same full hamiltonian but that will asymptotically in the far future become equal to the free packet $\Psi_{out}(t)$ with specified quantum numbers. The quantum-mechanical probability amplitude for finding a particle with the quantum numbers, energy, and momentum described by Ψ_{out} in the far future, given that in the distant past it had those specified by Ψ_{in}, is therefore equal to the probability amplitude that the particle in the state $\Psi_+(t)$ now is in the state $\Psi_-(t)$, which is the inner product of $\Psi_+(t)$ and $\Psi_-(t)$. This amplitude may also be expressed as the matrix element corresponding to Ψ_{in} and Ψ_{out} of the scattering operator S, which maps the initial free state onto the final free state into which it develops.

If the situation is idealized to a sharp energy E in the initial free state (in which case the wave packet cannot be normalized), the energy will also be equal to E in the final free state; this is the quantum version of the conservation of energy. The S matrix is then defined after factoring out the energy-conserving Dirac delta function. The resulting matrix has partly discrete indices (for the discrete quantum numbers) and partly continuous ones (for the momenta). Since the number of particles in nonrelativistic quantum mechanics is conserved, the norm of the final state must equal that of the initial one, and no possible final states are excluded. This implies that the S matrix is unitary. (In the case of wave scattering, the unitarity property is the result of the conservation of energy.) *See* CONSERVATION OF ENERGY; MATRIX THEORY.

The quantity of greatest physical interest is the probability amplitude for the presence of the target to cause scattering in a given direction. This, up to a kinematical factor, is the scattering or reaction amplitude: it is the deviation of the S matrix from the unit matrix (no scattering). The unitarity of the S matrix manifests itself as the optical theorem, which expresses the imaginary part of the scattering

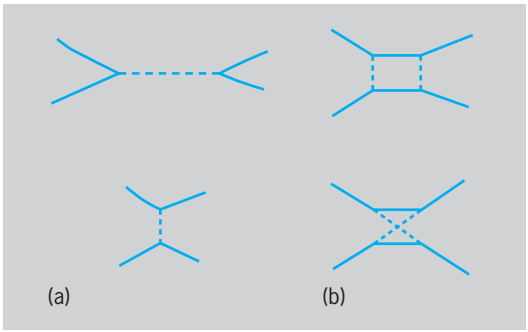
amplitude for forward scattering in terms of the total cross section.

Invariance properties of the physical system result in symmetries of the S matrix. An important example is the reciprocity relation, which, on the basis of invariance under a reversal of the sense of time, equates the scattering from the direction $\vec{k}$ to $\vec{k}'$ to the scattering from $-\vec{k}$ to $-\vec{k}'$. *See* SYMMETRY LAWS (PHYSICS); TIME REVERSAL INVARIANCE.

The mathematical properties of the S matrix in nonrelativistic quantum mechanics have been thoroughly studied and are, for the most part, well understood. If the potential energy in the Schrödinger equation, or the scattering obstacle, is spherically symmetric, the eigenfunctions of the S matrix are spherical harmonics and its eigenvalues are of the form $\exp(2i\delta_l)$, where the real number δ_l is the phase shift of angular momentum l. In the nonspherically symmetric case, analogous quantities are called the eigenphase shifts, and the eigenfunctions depend on both the energy and the dynamics. In the relativistic regime, without an underlying Schrödinger equation for the particles, the mathematical properties are not as well known. Causality arguments (no signal should propagate faster than light) lead to dispersion relations, which constitute experimentally verifiable consequences of very general assumptions on the properties of nature that are independent of the detailed dynamics. *See* ANGULAR MOMENTUM; CAUSALITY; DISPERSION RELATIONS; EIGENFUNCTION; SPHERICAL HARMONICS.

Calculation. In the nonrelativistic scattering of one particle by a fixed center of force, or of two particles by one another in their center-of-mass system, the S matrix is calculated by converting the Schrödinger equation with the scattering boundary condition into an integral equation known as the Lippmann-Schwinger equation. If the potential energy is invariant under rotation, angular momentum is conserved, each phase shift can be calculated separately by a separation of variables in the Schrödinger equation, and the S matrix can be computed as a Legendre series, which is known as the partial-wave series. For particles interacting via spin-dependent forces, the procedure is similarly based on solving systems of coupled Schrödinger equations that describe the spin-induced coupling between orbital angular momenta. Under certain assumptions on the structure of the forces, purely algebraic methods may be utilized for the calculation of S-matrix elements. In the more general case of noncentral forces, they have to be computed numerically from the solution of the three-dimensional Schrödinger equation. In the calculation of the S matrix for three or more interacting particles, the Lippmann-Schwinger equation does not determine the answer uniquely. This problem was overcome by the Faddeev equations and other subsequently developed methods. *See* LEGENDRE FUNCTIONS.

Actual calculations almost always must resort to approximation methods. These include low-energy approximations based on the rapid convergence of the partial-wave series at low energies; high-energy approximations analogous to ray optics with diffraction (WKB approximation); perturbation theory based on assumed weakness of the forces (Born approximation) or on a small deviation from forces for which the results are known (distorted-wave Born approximation); variational methods; and procedures based on physical intuition and insight. *See* PERTURBATION (QUANTUM MECHANICS); WENTZEL-KRAMERS-BRILLOUIN METHOD.



Typical Feynman diagrams for two-particle scattering. The solid lines are fermions, the broken lines bosons. (*a*) Second-order diagrams. (*b*) Fourth-order diagrams.

In quantum field theory, the S matrix, when it is practical to calculate it at all, is most commonly calculated by means of Feynman diagrams (see **illus.**). Each such diagram, which schematically represents the so-called virtual emission and reabsorption of force-transmitting bosons (in quantum electrodynamics, photons) can be translated by a general rule into a multiple integral that usually must be computed numerically. In the case of string theory, each line in such a diagram becomes a two-dimensional sheet. In all realistic cases, many of the integrals diverge and the S matrix has to be renormalized, a procedure, first introduced in quantum electrodynamics, in which a finite number of quantities that are infinite in the perturbation calculation are replaced by their experimental values. The strength of the coupling that describes the strong nuclear forces, or of the weak nuclear force at high energies, is such that straightforward expansions in terms of Feynman diagrams are impractical and it is necessary to resort to partial summations of infinitely many of them. In some cases, the requirement of unitarity of the S matrix is an important constraint. *See* FEYNMAN DIAGRAM; QUANTUM CHROMODYNAMICS; QUANTUM ELECTRODYNAMICS; QUANTUM FIELD THEORY; RENORMALIZATION; WEAK NUCLEAR INTERACTIONS.

Another method of calculation is to replace the continuous physical space-time by a lattice of discrete points. In that case, no infinities arise, but very large computers are required for meaningful results. The problem of approaching the continuum limit is difficult and not completely solved.

Inverse scattering. The inverse-scattering problem is to determine, if possible, the underlying force (or the shape of an obstacle) from the S matrix, which is assumed to be known from experiments. It was first solved for central potentials in the one-particle

Schrödinger equation on the assumption that what is known is one phase shift for all energies and the binding energies and so-called norming constants of the same angular momentum. The problem for central potentials and all phase shifts known at one energy was subsequently solved. The inverse-scattering problem for noncentral forces, in which the S matrix is given as a function of the energy and the directions of the incoming and scattered particles, is partially solved. *See* INVERSE SCATTERING THEORY.

Roger G. Newton

Bibliography. W. O. Amrein et al., *Scattering Theory in Quantum Mechanics*, 1977; K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory*, 1989; E. Leader and E. Predazzi, *An Introduction to Gauge Theories and the "New Physics,"* 1982; R. G. Newton, *Scattering Theory of Waves and Particles*, 2d ed., 1982; A. M. Polyakov, *Gauge Fields and Strings*, 1987; M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, vol. 3: *Scattering Theory*, 1979.

Scattering of electromagnetic radiation

The process in which energy is removed from a beam of electromagnetic radiation and reemitted with a change in direction, phase, or wavelength. All electromagnetic radiation is subject to scattering by the medium (gas, liquid, or solid) through which it passes. In the short-wavelength, high-energy regime in which electromagnetic radiation is most easily discussed by means of a particle description, these processes are termed photon scattering. At slightly longer wavelengths, the scattering of x-rays provides the most effective means of determining the structure of crystalline solids. In the visible wavelength region, scattering of light produces the blue sky, red sunsets, and white clouds. At longer wavelengths, scattering of radio waves determines their characteristics as they pass through the atmosphere. *See* LIGHT; METEOROLOGICAL OPTICS; RADIO-WAVE PROPAGATION.

It has been known since the work of J. Maxwell in the nineteenth century that accelerating electric charges radiate energy and, conversely, that electromagnetic radiation consists of fields which accelerate charged particles. Light in the visible, infrared, or ultraviolet region interacts primarily with the electrons in gases, liquids, and solids—not the nuclei. The scattering process in these wavelength regions consists of acceleration of the electrons by the incident beam, followed by reradiation from the accelerating charges. *See* ELECTROMAGNETIC RADIATION.

Scattering processes may be divided according to the time between the absorption of energy from the incident beam and the subsequent reradiation. True "scattering" refers only to those processes which are essentially instantaneous. Mechanisms in which there is a measurable delay between absorption and reemission are usually termed luminescence. If the delay is longer than a microsecond or so, the process may be called fluorescence; and mechanisms involving very long delays (seconds) are usually termed phosphorescence. *See* ABSORPTION OF ELECTROMAGNETIC RADIATION; FLUORESCENCE; LUMINESCENCE; PHOSPHORESCENCE.

Inelastic scattering. Instantaneous scattering processes may be further categorized according to the wavelength shifts involved. Some scattering is "elastic"; there is no wavelength change, only a phase shift. In 1928 C. V. Raman discovered the process in which light was inelastically scattered and its energy was shifted by an amount equal to the vibrational energy of a molecule or crystal. Such scattering is usually called the Raman effect. This term has been used in a more general way, however, often to describe inelastic scattering of light by spin waves in magnetic crystals, by plasma waves in semiconductors, or by such exotic excitations as "rotons," the elementary quanta of superfluid helium. *See* RAMAN EFFECT.

Brillouin and Rayleigh scattering. In liquids or gases two distinct processes generate inelastic scattering with small wavelength shifts. The first is Brillouin scattering from pressure waves. When a sound wave propagates through a medium, it produces alternate regions of high compression (high density) and low compression (or rarefaction). A picture of the density distribution in such a medium is shown in **Fig. 1**. The separation of the high-density regions is equal to one wavelength λ for the sound wave propagating. Brillouin scattering of light to higher (or lower) frequencies occurs because the medium is moving toward (or away from) the light source. This is an optical Doppler effect, in which the frequency of sound from a moving object is shifted up in frequency as the object moves toward the observer. *See* DOPPLER EFFECT.

The second kind of inelastic scattering studied in fluids is due to entropy and temperature fluctuations. In contrast to the pressure fluctuations due to sound waves, these entropy fluctuations do not generate scattering at sharp, well-defined wavelength shifts from the exciting wavelength; rather, they produce a broadening in the scattered radiation centered about the exciting wavelength. This is because entropy fluctuations in a normal fluid are not propagating and do not, therefore, have a characteristic frequency; unlike sound they do not move like

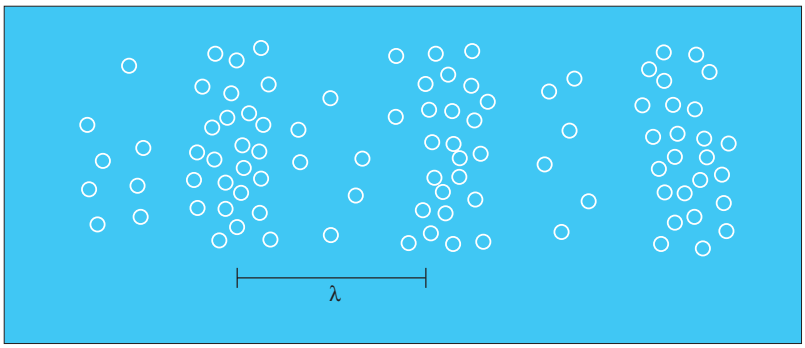


Fig. 1. Pressure-induced density fluctuations in a fluid.

waves through a liquid—instead, they diffuse. Under rather special circumstances, however, these entropy or temperature fluctuations can propagate; this has been observed in super-fluid helium and in crystalline sodium fluoride at low temperatures, and is called second sound. *See* ENTROPY; SECOND SOUND.

Scattering from entropy or temperature fluctuations is called Rayleigh scattering. In solids this process is obscured by scattering from defects and impurities. Under the assumption that the scattering in fluids is from particles much smaller than the wavelength of the exciting light, Lord Rayleigh derived in 1871 an equation, shown below, for such scattering; here $I(\theta)$ is the intensity of light scattered from an incident beam of wavelength λ and intensity I_0 at a distance r; d is the number of scattering particles; υ is the volume of the disturbing particle; and n is the index of refraction of the fluid. The cos θ term is present for unpolarized incident light, where θ is the scattering angle. Measurement of the ratio in Rayleigh's equation allows the determination of either Avogadro's number N or the molecular weight M of the fluid, if the other is known. The dependence of scattering intensity upon the inverse fourth power of the wavelength given in Rayleigh's equation is responsible for the fact that daytime sky looks blue and sunsets red: blue light is scattered out of the sunlight by the air molecules more strongly than red; at sunset, more red light passes directly to the eyes without being scattered.

$$\frac{r^2 I(\theta)}{I_0} = \pi d \lambda^{-4} \upsilon^2 (1 + \cos^2 \theta)(n - 1)^2$$

Large particles. Rayleigh's derivation of his scattering equation relies on the assumption of small, independent particles. Under some circumstances of interest, both of these assumptions fail. Colloidal suspensions provide systems in which the scattering particles are comparable to or larger than the exciting wavelengths. Such scattering is called the Tyndall effect and results in a nearly wavelength-independent (that is, white) scattering spectrum. The Tyndall effect is the reason clouds are white (the water droplets become larger than the wavelengths of visible light). *See* TYNDALL EFFECT.

The breakdown of Rayleigh's second assumption—that of independent particles—occurs in all liquids. There is strong correlation between the motion of neighboring particles. This leads to fixed phase relations and destructive interference for most of the scattered light. The remaining scattering arises from fluctuations in particle density discussed above and was first analyzed theoretically by A. Einstein in 1910 and by M. Smoluchowski in 1908.

Rayleigh's basic theory has been extended by several authors. Rayleigh in 1911 and R. Gans in 1925 derived scattering formulas appropriate for spheres of finite size, and in 1947 P. Debye extended the theory to include random coil polymers. The combined results of these three workers, generally called the Rayleigh-Gans-Debye theory, is valid for any size of particle, provided the refractive index n is near unity; whereas the Rayleigh theory is valid for any index, provided the particles are very small. The index n of any medium would be unity if there were no scattering.

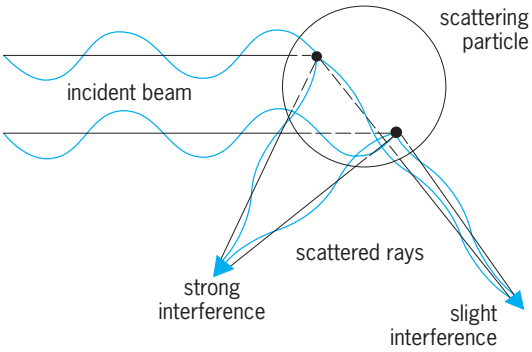


Fig. 2. Role of interference in determining the angular dependence of light scattering from particles that are comparable in size to the wavelength.

A more complete theory than Rayleigh-Gans-Debye was actually developed earlier by G. Mie in 1908; however, Mie's theory generally requires numerical solution. Mie's theory is valid even for particles larger than the wavelength of light used. For such particles large phase shifts occur for the scattered light. Mie scattering exhibits several maxima and minima as a function of scattering angle; the positions of these maxima depend upon particle size, as indicated in **Fig. 2**. These secondary maxima are essentially higher-order Tyndall scattering.

Critical opalescence. The most striking example of light scattering is that of critical opalescence, discovered by T. Andrews in 1869. As the critical temperature of a fluid is approached along the critical isochore, fluctuations in the density of the medium become larger than a wavelength and stay together for longer times. The independent particle approximation becomes very bad, and the intensity of light scattered in all directions becomes very large. The fluid, which might have been highly transparent at a temperature 1°C (1.8°F) higher, suddenly scatters practically all of the light incident upon it. The regions of liquidlike density in the fluid have molecules moving in phase, or "coherently," within them. Such coherent regions in fluids near their critical temperatures behave very much like the large spherical particles diagrammed in Fig. 2, thus establishing the relationship between Mie scattering of colloids and critical opalescence. The theory of scattering from correlated particles was developed primarily by L. S. Ornstein and F. Zernike in 1914–1926. *See* COHERENCE; CRITICAL PHENOMENA; OPALESCENCE.

Nonlinear scattering. At the power densities available with pulsed lasers, a variety of other scattering mechanisms can become important. (For nonlaser excitation these processes exist but are generally too weak to be detected.) Such processes include second-harmonic generation, in which the energy from two quanta (photons) in the incident beam produces one doubly energetic photon. A practical example is the production of green light from

the red light of a ruby laser. Third-harmonic generation (and similar higher-order processes) also occurs as an intense scattering mechanism at sufficiently high power densities. Whereas second-harmonic generation can occur only in certain crystals, for reasons of symmetry, third-harmonic generation occurs in any medium. Other more exotic nonlinear scattering mechanisms include sum generation and difference generation, in which two beams of light are combined to produce scattering having frequencies which are the sum or difference of the frequencies in the two incident beams, and rather weak processes such as the hyper-Raman effect, in which the scattered light has a frequency which is twice that of the incident beam plus (or minus) some vibrational energy. These effects are nonlinear because their intensities vary according to the square or higher power of the incident intensity. These nonlinear processes may be very important in astrophysics (for example, in stellar interiors), in addition to laboratory laser experiments. *See* NONLINEAR OPTICS. J. F. Scott

Bibliography. B. J. Berne and R. Pecora, *Dynamic Light Scattering*, 1976, reprint 1990; W. Brown (ed.), *Dynamic Light Scattering*, 1993; W. Hayes and R. Loudon, *Scattering of Light by Crystals*, 1978; K. S. Schmitz, *An Introduction to Dynamic Light Scattering by Macromolecules*, 1990.

Scent gland

A specialized skin gland of the tubuloalveolar or acinous variety found in many mammals. These glands produce substances having peculiar odors. In some instances they are large, in others small. Examples of large glands are the civet gland in the civet cat, the musk gland in the musk deer, and the castoreum gland in the beaver. The civet gland is an anal gland, whereas the musk and castoreum are preputial. Examples of small scent glands are the preputial or Tyson's glands in the human male which secrete the smegma, and the vulval glands in the female. The secretions in all of the above glands are sebaceous.

Many other vertebrates have glands whose secretions give off various types of odors. The mucus-secreting skin glands of fishes produce their fishy odor. Amphibia have glands which emit pungent, sweetish, or onionlike odors. Many urodeles have specialized courtship or hedonic glands. These amphibian glands are mucus-secreting. Scent glands are protective devices for many animals; in some species they serve to attract members of the same species or the opposite sex. The femoral glands on the inner aspect of the upper region of the hindlimbs of male lizards are specialized sebaceous glands associated with copulation. They give off a musty odor. The uropygial or preen glands opening upon the upper tail surface of birds produce an odorous, oily material used to waterproof feathers. *See* EPITHELIUM; GLAND; UROPYGIAL GLAND. Olin E. Nelsen

Scheduling

A decision-making function that plays an important role in most manufacturing and service industries. Scheduling is applied in procurement and production, in transportation and distribution, and in information processing and communication. A scheduling function typically uses mathematical optimization techniques or heuristic methods to allocate limited resources to the processing of tasks.

Application domains. There are many types of scheduling problems. The main categories include project scheduling, production or job shop scheduling, timetabling or rostering, and work-force scheduling. Each one of these categories can be subdivided.

Project scheduling is concerned with a set of activities that are subject to precedence constraints, specifying which jobs have to be completed before a given job is allowed to start its processing. All activities belong to a single (and typically large) project that has to be completed in a minimum time; for example, a large real estate development or the construction of an aircraft carrier.

Production or job shop scheduling is important in manufacturing settings, for example, semiconductor manufacturing. Customer orders have to be executed. Each order entails a number of operations that have to be processed on the resources or the machines available. Each order has a committed shipping date that plays the role of a due date. Production scheduling often also includes lot sizing and batching.

Timetabling occurs often in class room scheduling, scheduling of meetings, and reservation systems. In many organizations, especially in the service industries, meetings must be scheduled in such a way that all necessary participants are present; often other constraints have to be satisfied as well (in the form of space and equipment needed). Such problems occur in schools with classroom and examination scheduling as well as in the renting of hotel rooms and automobiles.

Work-force scheduling (crew scheduling, and so on) is increasingly important, especially in the service industries. For example, large call centers in many types of enterprises (airlines, financial institutions, and others) require the development of complicated personnel scheduling techniques.

Modeling. In order to determine satisfactory or optimal schedules, it is helpful to formulate the scheduling problem as a mathematical model. Such a model typically describes a number of important characteristics. One characteristic specifies the number of machines or resources as well as their interrelationships with regard to the configuration, for example, machines set up in series, machines set up in parallel. A second characteristic of a mathematical model concerns the processing requirements and constraints. These include setup costs and setup times, and precedence constraints between various activities. A third characteristic has to do with the objective that has to be optimized, which may be a

single objective or a composite of different objectives. For example, the objective may be a combination of maximizing throughput (which is often equivalent to minimizing setup times) and maximizing the number of orders that are shipped on time.

These mathematical models can often be formulated as mathematical programs, for example, linear programs, integer programs, or disjunctive programs. The mathematical programs typically exhibit a special structure that can be exploited when solving them. For example, the coefficient matrices of the integer programs used to formulate work-force scheduling problems typically consist of only zeroes and ones.

Computational complexity and algorithms. Most scheduling problems in theory as well as in practice are extremely hard to solve. In the theory of computational complexity, the hard problems are referred to as NP-hard problems (NP for nondeterministic polynomial time), whereas the easy problems are said to be polynomial time solvable. NP-hard implies that the computation time needed to find an optimal solution increases exponentially and not polynomially with the size of the instance of the problem (the size of an instance of a given problem could, for example, be measured by the number of jobs that have to be scheduled). The computation time needed to find an optimal solution for a problem that is polynomial time solvable increases polynomially in the size of an instance. The computational complexity of scheduling problems, that is, whether they are polynomial time solvable or NP-hard, can often be determined fairly easily. The majority of the scheduling problems studied in the literature are NP-hard.

Because of the many different types of scheduling problems and also because of the fact that any given problem type has instances of varying sizes, there are many different types of algorithms. The types can be categorized in several ways. One way is based on whether the algorithm is designed to find an optimal solution or whether it is designed to find just a good solution. An algorithm of the first type may have to consider all possible schedules and may be enumerative, whereas an algorithm of the second type is typically based on heuristics (algorithmic procedures that are based on common sense, but which do not guarantee an optimal solution).

A different way of categorizing the algorithms is based on the following: Some algorithms are of the constructive type, that is, they generate schedules one job at the time; whereas other algorithms are of the improvement type, that is, they start out with a complete schedule and then try to improve on it step by step, performing certain manipulations.

A class of algorithms that is widely used in practice consists of the so-called dispatching rules, typically heuristics that generate schedules constructively. A dispatching rule is a priority rule that sets priorities for the jobs available for processing according to a given function of the attributes of the jobs. A well-known priority rule is the so-called weighted shortest processing time first rule. This rule orders the jobs in a way that takes into account the processing time p of a job as well as the weight (or importance factor) w. The rule gives the job with the highest ratio of w/p the highest priority.

Another widely used class of algorithms uses so-called branch-and-bound procedures. Branch-and-bound is an established way of carrying out complete enumeration. An algorithm of this type is a constructive procedure designed to find an optimal solution.

A third class of algorithms consists of local search procedures. A local search routine is a heuristic of the improvement type. It starts out with a complete schedule that is feasible. Then it attempts to find better schedules through a series of well-defined manipulations, for example, adjacent pairwise interchanges. The three-best known types of local search procedures are simulated annealing, tabu search, and genetic algorithms. Simulated annealing attempts to find a better schedule than the best schedule generated so far by selecting a schedule in the neighborhood or proximity of the current schedule. Simulated annealing performs this neighborhood search in a random manner. Tabu search is similar to simulated annealing, but performs its neighborhood search in a nonrandom manner. A genetic algorithm selects, each time, multiple new schedules simultaneously.

Many algorithms that are implemented in practice are hybrids. For example, they may use a dispatching rule to find an initial feasible schedule and then feed this solution into a local search procedure that attempts to find solutions that are better than the initial solution.

Scheduling system design. The scheduling function is often incorporated in a system that is embedded in the information infrastructure of the organization. This infrastructure may be an enterprise-wide information system that is connected to the main databases of the company. Many other decision support systems may be plugged into such an enterprise-wide information system—for example, forecasting, order promising and due date setting, and material requirements planning (MRP).

The scheduling system may rely on its own piece of hardware, which may be a workstation with a monitor that has special graphics capabilities. The scheduling system may have its own local database that is linked with other databases in the company and possibly also with the Internet.

The database that the scheduling system relies on usually has some special characteristics. It has static data as well as dynamic data. The static data—for example, processing requirements, product characteristics, and routing specifications—are fixed and do not depend on the schedules developed. The dynamic data are schedule-dependent; they include the start times and completion times of all the operations on all the different machines, and the length of the setup times (since these may also be schedule-dependent).

The user interfaces of scheduling systems may often be very elaborate. They almost always include an electronic Gantt chart (horizontal bar chart displaying the time relationships between the different tasks in a project), which allows the user to click on operations, drag them, and drop them at other

machines and in other time periods. The system often has elaborate capabilities of readjusting itself after considering all the changes made by the scheduler.

The scheduling engine often has a library of algorithms. Such a library may contain priority rules and local search techniques, as well as more specialized techniques that take the bottleneck dynamics of the environment into account.

The scheduling engine often allows the user to provide some input. For example, the user may specify the amount of time (in seconds) that an algorithm, for example, a local search procedure, is allowed to run. The user may provide even more input such as specifying how one algorithm should be coupled with another. For example, first a priority rule is used in order to generate a schedule, and then this schedule is fed into a local search procedure that tries to find a better schedule.

Economic impact. The economic impact of scheduling is significant. In certain industries the viability of a company may depend on the effectiveness of its scheduling systems, for example, airlines and semiconductor manufacturing. Good scheduling often allows an organization to conduct its operations with a minimum of resources. In the case of a large airline, it may be possible to free-up the use of a plane just by more efficient scheduling. If a manufacturing company has established a reputation for on-time deliveries of customer orders, the company can charge a premium to its customers (which may be an additional 2–3%). *See* MATERIAL RESOURCE PLANNING; PRODUCTION PLANNING.

Michael Pinedo; Sridhar Seshadri

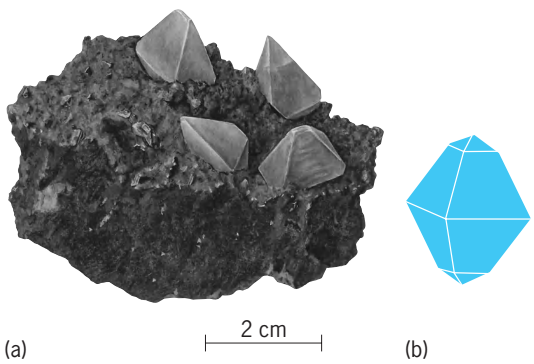
Bibliography. M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 1995; M. Pinedo and X. Chao, *Operations Scheduling with Applications in Manufacturing and Services*, 1999.

Scheelite

A mineral consisting of calcium tungstate, $CaWO_4$. Scheelite occurs in colorless to white, tetragonal crystals (see **illus.**); it may also be massive and granular. Its fracture is uneven, and its luster is vitreous to adamantine. Scheelite has a hardness of 4.5–5 on Mohs scale and a specific gravity of 6.1. Its streak is white. The mineral is transparent and fluoresces bright bluish-white under ultraviolet light.

Scheelite may contain small amounts of molybdenum. It is an important tungsten mineral and occurs principally in contact metamorphosed deposits (tactite) associated with garnet, diopside, tremolite, epidote, wollastonite, sphene, molybdenite, and fluorite, with minor amounts of pyrite and chalcopyrite. It also occurs in small amounts in vein deposits. The most important scheelite deposit in the United States is near Mill City, Nevada. *See* TUNGSTEN.

Edward C. T. Chao



Scheelite. (*a*) Crystals with chalcopyrite ore from Piedmont, Italy (*specimen from Department of Geology, Bryn Mawr College*). (*b*) Crystal habit (*after C. Klein and C. S. Hurlbut, Jr., Manual of Mineralogy, 21st ed., John Wiley and Sons, 1993*)

Schematic drawing

Concise, graphical symbolism whereby the engineer communicates to others the functional relationship of the parts in a component and, in turn, of the components in a system. The symbols do not attempt to describe in complete detail the characteristics or physical form of the elements, but they do suggest the functional form which the ensemble of elements will take in satisfying the functional requirements of the component. They are different from a block diagram in that schematics describe more specifically

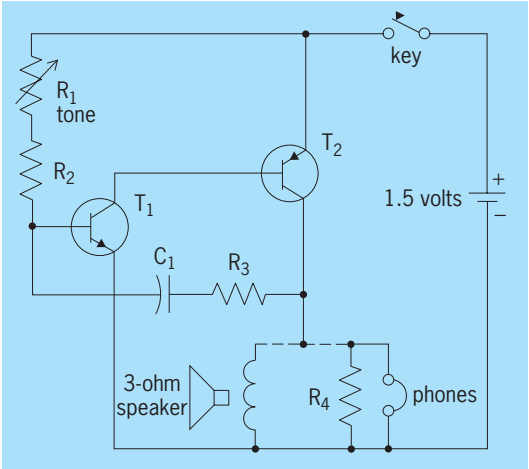


Fig. 1. Simple transistorized code practice oscillator, using standard symbols. (*After J. Markus, Sourcebook of Electronic Circuits, McGraw-Hill, 1968*)

the physical process by which the functional specifications of a block diagram are satisfied. Rather than expressing a mathematical relationship between, for example, an input and an output variable as in a block diagram, a schematic illustrates the physical principles and techniques by which the mathematical requirements of the element are realized. For instance, the schematic indicates whether electrical, hydraulic, mechanical, or pneumatic techniques are employed, and suitable symbols indicate the appropriate elements, such as batteries, resistors, valves, gearing, vacuum tubes, and motors.

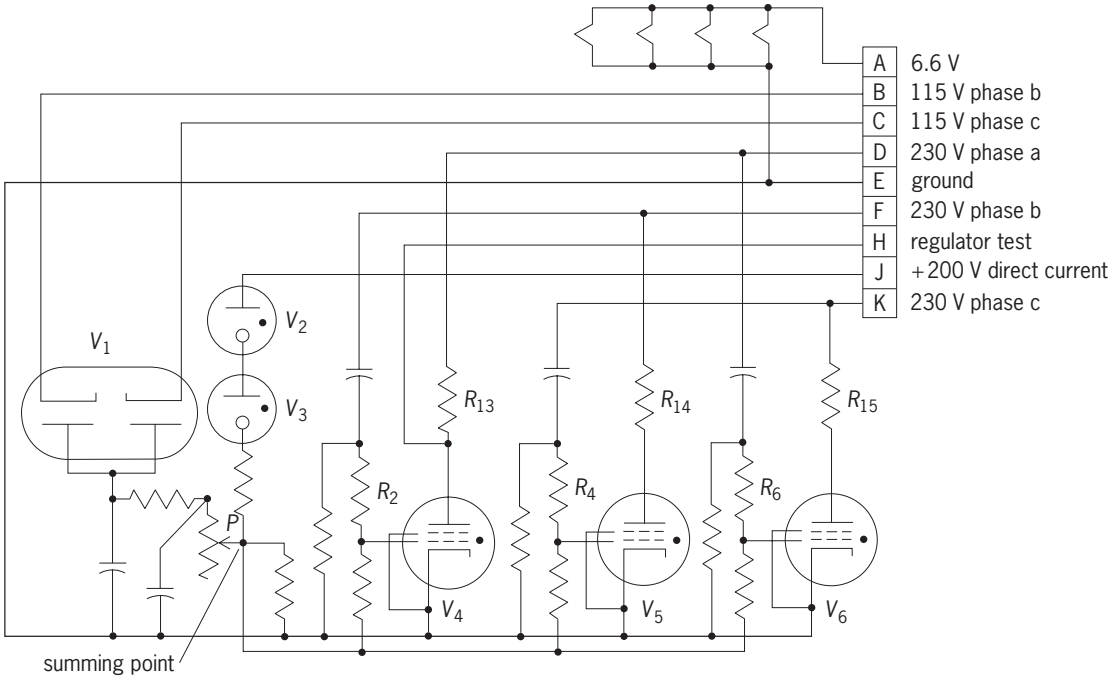
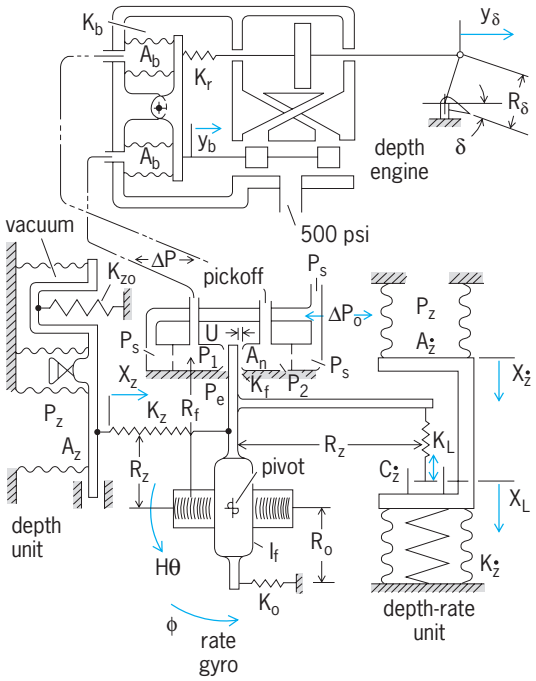


Fig. 2. Electrical schematic of voltage regulator.



Subscripts:

b = differential–pressure bellows
e = environment
f = flapper
L = depth-rate linkage
n = nozzles
o = ground, or reference
r = ram feedback
s = supply
z = depth unit
ż = depth-rate unit
δ = elevator

Fig. 3. Mechanical schematic of the depth-control mechanism of a torpedo.

Electrical schematic. An electrical schematic is a functional schematic which defines the interrelationship of the electrical elements in a circuit, equipment, or system. The symbols describing the electrical elements are stylized, simplified, and standardized to the point of universal acceptance (**Figs. 1** and **2**).

The simple character of the element symbol makes it possible to represent in a small area the interrelationship of the electrical elements in complex systems. This has the double advantage of economy of space and an increased facility of understanding, because one experienced in the symbolism can easily follow the various functional paths in the electrical schematic. The tracing of a signal path through an electrical schematic is considerably enhanced by the existence of more or less accepted rules with regard to the arrangement of the symbols and of the interconnections between the symbols, all contrived to make more lucid the functional interrelationship of the elements.

Mechanical schematic. A mechanical schematic is also a functional schematic. The graphical descriptions of elements of a mechanical system are more complex and more intimately interrelated than the symbolism of an electrical system and so the graphical characterizations are not nearly as well standardized or simplified (**Fig. 3**). However, a mechanical schematic illustrates such features as components, acceleration, velocity, position force sensing, and viscous damping devices. The symbols are arranged in such a manner and with such simplification as to economize on space and to facilitate an understanding of the functional interrelationship of the various components which make up the system. *See* DRAFTING; ENGINEERING DRAWING.

Robert W. Mann

Schist

Medium- to coarse-grained, mica-bearing metamorphic rock with well-developed foliation termed schistosity. Schist is derived primarily from fine-grained, mica-bearing rocks such as shales and slates. The schistosity is formed by rotation, recrystallization, and new growth of mica; it is deformational in origin. The planar to wavy foliation is defined by the strong preferred orientation of platy minerals, primarily muscovite, biotite, and chlorite. The relatively large grain size of these minerals (up to centimeters) produces the characteristic strong reflection when light shines on the rock. *See* BIOTITE; CHLORITE; MUSCOVITE.

The metamorphic grade of schists typically is in the middle to upper part (~660–930°F or ~350–500°C) of the greenschist facies and the lower part (~930–1100°F or ~500–600°C) of the amphibolite facies. In the amphibolite facies, micas react to form the mineral feldspar, which results in a coarsely foliated rock type called a gneiss. The wavy foliation pattern that is common in many schists reflects the presence of large, secondary minerals (such as garnet, staurolite, and chloritoid), which are grouped under the term porphyroblast. *See* AMPHIBOLITE; FACIES (GEOLOGY); FELDSPAR; GNEISS; PORPHYROBLAST; SHALE; SLATE.

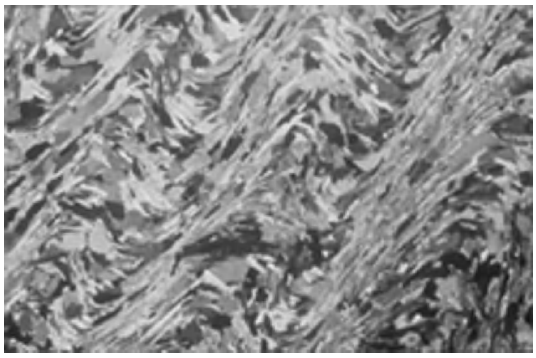
Schists are named by the assemblage of minerals that is most characteristic in the field; for example, a garnet-biotite schist contains porphyroblasts of garnet and a schistosity dominated by biotite. Under the microscope, schists commonly show a crenulation fabric (see **illus.**), reflecting the presence of an older foliation that may be depositional in origin or may represent an earlier deformation history. *See* PETROFABRIC ANALYSIS.

Schists can provide important information on the relationship between metamorphism, the occurrence of mineral reactions in response to changing pressure and temperature conditions, and deformation. When porphyroblasts grow, minerals that are either not involved in the reaction (accessory minerals such as graphite, zircon, and monazite) or minerals that are left over because the rock no longer contains the right mix of ingredients (such as quartz and feldspar) may become inclusions. These inclusions may form trails in the porphyroblast defining an internal foliation. The relationship between the internal foliation and the foliation in the matrix surrounding the blast, called the external foliation, gives information on the temporal relationship between deformation and metamorphism.

Predeformational metamorphism is characterized by an internal foliation whose shape is unrelated to the external foliation, and typically the two do not connect at the porphyroblast-matrix boundary. At the other extreme lies postdeformational metamorphism, in which an external foliation connects with the internal foliation within the grain without any disruption. For syndeformational growth, that is, when metamorphism and deformation overlap, the external foliation can be traced into a porphyroblast; but often the internal foliation is wavy to sigmoidal in shape. The classic example of the latter relationship is a snowball garnet. After the garnet nucleates, it rapidly grows by consumption and inclusion of minerals in the matrix. Simultaneously, the blast is affected by a shearing motion, much like ball bearings rotating between the hands. As the garnet continues to grow, it produces an internal foliation that appears to be rolled up, forming the characteristic snowball geometry. The rotation that is preserved from this inclusion pattern furthermore provides information on the type of deformation history of the rock (shear-sense indicator). *See* METAMORPHIC ROCKS; METAMORPHISM.

Ben A. van der Pluijm

Bibliography. B. A. van der Pluijm and S. Marshak, *Earth Structure: An Introduction to Structural Geology and Tectonics*, Ch. 11, 1997.



Photomicrograph of a mica schist from the Swiss Alps showing a crenulation fabric. The microscope image was taken under crossed polarized light with a field of view of ~1.5 mm across.

Schistosomiasis

A disease in which humans are parasitized by any of four species of blood flukes: *Schistosoma mansoni*, *S. haematobium*, *S. japonicum*, and *S. mekongi*. In contrast to other trematodes, the sexes are separate; that is, male and female reproductive systems occur in individual worms. The disease is also known as bilharziasis. *See* DIGENEA.

The approximate dimensions of the egg and male and female adult in each of the four species are shown in the **table**.

Dimensions of human blood flukes, *Schistosoma*

Species	Egg, μm	Adult	
		Length, cm	Diameter, mm
S. mansoni	150 × 65	Female 1–1.5	0.25
		Male 2	1
S. haematobium	150 × 60	Female 1–1.5	0.25
		Male 2	1
S. japonicum	80 × 65	Female 1–2	0.25
		Male 1–2.5	1
S. mekongi	45 × 40	Female 1.2	0.23
		Male 1.5	0.41

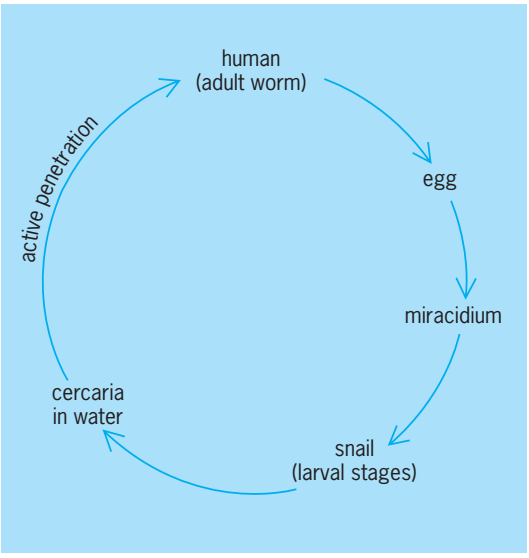
Distribution. *Schistosoma mansoni* is found in Africa, Brazil, Venezuela, Surinam, the Lesser Antilles, Puerto Rico, and the Dominican Republic. *Schistosoma haematobium*, originally found in Africa, the Near East, and the Mediterranean basin, was introduced into India during World War II. *Schistosoma japonicum* is widely spread in eastern Asia and the southwestern Pacific; however, in Taiwan it infects only animals, not humans. *Schistosoma mekongi* is reported from Laos and Cambodia.

Biology. An embryonated egg passed in feces or urine hatches in fresh water, liberating a miracidium. The ciliated miracidium penetrates into specific gastropod snails. There the miracidium is transformed into a mother sporocyst that gives rise by pedogenesis, or reproduction by larval stages, to several hundred daughter sporocysts. These move away to settle elsewhere in the snail and produce several thousand cercariae. The larval cycle lasts for about 1 month. The cercaria emerges from the mollusk, swims for up to 48 h, and penetrates the skin of the final host upon coming in contact with it. After a phase of growth in the lungs, the schistosomule moves to the intrahepatic veins for further development. Then it moves to its final habitat.

Schistosoma mansoni prefers to live in the inferior mesenteric vein of the lower bowel, particularly the branches of the sigmoid and rectum; *S. haematobium*, in the veins of the urinary bladder; *S. japonicum* and *S. mekongi*, in both the superior and inferior mesenteric veins. *Schistosoma mekongi* is related to *S. japonicum* but differs in that it has smaller eggs, a different intermediate host, and a longer prepatent period in the mammalian host.

For all schistosomes of humans, eggs start to appear within 2 months after exposure. The adult may live for many years. The generalized life cycle is shown in the **illustration**.

Epidemiology. As an intermediate host, *S. mansoni* uses snails of the genus *Biomphalaria*; *S. haematobium*, the genera *Bulinus and Physopsis*; *S. japonicum*, the genus *Onchomelania*; and *S. mekongi*, the genus *Tricula*. *Schistosoma mansoni* seems incapable of infecting *Oncomelania* ssp. *Schistosoma mansoni* is essentially a human parasite. Monkeys may act as natural reservoirs in Africa, and rodents in Brazil and Guadeloupe. *Schistosoma haematobium* has no reservoirs, while *S. japonicum* and *S. mekongi* possess many. Various animals may be infected in the laboratory with the four species.



Life cycle of the schistosomes.

Schistosomiasis is an agricultural hazard for humans of all ages in irrigated lands or swamps. Elsewhere fluvial waters are the main source of infection, in which case incidence is marked in human beings who are less than 15 years old, and is higher among boys than among girls. Since human activity is what causes the spread of the disease, it is epidemiologically sensible to say that schistosomiasis is caused by humans, not by snails.

Schistosomiasis is endemic in 74 developing countries, infects 200 million persons, and daily threatens 600 million. It persists as one of the most important parasitic diseases of humans.

Pathology. The disease changes in schistosomiasis may be described as follows.

Prepatent period. The penetration of the cercariae may or may not produce skin irritation. Heavy initial exposures result in inflammatory disease of the skin (urticaria), toxic symptoms, an increase in the size of the liver and spleen, and eosinophilia before egg laying starts.

Patent period. Schistosoma mansoni may lay 350 eggs daily, *S. japonicum*, 3000. About 10% reach the feces; the rest remain imprisoned in tissues, particularly the liver and intestine or urinary bladder, where they provoke fibrotic reaction (scars). Egg extrusion may result in bloody diarrhea or urine over a protracted period. The liver and spleen enlarge, and the lungs and spinal cord may become involved.

Postpatent period. The fibrotic changes may result eventually even when eggs are absent in the feces. The changes interfere with intestinal, hepatic, and bladder functions with serious consequences. Liver fibrosis leads to splenomegaly and esophageal varices. Death may occur from profuse hematemesis. Lung fibrosis may cause cardiac failure. Spinal cord lesions may result in paralysis.

Diagnosis. Rectal biopsy is the best definitive method to show the presence of eggs in the intestinal schistosomiasis. Bladder biopsy, or cystoscopy, is indicated in the urinary type. Fecal examination is of value in the patent period. Serological methods are of great complementary value; the complement-fixation test is the most sensitive, while the circumoval-precipitin one is highly specific. The method consists in incubating live eggs with blood serum of the suspect and searching for a hyaline, earlike precipitate protruding from the shell. Enzyme immunoassays have been developed that may prove superior to other methods as purified antigens become available. *See* COMPLEMENT-FIXATION TEST; IMMUNOASSAY.

Prevention. Measures to avoid contact with the infective stage are the only means of preventing

infection. In agricultural societies, this is virtually impossible. Latrines, rubber boots, skin repellants, education, and snail control have been tried, but with limited success. In more advanced industrialized countries, controlled urbanization has reduced exposure sites, with a consequent decrease in new infections.

Treatment. Chemotherapy has been difficult and dangerous, particularly in severe infections, but drugs have become available that promise effective control. Praziquantel, for example, with a 1-day oral administration is well tolerated and reportedly attains 95% cure of all schistosomes of humans at any stage of development. *See* EPIDEMIOLOGY; MEDICAL PARASITOLOGY. José F. Maldonado-Moll

Bibliography. P. C. Beaver and R. C. Jung, *Animal Agents and Vectors of Human Disease*, 5th ed., 1985; S. Cohen and K. S. Warren (eds.), *Immunology of Parasitic Infections*, 2d ed., 1982; M. Katz, D. D. Despommier, and R. W. Gadz, *Parasitic Diseases*, 1988; K. S. Warren, *Tropical and Geographical Medicine*, 1989.

Schistostegales

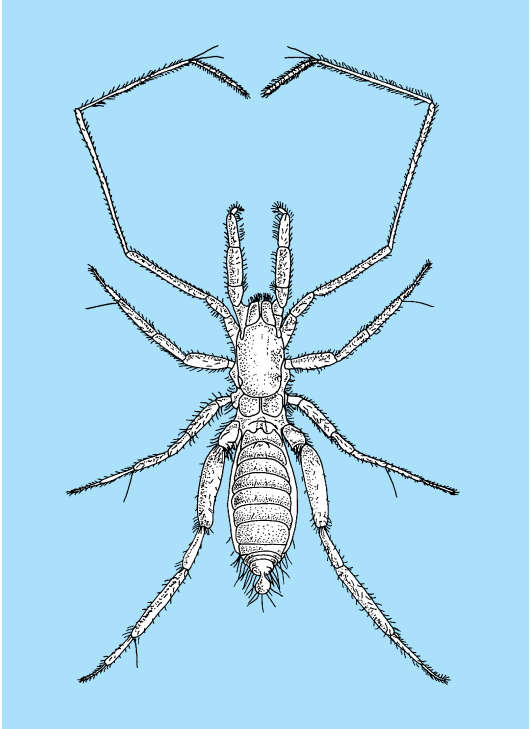
An order of the true mosses (subclass Bryidae), consisting of a single species, *Schistostega pennata*, the cave moss, which is especially characterized by its leaf arrangement and the form of its protonema. The plants grow in dimly lit, cavelike places, such as the undersides of upturned tree roots. The cave moss is widely distributed in north temperate regions. An unrelated moss in Australia and New Zealand, *Mittenia plumula*, has an identical protonema and habitat.

The protonema is much branched, with some branches consisting of spherical cells acting like lenses to concentrate weak rays of light on a backing of chloroplasts which reflect a greenish glow. The leafy plants are erect, unbranched, and dimorphous: sterile shoots have leaves wide-spreading in two rows and confluent because of broad basal decurrencies. Fertile plants have leaves in five rows and erect-spreading in a terminal tuft. All leaves are ecostate, entire, and unbordered, with thin-walled, laxly oblong-rhomboidal cells. The inflorescences are terminal, the setae elongate, and the capsules subglobose, with a flat operculum but no peristome. *See* BRYIDAE; BRYOPHYTA; BRYOPSIDA. Howard Crum

Bibliography. K. A. Lye, Studies in Norwegian bryophytes, 1: The family Schistostegaceae, *Lindbergia*, 1:205–213, 1972; H. Robinson, Observations on the origin of the specialized leaves of *Fissidens* and *Schistostega*, *Rev. Bryol. Lichen.*, 37:941–947, 1970.

Schizomida

An order of Arachnida with two families, Protoschizomidae (*Agastoschizomus*, *Protoschizomus*) and Schizomida (*Schizomus*, *Trithyreus*, *Megaschizomus*), including about 130 described species, mainly in the genus *Schizomus*. Schizomids occur



***Schizomus sawadai*. (*After K. Sekiguchi and T. Yamasaki, A redescription of "Trithyreus sawadi" (Uropygi: Schizomidae) from the Bonin Islands, Acta Arachnol., 24(2):73–81, 1972*)**

in most tropical and warm temperate regions, but each species has a restricted distribution. Many undescribed species probably occur in tropical leaf litter and caves. Schizomids are most closely related to Uropygi, with which they share several evolutionary novelties. *See* UROPYGI.

Schizomids are 0.12–0.44 in. (3–11 mm) in size, are white to tan, and lack eyes, although some retain vestigial "eyespots" (see **illus.**). The carapace is distinctively tripartite, and the abdomen has a short, terminal flagellum. Pedipalpal coxae are fused to form a trough, in which the animal digests its food externally. Pedipalps are raptorial, and first legs are elongated to serve as tactile organs. Pedipalps and flagellum are often sexually dimorphic.

Because schizomids are vulnerable to desiccation, they inhabit only moist places such as leaf litter, soil, caves, or beneath stones and logs. They are strict carnivores and fast runners. Sperm is transferred between sexes by a sperm capsule (spermatophore). The male deposits the spermatophore on the ground, the female grasps his modified flagellum in her chelicerae, and they advance in tandem until she contacts the spermatophore. Females apparently construct earthen brood chambers and stay with the young for a period of time. *See* ARACHNIDA. Jonathan A. Coddington

Bibliography. A. Kaestner, H. W. Levi, and L. R. Levi, *Invertebrate Zoology*, 2d ed., vol. 2, 1980; J. R. Reddell and J. C. Cokendolpher, Redescription of *Trithyreus grassii* (Arachnida: Schizomida: Schizomidae), *Oriental Insects*, 18:43–52, 1985;

J. M. Rowland, Breeding habits and early development of *Trithyreus pentapeltis* (Arachnida: Schizomida), *Entomol. News*, 83(3):69–74. 1972; J. M. Rowland and J. R. Reddell, The order Schizomida (Arachnida) in the New World, IV. *Goodnightorum* and *briggsi* groups and unplaced species (Schizomidae: *Schizomus*), *J. Arachnol.*, 9(1):19–46, 1981.

Schizophrenia

A brain disorder that is characterized by unusual mental experiences, such as hallucinations, changes in emotional experience, and severe decrements in social, cognitive, and occupational functioning. The popular belief that schizophrenia is characterized by multiple personalities is a misconception.

Schizophrenia is the most financially costly of all psychiatric conditions. Patients with schizophrenia use a disproportionately high percentage of mental health services. The prevalence of schizophrenia in the United States is approximately 1% of the population, but annual U.S. mental health care expenditures for its treatment have been estimated to be higher than 2.5% of the total cost of all health care in the country. Schizophrenia is one of the top five causes of disability for males and females, which is particularly striking since other causes of disability are far more common (for example, alcohol abuse and motor vehicle accidents). Schizophrenia is one of the top three most costly illnesses in the United States (following vascular illnesses and cancer). This high cost is underscored by the fact that many more Americans are affected annually by heart disease and cancer than schizophrenia. The indirect cost (lost wages and decreased productivity) of schizophrenia has been estimated to be approximately three times as great as the direct costs (medicine, hospitalization, and mental health care) of treating the illness. This article describes the characteristics of schizophrenia, its possible causes, and current treatments.

Diagnosis and Clinical Characteristics

Patients with schizophrenia demonstrate a series of biological differences when compared as a group with healthy individuals without schizophrenia. At present, however, there is no single biological marker available to indicate the presence of schizophrenia. A diagnosis is made on the basis of a cluster of symptoms reported by the patient and signs identified by the clinician. The most recent version of the *Diagnostic and Statistical Manual of Mental Disorders*, referred to as DSM-IV, includes criteria to help distinguish schizophrenia from other psychiatric disorders (see **table**).

Diagnostic criteria for schizophrenia (DSM-IV)

Criterion category	Description
Symptoms	Two or more of the following (one if hallucinations are characteristic or delusions are bizarre): delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms
Social/occupational dysfunction	Significant decline in functioning in areas such as work, interpersonal relations, or self-care
Duration	Continuous signs for at least 6 months, including 1 month of symptoms
Exclusions	Symptoms are not better explained by other disorders such as mood disorders, other brain disorders, substance abuse, autism or pervasive developmental disorder

Hallucinations. People with schizophrenia may report perceptual experiences in the absence of a real environmental stimulus. The most common of these is auditory hallucinations, most often reported in the form of words spoken to the person with schizophrenia. The voices are often derogatory in nature, and can be tremendously frightening, especially in people experiencing hallucinations for the first time. Visual hallucinations, such as of human faces, are also possible. Tactile hallucinations, such as the experience of something moving on the skin, are less likely. Hallucinations through the senses of smell (olfactory) and taste (gustatory) are rare but can occur in some patients. *See* HALLUCINATION.

Delusions. People with schizophrenia often maintain beliefs that are not held by the overwhelming majority of the general population. To be considered delusions, the beliefs must be unshakable. In many cases, these beliefs may be "bizarre." A bizarre delusion is defined as an idea that is completely impossible (for example, "a computer has been inserted into my skull"), in contrast to nonbizarre delusions that are plausible ("The FBI is watching me"), but still not true. Many of these bizarre beliefs stem from odd experiences. For instance, a person may report that new thoughts are placed inside his or her head via an outside force (thought insertion), or that his or her movements are being controlled by an outside agency, such as a satellite (passivity experiences). In some instances, the delusions have an element of suspiciousness to them, such as the incorrect belief that others are planning to cause the person with schizophrenia harm. The delusions may or may not be related to hallucinatory experiences.

Negative symptoms. Delusions and hallucinations are often referred to as "positive" symptoms, since they represent features of experience that are present in people with schizophrenia and absent in the general population. The phrase "negative" symptoms refers to human behavior that is found in most people, yet absent in people with schizophrenia. Included among these negative symptoms are social isolation, lack of motivation, lack of energy, slow or delayed speech, and diminished emotional expression, often referred to as "blunted affect."

Disorganization. People with schizophrenia may manifest an odd outward appearance due to disorganization. This presentation may include speech that does not follow logically or sensibly, at times to the point of being incoherent. Facial expression may be odd or inappropriate, such as laughing for no reason.

In some cases, people with schizophrenia may move in a strange and awkward manner. The extreme of this behavior, referred to as catatonia, has become very rare since pharmacological treatments became available.

Cognitive deficits. Although less striking than delusions and hallucinations, perhaps the most devastating feature of schizophrenia is the cognitive impairment found in most people with the disorder. On average, people with schizophrenia perform in the lowest 2–10% of the general population on tests of attention, memory, reasoning and problem solving, motor skills, and language abilities. Performance on tests of intelligence is lower than the population average, but not as impaired as these other cognitive domains. These cognitive deficits are perhaps the most important explanation for the difficulties that people with schizophrenia have functioning in everyday society. In addition, people with schizophrenia often have reduced awareness (referred to as "insight") regarding their illness. In fact, some may deny all symptoms of their disorder to the point that they will refuse treatment. *See* COGNITION.

Mood symptoms. Schizophrenia is associated with depression or bipolar disorder (manic-depressive illness) in about 10–15% of people with the disorder. The presence of both of these illnesses is referred to as schizoaffective disorder, and is generally associated with a more favorable outcome. In addition, as many as 50% or more of patients with schizophrenia experience episodes of depression that do not meet the full criteria for major depression. Up to 12% of patients with schizophrenia commit suicide, often during periods when they are depressed. *See* AFFECTIVE DISORDERS.

Functional impairment. People with schizophrenia are far less likely than the general population to work, marry, have offspring, and live independently. About 10–15% are able to sustain full-time employment, while 25–40% marry and have children. While the deinstitutionalization movements of the 1950s and the 1980s drastically reduced the number of people in long-term inpatient facilities, only a small minority of people diagnosed with schizophrenia in the United States are able to live without some form of public assistance. This functional impairment develops very early in the illness, and as many as 50% of patients with schizophrenia receive disability assistance within 6 months of their first treatment for the illness.

Course. The onset of schizophrenia is generally in the late teens to early twenties; however, onset is possible throughout the life span, including childhood, which is referred to as childhood schizophrenia, and in the later stages of life (after age 40), which is referred to as late-onset schizophrenia. Recent research on the early predictors of schizophrenia suggest that social isolation and cognitive deficits appear to be present years prior to the onset of delusions and hallucinations. While the onset of symptoms is abrupt in some people, others experience a more insidious process, including extreme social withdrawal, reduced motivation, mood changes, and cognitive and functional decline prior to the onset of full-blown schizophrenia symptoms. Recent interest has focused on the early indicators of impending psychosis, referred to as the "prodrome." If the presence of a prodrome could be reliably identified, then treatments for patients with schizophrenia could potentially be offered earlier and maybe even prevent the onset of illness in some cases.

Following the onset of illness, the course of schizophrenia is normally characterized by episodes of relative remission, in which only subtle residual and negative symptoms remain, and episodes of exacerbation of symptoms, which are often caused by failure to continue with treatment. Some long-term studies have suggested that some patients may show a reduced tendency to have exacerbations beyond their fifties.

Medical comorbidities. Patients with schizophrenia have a reduced life expectancy, above and beyond the statistical effects of suicide. Much of the increased mortality is due to cardiovascular illnesses related to smoking and obesity, both of which are more common in patients with schizophrenia than the population as a whole. Approximately 50–80% of people with schizophrenia smoke cigarettes, in contrast to the general population estimate of 25%. Due to their functional impairments and poverty, many patients with schizophrenia also receive substandard medical care.

Causes

The cause of schizophrenia has been a matter of concern and controversy for the last century. In a sense, the argument of "nature versus nurture" has been acted out repeatedly in this domain. It is likely that there are various forms of schizophrenia, perhaps all with different causes. Although schizophrenia appears to be inherited, at least in some cases, the influence of genes is far from complete. Although many arguments have been put forth regarding environmental factors that could cause schizophrenia, very few of these theories are consistently supported.

Genetics. Schizophrenia runs in families. The children of a parent with schizophrenia have a risk of developing the illness of about 10%, which is greater than 10 times the risk of developing schizophrenia with no relatives with the illness. This risk appears to be similar regardless of whether the children are raised by their parents with schizophrenia or are adopted away. At the same time, if one identical twin has schizophrenia there is only about a 50% chance that the other twin (who has exactly the same genes) will also be affected. These data suggest that a strictly genetic explanation is not adequate to describe the causes of the illness.

There have been extensive searches for genetic markers for the predisposition to develop schizophrenia. Other illnesses, such as Huntington's disease, which follow a pattern of inheritance associated with a strictly genetic transmission (50% of the children of any affected parent will develop the condition), have been found to have specific genetic markers. Recently, several genetic markers have been

found to be associated with schizophrenia. However, some of these influences appear similar in the healthy population and individuals with schizophrenia. It is likely, therefore, that each of these markers predispose a certain subgroup of the population to schizophrenia, but that a combination of genetic and environmental influences are needed to lead to the actual disorder. At this time, the most reasonable conclusion is that schizophrenia has a genetic component, albeit one that is not a classic pattern of single-gene inheritance, either recessive or dominant in nature. Research is continuing the search for better indicators of the genetic predisposition for schizophrenia. At present, too little is known to warrant the use of genetic counseling or other interventions. *See* BEHAVIOR GENETICS; HUMAN GENETICS.

Obstetrical complications. Several studies have found that the birth and intrauterine development of someone who eventually develops schizophrenia is more likely than normal to be associated with complications, such as forceps delivery or second-trimester maternal illness such as influenza. In addition, children of parents with schizophrenia who experience these complications are more likely than their siblings to manifest schizophrenia when they grow up. However, the number of documented cases of schizophrenia associated with pregnancy and birth complications in individuals without a schizophrenic relative is quite low. Thus, the simple experience of complicated pregnancy and delivery does not mean that any individual child has any meaningful risk for developing schizophrenia.

Family and social interaction. Many theories have tried to explain schizophrenia in terms of patterns of family interaction. There is no credible evidence that interaction patterns cause schizophrenia in someone who would not have developed it anyway. A small minority of mental health professionals still adhere to these theories, but they have been discarded by the majority since the late 1970s. It is clear, however, that social interaction patterns in the environment of individuals who already have schizophrenia can have an effect on their course of illness. High levels of criticism directed at newly discharged people with schizophrenia, like other stressors, is associated with increased frequency of reemergence of symptoms and rehospitalization.

Regional brain dysfunction. One of the new developments in the study of schizophrenia is associated with imaging technologies. Magnetic resonance imaging (MRI) allows for the visualization of the brain at high levels of resolution, allowing the identification of structures as small as 1 mm. People with schizophrenia often have changes in the structure of their brain, such as enlargement of the cerebral ventricles (fluid-filled spaces in the brain close to the midline). Various brain regions have been found to be smaller in patients with schizophrenia, including the frontal cortex, temporal lobes, thalamus, and the hippocampi. These brain structure abnormalities appear not to be isolated to the gray matter in the cortex, but may include the white matter that connects brain regions.

An additional development in the study of schizophrenia is the ability to study brain functions. Using MRI machines that take a rapid series of images of the brain, it is possible to capture patterns of blood flow in the brain. Since blood flow correlates with brain activity, it is possible to examine brain activity while people perform mental tasks. Studies of patients with schizophrenia have found patterns of abnormal activation of the brain while performing tests of memory and problem solving. While healthy individuals' brains use the frontal cortex to perform these tasks, patients with schizophrenia tend to have a less organized and coherent pattern of activation. Consistent with their generally less organized approach to learning new information and solving problems, brain activity itself in schizophrenia appears to be organized in a less efficient manner. *See* BRAIN; MEDICAL IMAGING.

Neurotransmitter changes. Individual cells in the brain (that is, neurons) communicate with each other using electrical-chemical means. These communications occur when different chemicals (called neurotransmitters) produced by the brain are released in proximity to other neurons. For years it has been suspected that neurotransmitter activity in schizophrenia is abnormal, because the medications used to treat schizophrenia act on dopamine receptors in the brain. Also, psychotic conditions can be caused by overdoses of medications that stimulate dopamine, such as amphetamine. The medications that are used to treat psychotic conditions such as schizophrenia act, in part, to reduce the activity of dopamine in the corpus striatum. This brain region is located at the juncture of the newer (cortex) and older (subcortical) parts of the brain.

Although dopamine is clearly implicated in schizophrenia, it is not simple overactivity of this neurotransmitter that causes schizophrenia. Measures of dopamine functions often do not suggest increased activity in patients with schizophrenia, and in some patients there is evidence of decreased activity. In addition, genes that influence levels of dopaminergic transmission have similar effects in healthy individuals and in patients with schizophrenia. Finally, dopamine interacts with several other neurotransmitters, meaning that the effects of dopamine might be indirect. With advances in technology that allow for the direct measurement of neurotransmitter activity in the central nervous system, more information in this area will be available in the near future. The current view of the dopamine hypothesis is that levels of transmitter activity are improperly balanced, with too little activation in the cortex and too much in the striatum. *See* DOPAMINE; NEUROBIOLOGY; NEURON.

Treatment

Both pharmacological and behavioral approaches are used to treat schizophrenia.

Pharmacological treatment. Chlorpromazine was the first pharmacological intervention to substantially reduce delusions and hallucinations in patients with schizophrenia. Several other similar medications were developed in the following decades. This class of medications, referred to as typical neuroleptics or conventional antipsychotics, act primarily to block dopamine receptors in the brain. These medications were able to have an impact on delusions and hallucinations in some patients, but treatment was accompanied by a variety of side effects, including sedation, muscular rigidity, and restlessness. Adjunctive medications such as benztropine are used to reduce these side effects, but such anticholinergic medications have side effects of their own, such as dry mouth, dizziness, and further cognitive impairment. Perhaps the worst side effect of typical antipsychotics is tardive dyskinesia, an irreversible movement disorder. The most effective dosage of these medications was initially believed to be very high. More recent studies have suggested that lower doses (for example, 2–4 milligrams per day of haloperidol) may be as effective with fewer side effects. The disadvantage of these medications is that they have minimal impact on cognitive deficits and negative symptoms such as social isolation. In the United States, only a small group of patients with schizophrenia receive treatment with these medications. In other areas of the world, particularly in economically disadvantaged countries, these medications continue to be the norm.

Since the approval of clozapine in the United States in 1988, clinicians have referred to a new group of medications as atypical neuroleptics or novel antipsychotics. The side effects associated with these medications are far different than the conventional antipsychotic medications, and in many cases are believed to be minimized. While clozapine had been available in Europe for years, it was banned in the United States due to the presence of a severe side effect, agranulocytosis (loss of white blood cells), which can be fatal. However, clozapine was the first medication that improved negative symptoms and cognitive deficits as well as delusions and hallucinations. Its approval in the United States was dependent upon a strict monitoring procedure that required patients to receive frequent blood tests if they were to continue on the medication.

Over the past decade, the Food and Drug Administration (FDA) approved several newer, atypical antipsychotic medications for use in the United States, including risperidone (1994), olanzapine (1996), quetiapine (1999), ziprasidone (2000), and aripiprazole (2002). The advantage of these medications is that they appear to have few of the side effects of the conventional antipsychotic medications, and they improve negative symptoms and cognitive function. The disadvantage of these medications is that they are more expensive than the conventional medications, whose patents have expired. The novel antipsychotic medications also have side effects. At higher doses, some of them have some of the side effects of the older medications, and many of them cause significant weight gain and diabetes. The efficacy and side-effect profiles have not been fully determined for the newest of these medications. *See* PSYCHOPHARMACOLOGY.

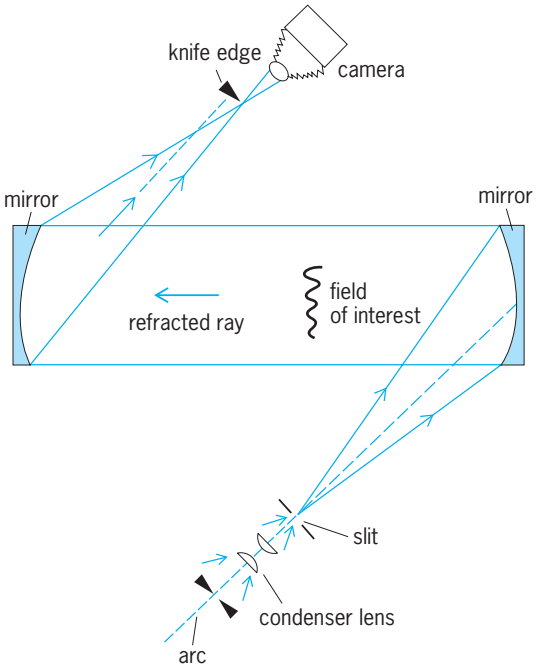
Behavioral treatment. There are several targets for behavioral treatments in schizophrenia. Patients with schizophrenia have difficulty acquiring skills in social, occupational, and independent living domains. Structured training programs have attempted to teach patients with schizophrenia how to function more effectively in these areas. Family interventions have also been designed to provide a supportive environment for patients with schizophrenia. These interventions have been demonstrated to reduce risk of relapse. Direct treatment of cognitive impairments, referred to as cognitive remediation, has also been implemented for patients with schizophrenia. While these interventions have not had tremendous success, their efficacy in combination with newer antipsychotics or other medications that potentially enhance cognition is promising. Teaching patients how to cope with their hallucinations and delusions is another target of current behavioral treatment. Most patients with schizophrenia do not spontaneously recognize their symptoms as unusual and unreal. Cognitive-behavioral treatments have been employed to help patients realize the nature of their symptoms and to develop plans for coping with them. Given the numerous difficulties confronted by people with schizophrenia, they often benefit greatly from the opportunity to discuss the exigencies of their daily life with someone who does not turn away from the unusual nature of their disorder. *See* PSYCHOTHERAPY.

Richard S. E. Keefe; Philip D. Harvey

Bibliography. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV), American Psychiatric Association, 1994; M. F. Green, *Schizophrenia Revealed*, Norton, New York, 2003; R. S. E. Keefe and P. D. Harvey, *Understanding Schizophrenia: A Guide to the New Research on Causes and Treatment*, Free Press, New York, 1994; E. F. Torrey, *Surviving Schizophrenia: A Manual for Families, Consumers, and Providers*, 4th ed., HarperCollins, New York, 2001; P. J. Weiden, *Breakthroughs in Antipsychotic Medications: A Guide for Consumers, Families, and Clinicians*, Norton, New York, 1999.

Schlieren photography

Any technique for the photographic recording of schlieren, which are regions or stria in a medium that is surrounded by a medium of different refractive index. Refractive index gradients in transparent media cause light rays to bend (refract) in the direction of increasing refractive index. This is a result of the reduced light velocity in a higher-refractive-index material. This phenomenon is exploited in viewing the schlieren, with schlieren photographs as the



Knife-edge method of viewing schlieren, employing the "z" configuration.

result. Modern schlieren systems are rarely limited to photography; electronic video recorders, scanning diode array cameras, and holography are widely used as supplements. *See* HOLOGRAPHY; OPTICAL DETECTORS; PHOTOGRAPHY; REFRACTION OF WAVES; SHADOWGRAPH.

Knife-edge method. There are many techniques for optically enhancing the appearance of the schlieren in an image of the field of interest. In the oldest of these, called the knife-edge method (see **illus.**), a point or slit source of light is collimated by a mirror and passed through a field of interest, after which a second mirror focuses the light, reimaging the point or slit where it is intercepted by an adjustable knife edge (commonly a razor blade is used). The illustration shows the "z" configuration which minimizes the coma aberration in the focus. Mirrors are most often used because of the absence of chromatic aberration. *See* ABERRATION (OPTICS).

Rays of light that are bent by the schlieren in the direction of the knife edge are intercepted and removed from the final image of the region of interest, causing those regions to appear dark. Consequently, the system is most sensitive to the density gradients that are perpendicular to the knife edge. The knife edge is commonly mounted on a rotatable mount so that it can be adjusted during a measurement to optimally observe different gradients in the same field of interest. The intensity in the processed image is proportional to the refractive index gradient. A gradient in the same direction as the knife edge appears dark. Gradients in the opposite direction appear bright. This method, employed with arc light sources, is still one of the simplest ways to view refractive index changes in transparent solids, liquids, and gases. A well-designed schlieren system can easily detect the presence of a refractive index gradient that causes 1 arc-second deviation of a light ray.

Color schlieren. Except for locating and identifying schlieren-causing events such as turbulent eddies, shock waves, and density gradients, schlieren systems are usually considered to be qualitative instruments. Quantitative techniques for determining density are possible but are much more difficult to employ. The most common of these is color schlieren. The knife edge is replaced with a multicolored filter. Rays of light refracted through different angles appear in different colors in the final image.

Use of coherent light. The availability of lasers and new optical components has expanded the method considerably. When a coherent light source such as a laser is used, the knife edge can be replaced by a variety of phase-, amplitude-, or polarization-modulating filters to produce useful transformations in the image intensity. For example, replacing the knife edge with a 180° phase step produces the Hilbert transform. Adding a graded-amplitude wedge to this performs differentiation of the image. Phase-shifting only the central focused spot produces phase-contrast interferometry. A variety of filters produce images referred to as schlieren interferometry. These include the insertion of stops, gratings, and Wollaston prisms in place of the knife edge. *See* INTERFEROMETRY; LASER; POLARIZED LIGHT.

Applications. Schlieren systems are common in aerodynamic test facilities and optical test laboratories. A schlieren system was used in the space shuttle *Space Lab 3* flights to observe fluid densities around crystals growing in zero gravity. *See* SHOCK WAVE; TRANSONIC FLIGHT; WIND TUNNEL.

James D. Trolinger

Bibliography. G. S. Settles, *Schlieren and Shadowgraph Techniques*, Springer-Verlag, 2001.

Schmidt camera

A fast, compact, wide-field optical system that uses a thin aspheric front lens at the center of curvature of a larger concave spherical mirror. The image is focused onto a curved focal plane between the two elements (**Fig. 1**). Schmidt cameras are sometimes

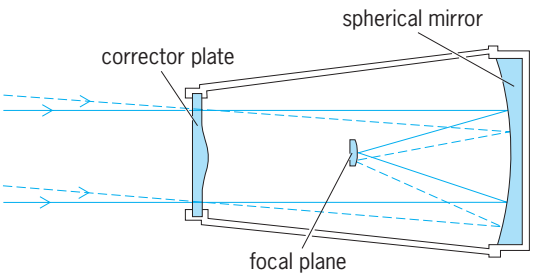


Fig. 1. Cross section of Schmidt camera with aspherical corrector plate. (*After J. M. Pasachoff, Astronomy: From the Earth to the Universe, 6th ed., Brooks/Cole Publishing, 2002*)

Largest Schmidt telescopes

Telescope or institution	Location	Diameter: Corrector plate, in. (m)	Diameter: Spherical mirror, in. (m)	Focal ratio	Date completed
Karl Schwarzschild Observatory	Tautenberg (near Jena), Germany	53 (1.3) (removable)	79 (2.0)	*f*/2	1960
Samuel Oschin Telescope, Palomar Observatory	Palomar Observatory, California	48 (1.2)	72 (1.8)	*f*/2.5	1948, 1987, 2000
United Kingdom Schmidt, Anglo-Australian Observatory	Siding Spring Observatory, New South Wales, Australia	48 (1.2)	72 (1.8)	*f*/2.5	1973
Tokyo Astronomical Observatory	Kiso Mountains, Japan	41 (1.1)	60 (1.5)	*f*/3.1	1975
European Southern Observatory Schmidt*	La Silla, Chile	39 (1.0)	64 (1.6)	*f*/3	1972

*Decommissioned in 1998.

used in microscopes, astronomical spectrographs, and projection televisions. The largest such devices can image several degrees of the sky and are known as Schmidt telescopes. The Estonian optician Bernhard Schmidt devised the scheme in 1930. The field of best focus is located midway between the lens and the mirror and is curved toward the mirror, with a radius of curvature equal to the focal length. Historically film or photographic plates were bent to match this curved focus, and used to make the first all-sky survey pictures. Field-flattening lenses permit the use of charge-coupled devices (CCDs) in newer applications. Large mosaics of CCDs have replaced photographic emulsions in some applications, and large numbers of optical fibers have also been used in multiobject spectrograph configurations. *See* GEOMETRICAL OPTICS; LENS (OPTICS); MIRROR OPTICS.

Schmidt telescopes are very fast, some having focal ratios in the vicinity of *f*/1. Their distinguishing feature is the very wide angle over which good images are obtained. Thus, under a dark sky they are sensitive to extended objects of low surface brightness. Schmidt telescopes have no coma; because the only lens element is so thin, they suffer only slightly from chromatic aberration and astigmatism. The front element, smaller than the main mirror, is sometimes known as a corrector plate. Often the outer surface of the corrector plate is plane, with the inner surface bearing the figure; however, a slight curvature is sometimes introduced to prevent misleading ghost images. This corrector plate almost eliminates the spherical aberration, giving extremely sharp images. For the most critical work, over a wide range of wavelengths, it must be made achromatic. *See* ABERRATION (OPTICS); OPTICAL SURFACES.

The Schmidt design became widely used after 1936, when the secret of the fabrication of the corrector plate was released. Schmidt is credited as much with his skill in figuring the fourth-degree curves on the corrector plate as with the design itself. The largest Schmidt telescopes (see **table**; **Fig. 2**) are 39–53 in. (1.0–1.3 m) in diameter. Larger corrector plates would sag too much, as they can only be supported around their circumference.

Increasingly, Schmidt telescopes, especially several 24 in. (0.6 m) in diameter, are equipped with

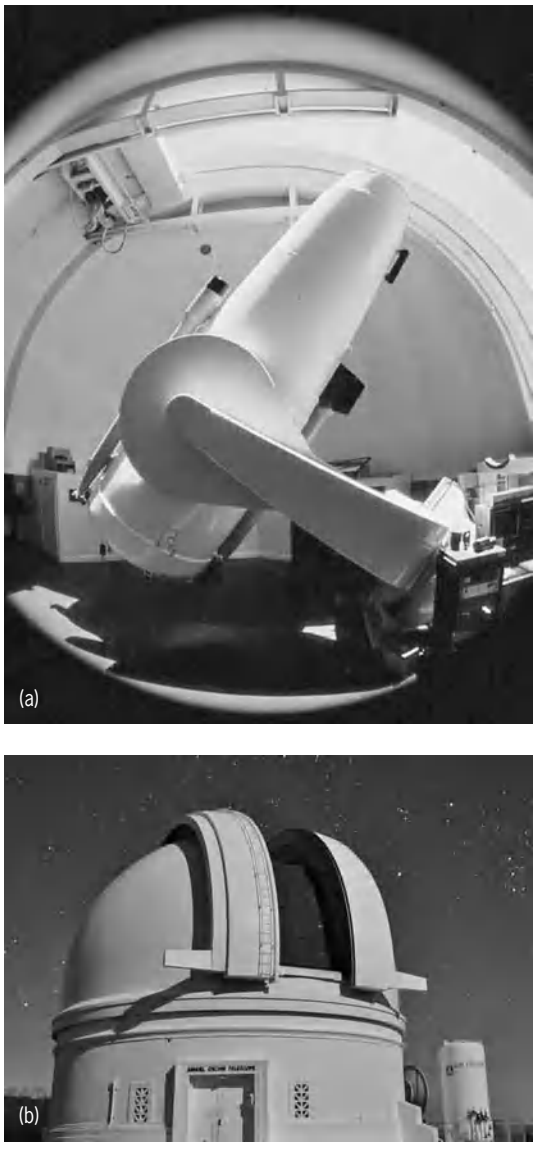


Fig. 2. Samuel Oschin Telescope, a 48-in. (1.2-m) Schmidt camera at Palomar Observatory. (*a*) Interior view (*Alain Maury*). (*b*) Exterior view. A large liquid nitrogen tank to the right of the open dome is used for cooling the instrument. (*Scott Kardel*)

charge-coupled devices. An upgrade to the Samuel Oschin Telescope (48-in. or 1.2-m Schmidt) at Palomar Observatory in California installed the Yale-QUEST mosaic of 112 CCDs, which images an area of the sky 4° square. While smaller than the original photographic field (6°), it is still enormous when compared with the fields for charge-coupled devices at most current large telescopes. The telescope is now fully robotic, and operates automatically and unattended every clear night. The Palomar-QUEST camera is used about half the time to survey the sky for time-variable objects such as Supernovae and Gamma-Ray Bursters, and the other half to search for moving objects such as near-Earth asteroids and distant solar system objects such as Sedna. The use of charge-coupled devices provides higher sensitivity and allows advanced image processing and easy distribution of data and results. These large-format cameras require cooling to reduce electronic and readout noise (Fig. 2*b*). *See* CHARGE DEVICES.

Sky surveys. The Schmidt telescope at Palomar Observatory, now the Samuel Oschin Telescope, was used in the 1950s to survey the northern two-thirds of the sky in the Palomar Observatory–National Geographic Society Sky Survey (POSS). The plate pairs—one photographed in red light and one in blue light—provided a first-epoch coverage on which many nebulae, galaxies, clusters of galaxies, and other objects were discovered. Each plate, about 14 in. (35 cm) square, covers an area about 6° square. The POSS and later Schmidt sky surveys are basic references. An infrared survey north of the equator was carried out in the 1970s. A smaller Schmidt telescope (18 in. or 46 cm) at the Palomar Observatory has discovered many asteroids with orbits that cross that of the Earth, as well as comet Shoemaker-Levy 9 and other comets. *See* ASTEROID; COMET.

The European Southern Observatory (ESO) Schmidt in Chile (decommissioned in 1998) and the United Kingdom Schmidt at the Anglo-Australian Observatory in Australia were used to survey the southern hemisphere, with improved emulsions, from declination +20° to the south celestial pole. The United Kingdom Schmidt was provided with an achromatic corrector plate in 1977. It completed an infrared survey of the Milky Way and Magellanic Clouds, and extended this survey to the whole southern sky, and to a blue-red study of the region just south of the celestial equator to provide images showing stars about 1.5 magnitudes fainter than those on the Palomar survey. In addition to mapping, the plates are used to help make optical identifications of southern radio and x-ray sources. A second-epoch red survey was carried out in the 1990s, largely for Hubble guide-star purposes. The United Kingdom Schmidt Telescope is carrying out a survey of the Milky Way Galaxy and Magellanic Clouds in the red light of hydrogen (H-alpha) to reveal supernova remnants, planetary nebulae, and the fine structure of gas clouds in the Galaxy. Study of the plates has been speeded by the development of automatic plate-measuring machines. Three-color U.K. Schmidt photographs of celestial objects by D. F. Malin are widely viewed.

In 1987, an achromatic doublet corrector plate was installed in the Samuel Oschin Telescope. With it, a second-epoch sky survey was undertaken. An infrared color (800–900 nanometers, limiting magnitude 19) was added to the blue (385–550 nm) and red (610–690 nm) colors, which are slightly different from those of the first survey but are similar to those used by the United Kingdom and European Southern Observatory Schmidts for their southern survey. Advances in film technology and thermal control of the telescope, to improve the focus, allowed a limiting magnitude fainter by a factor of 6 in the blue (to magnitude 22.5) and 2 in the red (to magnitude 20.8) to be reached and images to be recorded with finer resolution. The infrared survey was made possible by developments in hypersensitization of infrared emulsions, which increased their speed by a factor of 200. The survey includes 894 fields between the celestial equator and the north celestial pole. The second-epoch plates were scanned to produce the Digital Sky Survey (DPOSS). As parts of the sky were completely surveyed, the telescope was used to search for Kuiper Belt objects, and this work has continued with the large digital camera. *See* KUIPER BELT.

Second-epoch Schmidt survey plates were scanned to make the Hubble Space Telescope Guide Star Catalog of 16,000,000 objects, and also form the basis of the U.S. Naval Observatory's USNO-B astrometric catalog of over 10^9 stars. The catalogs, scanned images, and database are available on the World Wide Web. *See* ASTRONOMICAL ATLASES; ASTRONOMICAL CATALOGS; ASTRONOMICAL PHOTOGRAPHY; HUBBLE SPACE TELESCOPE.

Spectroscopy. In addition to direct photography, Schmidt telescopes can be used to obtain the spectra of celestial objects. The United Kingdom Schmidt can be used with two 48-in.-diameter (1.2-m) objective prisms, either singly or in combination, to spread each image out into a short spectrum. The telescope also has a facility called FLAIR, which uses optical fibers to couple the focal surface to a charge-coupled-device spectrograph, allowing detailed spectra to be obtained of up to 100 objects simultaneously in the wide field of view. FLAIR is widely used for galaxy redshift surveys and surveys of particular star types. A fully automated version of FLAIR, known as 6dF (for 6-degree field) is being built to allow a galaxy-redshift survey of the entire southern hemisphere to be undertaken. The United Kingdom Schmidt and the Curtis Schmidt Telescope, a Schmidt telescope with a 24-in.-diameter (0.61-m) corrector plate and a 36-in.-diameter (0.91-m) spherical mirror at the Cerro Tololo Inter-American Observatory in Chile, also have objective-prism capabilities. *See* ASTRONOMICAL SPECTROSCOPY.

Related systems. The success of the Schmidt design has led to many other types of catadioptric systems, with a combination of lenses and mirrors. The desire for a wide field has led to most modern large telescopes being built to the Ritchey-Chrétien

design, which use two hyperboloid mirrors instead of the traditional paraboloids to achieve coma-free fields of up to one degree, with a small amount of focal plane curvature. These fields are much smaller than those of Schmidts, but the telescope aperture can be larger, because no large corrector lens is needed.

Schmidt-Cassegrain design and amateur use. Telescopes employing some features of the Schmidt design are also available for amateurs. Many amateur astronomers use telescopes of the Schmidt-Cassegrain design, in which a small mirror attached to the rear of the corrector plate reflects and refocuses the image through a hole in the center of the primary mirror. These telescopes, in sizes ranging from 5 in. (12.5 cm) to 16 in. (41 cm), are portable and relatively inexpensive for the quality of image. Their fields are much narrower than a standard Schmidt design, however, and the focal ratios much slower, usually about *f*/10. A relatively inexpensive range of telescopes are manufactured based on a modified Ritchey-Chrétien design. Because aspheric hyperboloids are expensive to manufacture, these telescopes combine a spherical primary with full-aperture corrector lens to emulate the necessary primary figure. Together with a hyperboloidal secondary attached to the rear of the corrector, they produce a wide angle (curved) coma-free field. *See* TELESCOPE.

Jay M. Pasachoff; Andrew Pickles

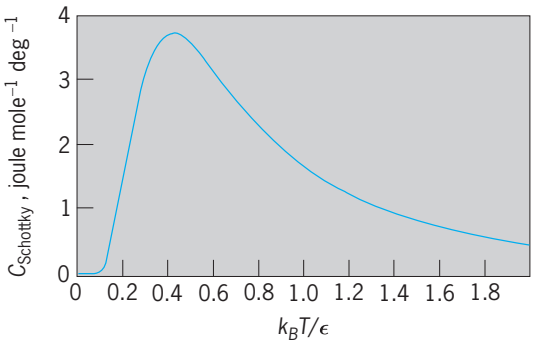
Bibliography. M. Capaccioli (ed.), *Astronomy with Schmidt-Type Telescopes*, IAU Colloq. 78, D. Reidel, 1983; J. Chapman et al. (eds.), *The Future Utilisation of Schmidt Telescopes*, IAU Colloq. 148, Astronomical Society of the Pacific, 1995; D. di Cicco, A preview of Meade's RCX400, *Sky Telesc.*, 109(5):86–89, May 2005; H. T. MacGillivray and E. B. Thomson (eds.), *Digitized Optical Sky Surveys*, Kluwer Academic Publishers, 1992; D. F. Malin, *A View of the Universe*, Sky Publishing Co./Cambridge University Press, 1993; S. Marx and W. Pfau, *Astrophotography with the Schmidt Telescope*, Cambridge University Press, 1992.

Schottky anomaly

A contribution to the heat capacity of a solid arising from the thermal population of discrete energy levels as the temperature is raised. The effect is particularly prominent at low temperatures, where other contributions to the heat capacity are generally small. *See* SPECIFIC HEAT OF SOLIDS.

Discrete energy levels may arise from a variety of causes, including the removal of orbital or spin degeneracy by magnetic fields, crystalline electric fields, and spin orbit coupling, or from the magnetic hyperfind interaction. Such effects commonly occur in paramagnetic ions. *See* LOW-TEMPERATURE THERMOMETRY; PARAMAGNETISM.

The thermal population of an energy level E_i is proportional to the Boltzmann factor, $\exp(-E_i/k_BT)$,



Theoretical curve for the Schottky heat capacity when there are two energy levels separated by a splitting ϵ. (*After H. M. Rosenberg, Low Temperature Solid State Physics, Clarendon Press, 1963*)

where T is the temperature and k_B is Boltzmann's constant ($k_B = 1.38 \times 10^{-23}\,\text{J} \cdot \text{K}^{-1}$). The largest contributions to the Schottky heat capacity, C_{Schottky}, will therefore occur at temperatures where k_BT is comparable with the E_i. Quantitative estimates of C_{Schottky} may be made in terms of the Boltzmann factors. For an assembly of N, identical noninteracting systems, each with n low-lying energy levels E_i ($i = 1, \ldots, n$), the total internal energy is given by the equation below. The corresponding heat capacity is then

$$U = \frac{N \sum_{i=1}^{n} E_i \exp(-E_i/k_BT)}{\sum_{i=1}^{n} \exp(-E_i/k_BT)}$$

given by $C_{\text{Schottky}} = dU/dT$. The simplest example is the case $n = 2$, corresponding to a single splitting ϵ between the ground state ($i = 1$) and the first excited state ($i = 2$). The **illustration** shows the heat capacity for this case plotted as a function of k_BT/ϵ. For more complex energy-level schemes, the curves are generally similar, with a single broad peak near $k_BT/\bar{\epsilon} \sim 1$, where $\bar{\epsilon}$ is the mean energy above the ground state. However, more structured curves, with more than one peak, are also possible if the energy-level scheme includes excitation energies of very different relative magnitudes. *See* BOLTZMANN STATISTICS.

Corresponding to the Schottky heat capacity, there is a contribution to the entropy, $S = \int (C_{\text{Schottky}})/dT$. This can act as a barrier to the attainment of low temperatures if the substance is to be cooled either by adiabatic demagnetization or by contact with another cooled substance. Conversely, a substance with a Schottky anomaly can be used as a heat sink in experiments at low temperatures (generally below 1 K) to reduce temperature changes resulting from the influx or generation of heat. *See* ADIABATIC DEMAGNETIZATION; LOW-TEMPERATURE PHYSICS.

W. P. Wolf

Bibliography. E. S. Gopal, *Specific Heats at Low Temperatures*, 1966; C. Kittel, *Introduction to Solid State Physics*, 7th ed., 1995; H. M. Rosenberg, *Low Temperature Solid State Physics*, 1963.

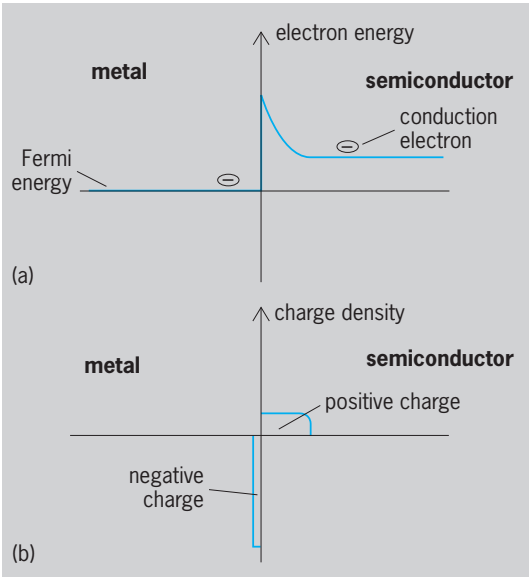
Schottky barrier diode

A metal-semiconductor diode (two terminal electrical device) that exhibits a very nonlinear relation between voltage across it and current through it; formally known as a metallic disk rectifier. Original metallic disk rectifiers used selenium of copper oxide as the semiconductor coated on a metal disk. Today, the semiconductor is usually single-crystal silicon with two separate thin metal layers deposited on it to form electrical contacts. One of the two layers is made of a metal which forms a Schottky barrier to the silicon. The other forms a very low resistance, so-called ohmic, contact. The Schottky barrier is an electron or hole barrier caused by an electric dipole charge distribution associated with the contact potential difference which forms between a metal and a semiconductor under equilibrium conditions. The barrier is very abrupt at the surface of the metal because the charge is primarily on the surface. However, in the semiconductor, the charge is distributed over a small distance, and the potential gradually varies across this distance (see **illus.**). *See* CONTACT POTENTIAL DIFFERENCE.

A basic useful feature of the Schottky diode is the fact that it can rectify an alternating current. Substantial current can pass through the diode in one direction but not in the other. If the semiconductor is *n*-type, electrons can easily pass from the semiconductor to the metal for one polarity of applied voltage, but are blocked from moving into the semiconductor from the metal by a potential barrier when the applied voltage is reversed. If the semiconductor is *p*-type, holes experience the same type of potential barrier but, since holes are positively charged, the polarities are reversed from the case of the *n*-type semiconductor. In both cases the applied voltage of one polarity (called forward bias) can reduce the potential barrier for charge carriers leaving the semiconductor, but for the other polarity (called reverse bias) it has no such effect. *See* DIODE; SEMICONDUCTOR; SEMICONDUCTOR RECTIFIER.

James E. Nordman

Bibliography. R. H. Bube, *Electrons in Solids: An Introductory Survey*, 3d ed., 1992; G. Neudeck, *The PN Junction Diode*, 2d ed., 1989.



Distributions of (*a*) electric potential and (*b*) electric charge for a Schottky diode without voltage bias.

Schottky effect

The enhancement of the thermionic emission of a conductor resulting from an electric field at the conductor surface. Since the thermionic emission current is given by the Richardson formula, an increase in the current at a given temperature implies a reduction in the work function ϕ of the emitter. *See* THERMIONIC EMISSION.

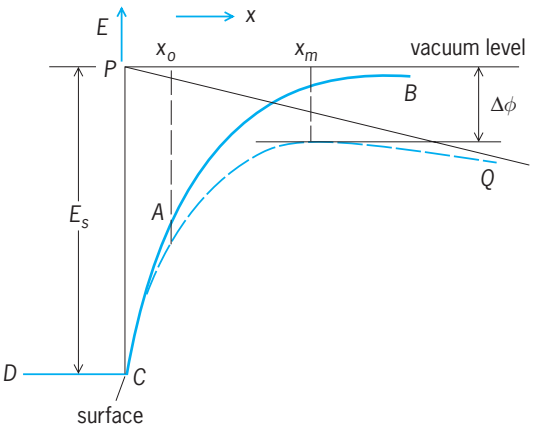


Fig. 1. Surface potential barrier and Schottky effect; the lowering of the work function $\Delta\phi$ is highly exaggerated.

With reference to **Fig. 1**, let the vacuum level represent the energy of an electron at rest in free space and let *CD* be the energy of a conduction electron at rest in a metal. If an electron approaches the metal surface from infinity, its potential energy *V* relative to the vacuum level is given by the well–known image potential $V(x) = -e^2/4x$, where x is the distance from the surface, and $e = 1.6 \times 10^{-19}$ coulomb is the magnitude of the electron's charge. The image potential is valid only for $x > x_0$, where x_0 is of the order of the distance between neighboring atoms in the metal; that is, x_0 is a few angstroms. In the absence of an applied field, *CAB* then represents the potential energy of an electron as a function of x. *AB* corresponds to the image potential; the exact shape of the curve between *C* and *A* is uncertain.

Suppose now a constant field *F* is applied externally between the surface of the emitting cathode and an anode; this produces a potential energy of an electron of $-eFx$ (line *PQ* in Fig. 1), and hence the total potential energy of an electron for $x > x_0$ is

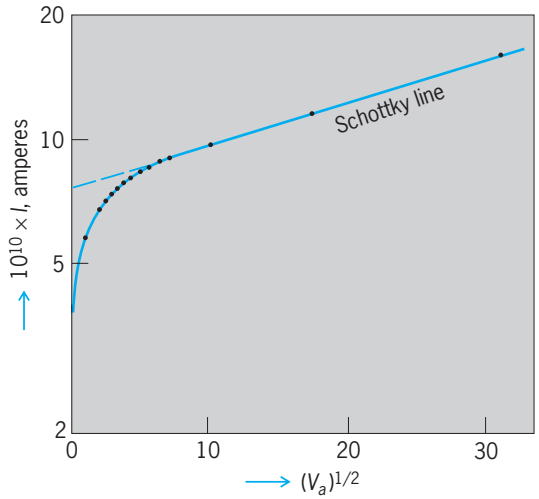


Fig. 2. Logarithm of thermionic emission current *I* of tungsten as function of square root of anode voltage V_a. (*After W. B. Nottingham, Phys. Rev., 58:927–928, 1940*)

given by Eq. (1), indicated by the broken line *CAQ*

$$V(x) = \frac{e^2}{4x} - eFx \tag{1}$$

in Fig. 1. This function has a maximum value given by Eq. (2).

$$x = x_m = {}^1/_2\left(\frac{e}{F}\right)^{1/2} \tag{2}$$

The maximum lies below the vacuum level by an amount $\Delta\phi$, given in Eq. (3), which represents the

$$\Delta\phi = V(x_m) = -e(eF)^{1/2} \tag{3}$$

reduction in the work function of the metal. For $F = 1000$ volts/cm, $x_m \equiv 10^{-5}$ cm and $\Delta\phi \equiv 10^{-2}$ eV; the actual change in the work function is thus small. If a field is present, the work function ϕ in the Richardson formula should be replaced by $(\phi - \Delta\phi)$. Hence, the current increases by the factor given by notation (4).

$$\exp\left[\frac{e}{kT}(eF)^{1/2}\right] \tag{4}$$

According to this interpretation, a plot of the logarithm of the current versus the square root of the anode voltage should yield a straight line. An example is given in **Fig. 2** for tungsten; the deviation from the straight line for low anode voltages is due to space-charge effects. *See* SPACE CHARGE.

The straight portion of the line (the Schottky line) confirms the interpretation; the true saturation current for zero field is obtained by extrapolation of the Schottky line as indicated. Detailed studies have shown extremely small periodic deviations with reference to the Schottky line; these deviations are interpreted on the basis of the wave-mechanical theory describing the motion of electrons across the image potential barrier shown in Fig. 1. Adrianus J. Dekker

Schrödinger's wave equation

A linear, homogeneous partial differential equation that determines the evolution with time of a quantum-mechanical wave function.

Quantum mechanics was developed in the 1920s along two different lines, by W. Heisenberg and by E. Schrödinger. Schrödinger's approach can be traced to the notion of wave-particle duality that flowed from A. Einstein's association of particlelike energy bundles (photons, as they were later called) with electromagnetic radiation, which, classically, is a wavelike phenomenon. For radiation of definite frequency f, each bundle carries energy hf. The proportionality factor, $h = 6.626 \times 10^{-34}$ joule-second, is a fundamental constant of nature, introduced by M. Planck in his empirical fit to the spectrum of blackbody radiation. It is the quintessential parameter of quantum mechanics. This notion of wave-particle duality was extended in 1923 by L. de Broglie, who postulated the existence of wavelike phenomena associated with material particles such as electrons. *See* PHOTON; WAVE MECHANICS.

Energy equation. This was the direction pursued by Schrödinger. There are certain purely mathematical similarities between classical particle dynamics and the so-called geometric optics approximation to propagation of electromagnetic signals in material media. For the case of a single (nonrelativistic) particle moving in a potential $V(\mathbf{r})$, this analogy leads to the association with the system of a wave function, $\Psi(\mathbf{r})$, which obeys Eq. (1). Here m is the mass of

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = E\Psi \tag{1}$$

the particle, E its energy, $\hbar = h/(2\pi)$, and ∇^2 is the laplacian operator. *See* CALCULUS OF VECTORS; GEOMETRICAL OPTICS; LAPLACIAN.

This equation may be applied to the hydrogen atom, taking for V the Coulomb potential $V = -e^2/r$, where e is the electron charge. From a purely mathematical point of view, the equation of course has solutions for any value of the energy. However, even without a view as to what the wave function signifies, it may be postulated on intuitive grounds that, in order to be physically acceptable, solutions ought to be well behaved, that is, single-valued, nonsingular, and bounded as the distance r goes to infinity. With these conditions imposed, it is found, for energies less than 0, that only certain discrete energies are allowed. The energy spectrum that emerges in this way is exactly that of the older Bohr model of the atom. It is in good, though not quite perfect, agreement with experiment. *See* ATOMIC STRUCTURE AND SPECTRA; ENERGY LEVEL (QUANTUM MECHANICS).

The so-called energy equation (1) stands in analogy with the equation governing an optical (that is, electromagnetic) disturbance of definite frequency f propagating in a medium with a space-dependent index of refraction, $n(\mathbf{r})$. With $\Phi(\mathbf{r},t) = u(\mathbf{r}) \sin 2\pi ft$ representing the amplitude of the disturbance, the

equation for $u(\mathbf{r})$ is Eq. (2).

$$-\nabla^2 u = \frac{n^2}{c^2}\omega^2 u \qquad \omega = 2\pi f \tag{2}$$

See ELECTROMAGNETIC RADIATION.

Time-dependent equations. In the electromagnetic case, the disturbance of course need not be of definite frequency. For the general situation corresponding to an arbitrary superposition over frequencies, the amplitude obeys the wavelike equation (3). For

$$-\nabla^2 \Phi = \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\Phi \tag{3}$$

the particle case, a superposition over frequencies would correspond to a superposition of motions of different energies. That is without meaning classically. Classically, a particle in its motions has some definite energy. Nevertheless, it is possible to set that problem aside and ask what more general equation a time- as well as space-dependent wave function, $\Psi(\mathbf{r},t)$, might obey. Here the analogy with electromagnetism beckons. Instead of a wave equation involving second-order time derivatives, what suggests itself is Eq. (4), which is now called the Schrödinger

$$i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi \tag{4}$$

equation. It involves only a first-order time derivative.

The energy equation (1) corresponds to a monochromatic solution of the form $\Phi = u(\mathbf{r})e^{i\omega t}$, with $E = \hbar\omega$ the particle energy. A general solution of Eq. (4) corresponds to a superposition over energies. But such a superposition raises the question of what it means to superpose energies, and, for that matter, what is the physical significance of Ψ.

Multiparticle systems. The wave function can be generalized to a system of more than one particle, say N of them. A separate wave function is not assigned to each particle. Instead, there is a single wave function, $\Psi(\mathbf{r}_1,\mathbf{r}_2, \ldots, \mathbf{r}_N,t)$, which depends at once on all the position coordinates as well as time. This space of position variables is the so-called configuration space. The generalized Schrödinger equation is Eq. (5), where the potential V may now depend on

$$\mathrm{i}\hbar\frac{\partial \Psi}{\partial t} = -\sum_{i=1}^{N}\frac{\hbar^2}{2m_i}\nabla_i^2\Psi + V\Psi \tag{5}$$

all the position variables. Three striking features of this equation are to be noted:

1. The complex number i (the square root of minus one) appears in the equation. Thus Ψ is in general complex. *See* COMPLEX NUMBERS AND COMPLEX VARIABLES.

2. The time derivative is of first order. Thus, if the wave function is known as a function of the position variables at any one instant, it is fully determined for all later times.

3. The Schrödinger equation is linear and homogeneous in Ψ, which means that if Ψ is a solution so is $c\Psi$, where c is an arbitrary complex constant. More generally, if Ψ_1 and Ψ_2 are solutions, so too is the linear combination $c_1\Psi_1 + c_2\Psi_2$, where c_1 and c_2 are arbitrary complex constants. This is the superposition principle of quantum mechanics. *See* SUPERPOSITION PRINCIPLE.

Probability interpretation. The Schrödinger equation suggests an interpretation in terms of probabilities. Provided that the wave function is square integrable over configuration space, it follows from Eq. (5) that the norm, $\langle\Psi|\Psi\rangle$, is independent of time, where the norm is defined by Eq. (6). It is possible

$$\langle\Psi \mid \Psi\rangle = \int d^3x_1 d^3x_2 \ldots d^3x_N \Psi^*\Psi \tag{6}$$

to normalize Ψ (multiply it by a suitable constant) to arrange that this norm is equal to unity. With that done, the Schrödinger equation itself suggests that expression (7) is the joint probability distribution at

$$\Psi^*\Psi d^3x_1 d^3x_2 \ldots d^3x_N \tag{7}$$

time t for finding particle 1 in the volume element d^3x_1, particle 2 in d^3x_2, and so forth.

Quantum mechanics, that is, deals with probabilities. All that can be known about the state of a system at any instant is contained in its wave function; and the wave function, though it evolves deterministically with time, yields in general only probabilistic predictions about the outcomes of measurements that are to be made on the system. There are two broad classes of problems that arise within quantum mechanics: (1) to determine, in general, the possible outcomes of measurements of various physical observables (energy, angular momentum, linear momentum, particle location, and so forth); and (2) to determine, for any given state Ψ, the probability distribution over those outcomes. For position measurements all locations are possible, with the probability distribution as given above. For other kinds of observables the rules are more complicated, but they are definite. For instance, in the special case of energy the possible outcomes (as illustrated on the example of a one-particle system) are the energies, E, for which Eq. (1), Schrödinger's starting point, has well-behaved solutions. *See* PROBABILITY (PHYSICS).

Generalizations. The Schrödinger equation and the rules of quantum-mechanical interpretation can be easily extended to incorporate spin, a sort of intrinsic angular momentum property possessed by certain species of particles (for example, electrons, protons, and neutrons). Extensions that incorporate the dictates of special relativity pose much deeper issues. They lead to the Dirac equation for the relativistic electron, a generalization of the Schrödinger equation that embodies its fundamental characteristics (the superposition principle, probabilistic interpretations, and so forth). *See* NONRELATIVISTIC QUANTUM THEORY; QUANTUM MECHANICS; RELATIVISTIC QUANTUM THEORY; SPIN (QUANTUM MECHANICS); WAVE EQUATION; WAVE MOTION.

Sam Treiman

Bibliography. N. Bohr, *Atomic Physics and Human Knowledge*, 1958, reprint 1987; W. Heisenberg, *Physics and Philosophy*, 1959; M. Jammer, *The Conceptual Development of Quantum Mechanics*, 1966, reprint 1989; E. Schrödinger, *Collected Papers on Wave Mechanics*, 1928; J. Wheeler and W. Zurek (eds.), *Quantum Theory and Measurement*, 1983.

Schuler pendulum

Any apparatus which swings, because of gravity, with a natural period of 84.4 min, that is, with the same period as a hypothetical simple pendulum whose length is the Earth's radius. In 1923 Max Schuler showed that such an apparatus has the unique property that the pendulum arm will remain vertical despite any motions of its pivot. It is therefore useful as a base for navigational instruments. Schuler also showed how gyroscopes can be used to increase the period of a physical pendulum to the desired 84.4 min.

Gyrocompasses employ the Schuler principle to avoid errors due to ship accelerations. The principle has become the foundation of the science of inertial navigation.

Basic principle. The principle is most easily explained with the simple pendulum of **Fig. 1***a*. The period T for one complete swing of this device is given by Eq. (1), where l is the length of the pendulum and g is the acceleration, due to gravity, of a freely falling object (32.2 ft/s^2 or 9.8 m/s^2). If length l could be made equal to the Earth's radius R, as in Fig. 1*b*, gravity would hold the bob fixed, and the pendulum arm would stay vertical for all motions of point P. If l equals R, the period becomes 84. 4 min.

$$T = 2\pi\sqrt{\frac{l}{g}} \qquad (1)$$

While a pendulum of such length cannot be built, the same effect can be achieved using the solid pendulous rod of Fig. 1*c*. Although only a few feet long, this pendulum can be given a period of 84.4 min by placing the pivot close to the center of mass. (If the pivot were at the mass center, the period would be infinitely long; that is, the rod would remain indefinitely in any position.)

Figure 2*b* demonstrates that a Schuler-period pendulum will also stay locally vertical during accelerations of its pivot, just as the simple pendulum did in Fig. 1*b*. The proof is by comparison. In Fig. 2*a* the pivot is high, and the pendulum swings back from vertical when the pivot is accelerated. In Fig. 2*c* the pivot is at the mass center, and the pendulum does not swing at all, so that after traveling some distance around the Earth it is forward of the local vertical. In Fig. 2*b* the pivot is placed at just the right point, so that whatever the acceleration of the pivot, the pendulum swings back just enough to stay locally vertical.

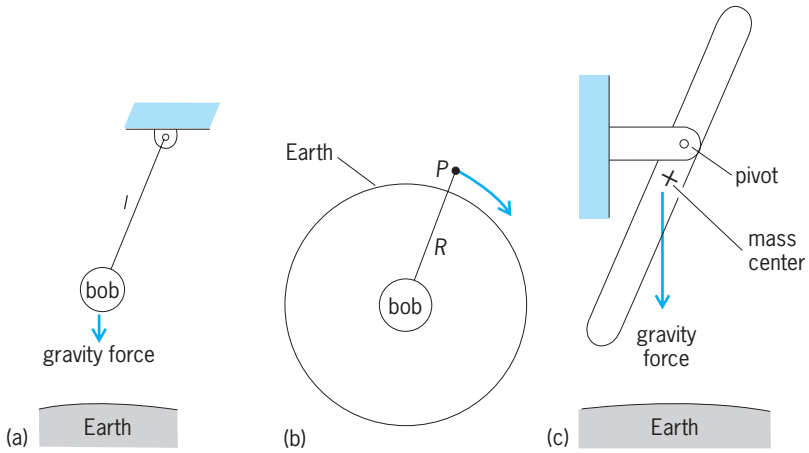


Fig. 1. Basic principles illustrated by (*a*) simple pendulum; (*b*) Earth-radius pendulum; (*c*) solid pendulous rod.

Unfortunately, for a 2-ft (0.6-m) pendulum like that of Fig. 1*c* to have an 84.4-min period, the pivot must be placed only about 0.2×10^{-6} in. (5 nanometers) above the mass center, an unrealizable physical tolerance.

Gyropendulum and gyrocompass. The period of a pendulous mass can be greatly increased by attaching to it a rapidly spinning gyro wheel as, for example, in **Fig. 3**. Extra freedom, provided by the thrust bearing in Fig. 3, is necessary to permit the gyroscopic action.

The gyrocompass is a special pendulous gyroscope that is sensitive to Earth rotation. If the pendulous mass is m, the horizontal spin axis has a natural period of oscillation, about north, given by Eq. (2),

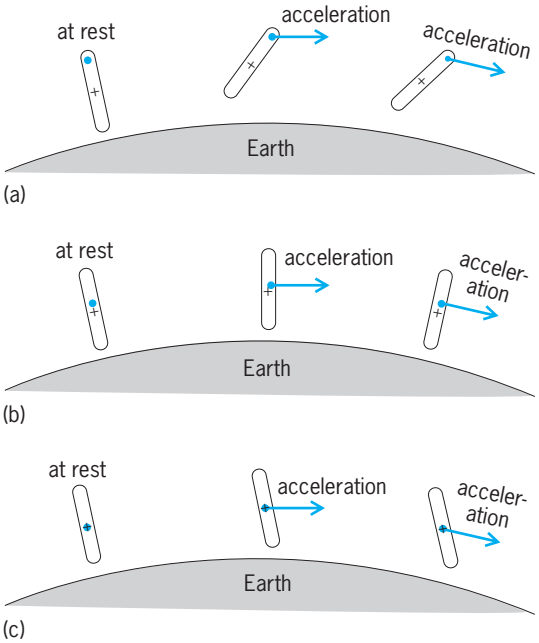


Fig. 2. Behavior of pendulous rod when accelerated about Earth. (*a*) Short-period pendulous rod. (*b*) Schuler-period pendulous rod. (*c*) Nonpendulous rod.

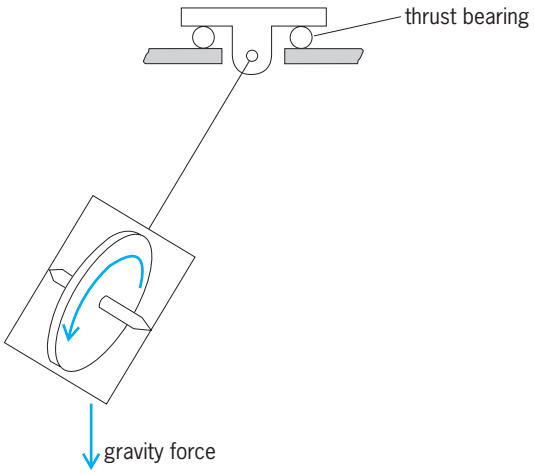


Fig. 3. Gyropendulum; thrust bearing provides extra freedom, necessary for gyroscopic action.

where H is gyro momentum, λ is latitude, and Ω_e is

$$T = 2\pi\sqrt{\frac{H}{mgl\Omega_e \cos\lambda}} \qquad (2)$$

Earth rate (360°/day). Ship-borne gyrocompasses are rendered insensitive to ship accelerations by making $T = 84.4$ min. *See* GYROCOMPASS.

Inertial guidance. The stable platform of an inertial guidance system is made to behave like a pendulum by the use of feedback from an accelerometer mounted on the platform. The accelerometer signal controls a restoring torque to the platform.

The principle is demonstrated in **Fig. 4**. The accelerometer consists of a spring-restrained mass. If the platform tilts, gravity pulls the mass to one side; the amount is measured electrically, integrated, and used to drive the platform motor. (Actual inertial platforms are controlled by a gyro, which is precessed at a rate by the integrator signal.) *See* ACCELEROMETER.

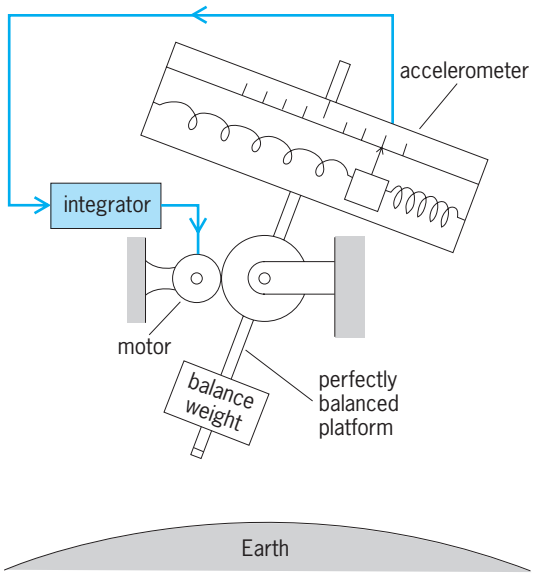


Fig. 4. Principle of Schuler-tuned stable platform.

The system of Fig. 4 behaves like the physical pendulum of Fig. 1*a*. When either device is tilted from vertical, gravity produces a restoring action so that the device swings back and forth about the vertical. The period of swing is $2\pi\sqrt{l/g}$ for the pendulum of Fig. 1*a*, and $2\pi\sqrt{1/Kg}$ for the platform of Fig. 4, where K is the loop gain from tilt angle to motor rate. The platform is accordingly given an 84.4-min period by making K equal to $1/R$.

If the platform of Fig. 4 is installed in a vehicle and is once leveled, it will always remain locally level despite motions of the vehicle about the Earth. Then the signal from the accelerometer will represent only vehicle acceleration, uncontaminated by gravity, and can be used to determine vehicle position. *See* INERTIAL GUIDANCE SYSTEM.

Robert H. Cannon

Sciatica

A specific pattern of leg pain that follows the course of the sciatic nerves. It is usually caused by injury to one or more of the sciatic root nerves within the lower spinal canal. The contributing nerves, the last two lumbar and first two sacral, exit sequentially through openings on each side of the spine. The individual nerves then enter common protective sheaths to become the sciatic nerves. Each one leaves the pelvis and runs down the back of the leg to the foot. Along the way branches supply the skin, particularly in the lower leg, and some of the muscles that move the hip and bend the knee. The sciatic nerves control all the muscles that enable walking on the heels and toes.

Sciatic pain is radicular pain, that is, pain that follows the distribution of a nerve. The typical pain locations are in the back of the thigh, behind the knee, over the calf, and into one or both sides of the foot as far as the toes. Due to the length of the sciatic nerve, the precise location of the pain varies, but it is always constant and dominant in the leg as opposed to the back. In the acute phase the intensity may fluctuate in response to movement or position, but the pain never disappears completely. Sciatica may be associated with diminished sensation or decreased power in the muscles supplied by the affected nerve roots. More often the pain occurs with signs of root irritation only but with otherwise normal function.

Sciatica results from the combination of mechanical pressure and chemical insult acting on a nerve root. Pressure alone produces a highly unpleasant tingling or burning sensation, but when accompanied by a local inflammatory response the result is excruciating pain. Because the physical deformation of a nerve root caused by a herniated disc is always associated with an inflammatory reaction, an acute disc rupture is by far the most frequent cause of sciatica. Other but rare possibilities include pyogenic infections with abscess formation, synovial cysts (cysts occurring in proximity to facet joints in the spine), diabetic neuropathy, vasculitis (inflammation of blood or lymph vessels), injected irritating agents,

and a range of metabolic and autoimmune disorders producing neuritis (nerve inflammation).

The term sciatica is frequently misapplied to a variety of other conditions that cause leg pain. It is most often confused with the referred leg pain that accompanies common backache. In these cases the back pain is dominant while the leg symptoms are typically intermittent and do not follow the path of the sciatic nerve. *See* PAIN; SPINE; VERTEBRA.

Hamilton Hall

Bibliography. D. Borenstein, *Back in Control*, M. Evans and Co., New York, 2001; J. Frymoyer (ed.), *The Adult Spine: Principles and Practice*, 2d ed., Lippincott–Raven, Philadelphia, 1997; H. Hall, *A Consultation with the Back Doctor*, McClelland & Stewart, Toronto, 2003.

Science

The study of nature and natural phenomena. Application of the term did not begin with any formal definition; rather the various disciplines arose independently, each in response to some particular need. It was then observed that certain of these disciplines had enough traits in common to justify classifying them together as one of the sciences. Usage is not, however, always unanimous as to whether some disciplines should be called sciences, and there is often lively controversy as to the propriety of speaking of the social or historical sciences.

Usually a science is characterized by the possibility of making precise statements which are susceptible of some sort of check or proof. This often implies that the situations with which the special science is concerned can be made to recur in order to submit themselves to check, although this is by no means always the case. There are observational sciences such as astronomy or geology in which repetition of a situation at will is intrinsically impossible, and the possible precision is limited to precision of description. There is also usually the implication that the subject matter of the individual science is something in the world of phenomena. Thus it is not usual to speak of the "science" of mathematics or the "science" of logic, even though both these disciplines are capable of the highest precision.

A common method of classifying sciences is to refer to them as either exact sciences or descriptive sciences. Examples of the former are physics and, to a lesser degree, chemistry; and of the latter, taxonomical botany or zoology. The exact sciences are in general characterized by the possibility of exact measurement. Measurement is fundamentally description by the use of numbers. Given the system of measurement and the measuring numbers for any special situation, that situation has been adequately described if it is possible to reconstruct a situation such that measurement on it gives the same numbers. Because mathematics operates to a large extent with numbers, systems that are subject to exact measurement are also susceptible of mathematical analysis; this susceptibility is one of the most important characteristics of the exact sciences. One of the most important tasks of a descriptive science is to develop a method of description or classification that will permit precision of reference to the subject matter. *See* PHYSICAL SCIENCE; SCIENTIFIC METHODS.

Percy W. Bridgman; Gerald Holton

Scientific and applications satellites

Scientific satellites are used to gain basic knowledge. They may be satellites of Earth or of other planets or their moons. Satellites used to apply space knowledge and techniques to practical purposes are called applications satellites. Many spacecraft are used for multiple purposes, combining space exploration, science, and applications in various ways. Scientific satellites are often called research satellites, and applications satellites are commonly designated by the application field, for example, navigation satellites. *See* SATELLITE (SPACECRAFT).

Many types of investigations are possible with scientific satellites. Some use the vantage point of space for Earth and astronomical observations. Others use the space environment for experiments not possible in the laboratory, such as determining the behavior of biological specimens or the growth of crystals under long periods of weightlessness. Astronomical telescopes, spectrometers, Geiger counters, ionization chambers, pressure gages, and magnetometers are among the instruments used in scientific satellites.

Space environment observations. The space environment was one of the first areas investigated in scientific satellites. In this environment are cosmic rays, dust, magnetic fields, and various radiations from the Sun and galaxies. A variable solar wind of charged particles blows continually throughout the solar system. More energetic solar particles are called solar cosmic rays. *See* COSMIC RAYS; SOLAR MAGNETIC FIELD; SOLAR WIND.

The famous Van Allen radiation belts were discovered by the first United States satellites. Soon thereafter satellites and space probes confirmed the existence of the solar wind. The continued investigation of these phenomena by satellites and space probes has led to the knowledge that a planet with a strong magnetic field, like Earth, is surrounded by a complex region, called a magnetosphere, which plays a major role in magnetic storms and auroras and in trapping solar particles to form radiation belts. *See* MAGNETOSPHERE; SPACE PROBE; VAN ALLEN RADIATION.

Meteoric particle studies seek to determine the size, velocity, and energy distributions of space dust particles, their composition and likely origin, and the physical processes that occur on impact of the particles with a surface. *See* INTERPLANETARY MATTER; METEOR.

Numerous investigations are made of the effect of the space environment on materials and processes, including effects of radiation damage, meteoric erosion processes, and extremely high vacuum for very

long periods. The influence of prolonged weightlessness on welding, alloying metals, growing crystals, and biological processes constitutes a promising field for future practical applications. *See* SPACE PROCESSING.

Earth science. Scientific satellites, sounding rockets, and space probes have had a profound effect on earth science. It is now possible to solve problems of the upper atmosphere and ionosphere that could not be resolved by observations from the ground. The Earth and its atmosphere can now be studied in comparison with the other planets and their atmospheres, providing greater insight into the formation and evolution of the solar system. *See* IONOSPHERE; PLANETARY PHYSICS; SOLAR SYSTEM.

The combined results of measurements from Earth, lunar, and other planetary satellites, space probes, and crewed lunar missions have added significantly to scientists' knowledge of Earth and other planets. For example, these observations have shown that for 5×10^8 years or more after their formation the planets were subjected to bombardment by huge asteroid-sized meteorites; that Mars is a very active planet; and that large bodies like the Moon and planets will necessarily undergo substantial, even violent, evolution after their formation, so that much of the record of the very earliest days of the solar system will have to be sought from smaller objects like comets and asteroids, primarily by using space probes such as *Deep Impact*, which triggered a substantial cometary explosion on July 4, 2005.

Satellites also make possible precision measurements of the size and shape of Earth and its gravitational field, and even measurements of the slow drifting of continents relative to each other. *See* CONTINENTAL DRIFT; EARTH, GRAVITY FIELD OF; GEODESY.

Astronomy. Satellites have also had a great influence on astronomy by making it possible to observe the Sun, stars, and other celestial objects in all the wavelengths that reach the vicinity of Earth, whereas the ionosphere and atmosphere prevent most of these radiations from reaching telescopes on the ground. Moreover, even in the visible wavelengths, the small-scale turbulence and continuously varying refraction in the lower atmosphere distort the image of a stellar object viewed at the ground, so that there is a limit to the improvement that can be achieved by increasing the size of a ground-based telescope. It has been estimated that a 12-in.-diameter (30-cm) optical telescope in a satellite is as effective in resolution as a 100-in.-diameter (254-cm) telescope at Earth's surface. Also, the Hubble Space Telescope with its 94-in. (2.4-m) mirror is living up to its promise of making observations of objects nearly 100 times fainter, or 10 times farther away, than can now be observed with the best ground-based instruments. *See* HUBBLE SPACE TELESCOPE; TELESCOPE.

The importance of being able to look deeper into space arises from the fact that the farther away an object is, the longer its light has taken to get to Earth. Since the radiation from a celestial object reveals the object as it was when emitting that radiation, the more distant objects will let scientists look farther back in time. By increasing the distance that can be seen into space, the chances are increased that the time when the universe was generated may be seen.

The importance of being able to observe in all wavelengths arises from the fact that theory shows that vital information about the birth, evolution, and demise of celestial objects is contained in the infrared, ultraviolet, and other wavelengths that do not reach the ground.

With satellite instruments a hitherto unobtainable ultraviolet map of the sky has been made. Hundreds of x-ray sources have been discovered. Huge clouds of hydrogen, millions of miles across, have been detected around comets. The Sun has been shown to be even more complex than previously supposed, and a great deal of information has been obtained about sunspots, solar flares, and solar magnetic fields. Numerous solar observatories have been launched, along with radio, infrared, ultraviolet, x-ray, and gamma-ray astronomy satellites. *See* ASTRONOMY; COMET; GAMMA-RAY ASTRONOMY; INFRARED ASTRONOMY; RADIO ASTRONOMY; SUN; ULTRAVIOLET ASTRONOMY; X-RAY ASTRONOMY.

Applications satellites. The applications satellites emphasize the continuing day-to-day practical utilization of the satellite. They are operational in nature, although some of them, either directly or as a by-product of their operational output, contribute significantly to research, for example, with meteorological or Earth survey satellites. An applied research program generally precedes establishment of an applications satellite system. Thus, *Tiros* and *Nimbus* laid the groundwork for operational meteorological satellites. Likewise, *Syncom* satellites preceded the *Intelsat* communications system.

Applications satellites fall generally into two distinct classes: (1) those using satellite radio communications as the essential element; and (2) those that are directed toward providing observational or survey data. In the first class are the now-familiar radio and television relay satellites like *Intelsat* and DirecTV; and navigation, geodetic, data collection and relay, disaster warning, air-traffic control, and maritime communications satellites. In the second class are meteorological satellites and satellites for the survey and monitoring of Earth's natural and cultural resources in such areas as geology and mining, hydrology, glaciology, sea ice, agriculture, forestry, land use, and civil engineering.

The earliest commercial space field to become fully established is that of communications satellites, with which countries all over the world are involved. In a very real sense the satellite for communications substitutes for a very high antenna mast. Because of its height the satellite is particularly useful for long-distance communications. Since stations far apart are within the line of sight of the satellite, microwave techniques can be extended from a distance of a few tens to many thousands of miles, eliminating dependence on the highly variable ionosphere and greatly extending the range of frequencies which may be used for long-distance communications. For

these purposes the geostationary orbit altitude of 22,237 mi (35,786 km) has been found especially effective. *See* COMMUNICATIONS SATELLITE.

Navigational satellites are used to determine position on the surface of Earth, supplementing celestial and radio navigation techniques. Marine and aeronautical navigational systems based on the Global Positioning System are in operation. *See* SATELLITE NAVIGATION SYSTEMS.

The meteorological satellite system obtains data for the preparation of weather forecasts and for the detection and surveillance of unusual weather conditions such as hurricanes and tornadoes. The complete system consists of a number of satellites in low-altitude—several-hundred-mile—Sun-synchronous, near-polar orbits, and a number in geosynchronous orbits over the Equator. The system affords a constant awareness of weather conditions over virtually the whole world. Optical and infrared imaging furnish large-scale pictures of cloud patterns and weather systems, while other instruments give temperature profiles of the lower atmosphere and stratosphere and some estimates of humidity and winds. *See* METEOROLOGICAL SATELLITES.

Cartographic and geodetic satellites are used as adjuncts to conventional techniques. With cartographic satellites the object is to prepare maps of Earth's surface and of the culture upon it. The geodetic satellite aids in determining the shape of Earth's surface—the geoid. Cartographic satellites use optical camera techniques, essentially identical to those of aerial photography. Geodetic satellites may use optical techniques in which angles of vision of the satellite as seen from Earth stations are measured; radio techniques in which both distance and angles are measured; or indirect techniques in which the shape of the geoid is inferred from the satellite orbit, in particular from the perturbations of this orbit resulting from the deviations of the geoid from a perfect sphere. *See* CARTOGRAPHY.

Imaging of Earth in various wavelengths from the ultraviolet to the infrared and radio regions yields a wide range of detail about Earth's surface structure and physical properties, vegetation, and cultural development. Most of the imaging is done by passive techniques using radiations emitted by the object being viewed, but much useful work can be done with active methods like radar. In the broad perspective of images covering thousands or tens of thousands of square miles, features stand out that were not recognized on foot or from aircraft. Numerous previously unknown faults and fracture systems of the kind associated with mineral ore and oil deposits have been identified. The combination of satellite and ground techniques for geological survey and prospecting promises to be especially fruitful.

A staging technique in which accurate information about some resource—a timber stand, for example—obtained on the ground over a small area is extrapolated to a larger area by aircraft imaging, and then further extrapolated to a vastly greater area by satellite imaging, is proving to be very accurate, economical, and time-saving for surveys of agricultural crops, grazing lands, and forests. Total crop productions can be estimated in advance. Tree or crop diseases can be spotted early and the total extent and location of the infection can be estimated. Watershed managers can use the satellite images to help in assessing the amount of water in reservoirs and in mountain snow and ice. Sea ice, hazardous to shipping, can be kept under surveillance. Many states use satellite images to produce land-use maps which are important for helping to guide the continuing development of the state and its resources.

Orbit. The orbits for scientific and applications satellites depend on the requirements of the investigation or use, and also on the nature of the local environment which may adversely affect the purpose or life of the satellite. For example, a satellite to study the radiation belts would be placed in an orbit traversing the belts, but an astronomical satellite with sensitive electronic instruments should be kept clear of the radiation in the belts.

Below about 100 mi (160 km) air drag on a satellite is prohibitively great. Between 100 and 200 mi (160 and 320 km) altitude air drag is still appreciable and can significantly affect the orbit of a satellite. Above 200 mi (320 km) air drag may generally be ignored except for precision measurements. Ionospheric effects on radio signals may have to be corrected for, when using satellites for such purposes as navigation or geodetic studies. Even solar radiation pressure may have to be taken into account in using satellites for air density or geodetic measurements.

In general, the remarks on orbits for Earth satellites apply to satellites of other planets as well, but specific characteristics of satellite orbits must be calculated for each planet.

Spacecraft. Scientific satellites divide conveniently into two groups: the Explorer class (named after the early *Explorer* satellites) and the observatory class. In general, the Explorer-class satellites are special-purpose and lighter than the observatory class. Typical weights are from 150 to 500 lb (70 to 230 kg), although some may be less, some more. Many dozens of Explorer-class scientific and applications satellites have been launched. Those that made the early discoveries in the radiation belts were of the Explorer class.

Observatory-class satellites tend to be heavier than Explorer-class satellites, ranging from below 1000 lb (450 kg) to many tons in weight. They also usually have more complicated systems for power, telemetering, and orientation, and often combine a large number of experimental objectives in one mission. The orbiting solar observatories (OSOs), the orbiting astronomical observatories, and the *Nimbus* meteorological satellites are all examples of the observatory class. *See* SPACE COMMUNICATIONS; SPACE POWER SYSTEMS; SPACE TECHNOLOGY; SPACECRAFT STRUCTURE.

International interest. Since the Soviet Union and the United States pioneered the launching of space rockets, interest in the use of satellites for science and applications has become worldwide. Other countries, including Italy, France, Japan, China, and

India, have launched their own satellites. Over 70 nations have been associated in various ways with the United States space program, many of them with scientific and applications satellites, sometimes furnishing instruments for United States satellites, other times providing complete satellites for launching by the United States. Homer E. Newell; John F. Clark

Bibliography. S. Q. Kidder and T. H. Vonder Haar, *Satellite Meteorology: An Introduction*, 1995; National Research Council, *Mission to Planet Earth*, 1988; National Research Council, *Solar and Space Physics*, 1988; C. B. Pease, *Satellite Imaging Instruments*, 1991; G. Seeber, *Satellite Geodesy*, 1993; K.-H. Szekielda, *Satellite Monitoring of the Earth*, 1989; J. Wertz and W. Larson (eds.), *Space Mission Analysis and Design*, 3d ed., 1999.

Scientific methods

Strategies or uniform rules of procedure used in some scientific research with a measure of success. Scientific methods differ in generality, precision, and the extent to which they are scientifically justified. Thus, whereas the experimental method can in principle be used in all the sciences dealing with ascertainable facts, the various methods for measuring the electron charge are specific. The search for increasing quantitative precision involves the improvement or invention of special methods of measurement, also called techniques. All scientific methods are required to be compatible with confirmed scientific theories capable of explaining how the methods work. The most general of all the methods employed in science is called the scientific method.

The scientific method. All the sciences referring to real or putatively real things are supposed to abide by the scientific method. The latter may be summarized as the following sequence of steps: identification of a knowledge problem; precise formulation or reformulation of the problem; examination of the background knowledge in a search for items that might help solve the problem; choice or invention of a tentative hypothesis that looks promising; conceptual test of the hypothesis, that is, checking whether it is compatible with the bulk of the existing knowledge on the matter (for it might be a wild conjecture not worth pursuing); drawing some testable consequences of the hypothesis; design of an empirical (observational or experimental) test of the hypothesis or a consequence of it; actual empirical test of the hypothesis, involving a search for both favorable and unfavorable evidence (examples and counterexamples); critical examination and statistical processing of the data (for example, calculation of average error and elimination of outlying data); evaluation of the hypothesis in the light of its compatibility with both the background knowledge and the fresh empirical evidence; if the test results are inconclusive, design and performance of new tests, possibly using different special methods; if the test results are conclusive, acceptance, modification, or rejection of the hypothesis; if the hypothesis is acceptable, checking whether its acceptance forces some change (enrichment or correction) in the background knowledge; identifying and tackling new problems raised by the confirmed hypothesis; and repetition of the test and reexamination of its possible impact on existing knowledge. The case of the introduction and checking of new procedures is parallel, except that in this case reliability and accuracy, not truth, are at stake. So is the case of the introduction and checking of artifacts such as machines or organizations, except that in this case the checking is for efficiency. *See* SYSTEMS ENGINEERING.

An alternative interpretation of the scientific method is as a sequence of problems. A typical example is the empirical problem of finding a particular value of a certain magnitude M, or the parallel theoretical problem of explaining why the value of M happens to be m. The experimenter's problem is to determine the measured value of M with relative error smaller than a preassigned number e. The theorist's problem is to determine what premises imply that the value of M is m, or whether m is to be found only by empirical means. The next experimental problem is to determine which experimental device would make it possible to measure M with error less than e. The corresponding theoretical problem is to determine which theory, subsidiary hypotheses, and data imply that M is worth m. A subsequent experimental problem is to determine the value of M that a measurement with the help of the chosen device yields. A corresponding theoretical problem is to determine the value of M according to the theory, subsidiary hypothesis, and data. From here on, the questions are addressed to both experimenter and theorist. They seek to find what the result implies or suggests; how the result could be corroborated independently, that is, by alternative procedures; whether the new result is more precise and plausible than the results obtained by alternative procedures; if so, what the result implies or suggests; and, if not, what may have to be altered in the preceding operations.

Experimental method. When the scientific method involves experiment, that is, the deliberate controlled modification of some properties (factors or variables), it is called the experimental method. This method involves the design and operation of an experimental setup. Such a setup often includes one or more measuring instruments. When the objects of experiment are very similar, they are compared before and after the variable or variables in question have been altered. When the objects of experiment exhibit significant variations, as is the case with organisms, in particular humans, a large collection of them must be studied. Every such collection is divided into two roughly equal parts: the experimental and the control groups. To avoid bias, the members of each collection are picked at random and, in the case of humans, the controls are given placebos. (In more refined experiments the control group is split into two, only one of which is given a placebo, and the experimenter does not know beforehand who will be given what.) Only the members of the experimental group are subjected to the stimulus whose

effects the experimenter wishes to find out. The simplest experimental design involves the variation of one variable at a time. More complex experimental designs allow for the simultaneous variation of two or more variables. After the stimulus has been applied, the variables of interest are observed or measured in the two groups. If a difference is observed, a statistical significance test is applied to find out whether the difference is genuine or due to small individual differences or to random errors. The procedure is usually repeated by the same observer or by an independent investigator to check for possible errors in design, execution, or interpretation. *See* EXPERIMENT; STATISTICS.

Discovering, inventing, and checking. The scientific method is not a recipe for making original discoveries or inventions; it does not prescribe the pathway that scientists must follow to attain success. Nor is it a substitute for creativity and resourcefulness. The actual pathway of scientific research is messy and partly at the mercy of unforeseen accidents, both lucky and unlucky. Moreover, this pathway depends on the subject matter as well as on the individual's imagination and experience. The goal of the scientific method is to ascertain whether a hypothesis is true to some degree. Indeed, the nucleus of the scientific method is the confrontation of an idea (hypothesis) with the facts it refers to, regardless of the source of the idea in question. In sum, the scientific method is a means for checking hypotheses for truth rather than for finding facts or inventing ideas.

The scientific method has not always been understood in this fashion. For instance, Francis Bacon and many others thought that the scientific method is a simple and guaranteed recipe for discovery and invention. But in fact no one has come up with any algorithms for having original ideas. (This is one of the reasons that it is impossible at present to design creative computers.) Others have denied the very existence of the scientific method.

Since this is a question concerning facts, it can only be resolved by an empirical investigation concerning the way that people actually conduct scientific research, regardless of what they say about method. Such investigation is bound to show that the answer depends on the kind of problem. Routine problems can indeed be handled with the help of well-tried special methods (techniques) that can be described almost exhaustively, so much so that their application can be entrusted to apprentices or technicians. Original problems are something utterly different. It requires originality and experience to find an open and interesting problem that can be investigated in a finite time with limited resources. Once a promising problem has been identified, it must be stated or perhaps reformulated in a precise way, and a tentative plan to work on it must be formulated—again, an original task. The invention of new hypotheses and new methods also takes inspiration and luck in addition to hard work and discussions with colleagues. Finally, the hypothesis must be checked in order to find out whether it is true to some degree, and the method must be tried out to ascertain whether it accomplishes what it purports do and, if so, whether it is better than rival techniques. It is here, in checking, that the scientific method plays a decisive and distinctive role. In short, the scientific method is the way that scientists proceed to check ideas and techniques, not invent them.

Reach of the scientific method. Whether the scientific method can legitimately and fruitfully be used in the social sciences, such as history, and in the socionatural or hybrid sciences, such as psychology and linguistics, has been a subject of controversy. The history of these sciences shows conclusively that the scientific method has been fruitful wherever it has actually been employed.

Test of scientificity. Since the use of the scientific method is one characteristic of scientific research, its absence is a sure indicator of nonscience. In other words, a discipline where the scientific method plays no role is not a science. Thus, such fields as theology, literary criticism, psychoanalysis, homeopathy, graphology, and palmistry can hardly be regarded as scientific. *See* EMPIRICAL METHOD; HYPOTHESIS; SCIENCE.

Mario Bunge

Bibliography. M. Bunge, *Epistemology and Methodology I: Exploring the World*, 1983; M. Bunge, *Epistemology and Methodology II: Understanding the World*, 1983; H. Feigl and M. Brodbeck (eds.), *Readings in the Philosophy of Science*, 1953; E. Nagel, *The Structure of Science*, 1961, reprint 1979; A. Zucker, *Introduction to the Philosophy of Science*, 1996.

Scintillation counter

A particle or radiation detector which operates through emission of light flashes that are detected by a photosensitive device, usually a photomultiplier or a silicon PIN diode. The scintillation counter not only can detect the presence of a particle, gamma ray, or x-ray, but can measure the energy, or the energy loss, of the particle or radiation in the scintillating medium. The sensitive medium may be solid, liquid, or gaseous, but is usually one of the first two. The scintillation counter is one of the most versatile particle detectors, and is widely used in industry, scientific research, medical diagnosis, and radiation monitoring, as well as in exploration for petroleum and radioactive minerals that emit gamma rays. Many low-level radioactivity measurements are made with scintillation counters. *See* LOW-LEVEL COUNTING.

The scintillation phenomenon was used by E. Rutherford and colleagues, who observed the light flashes from screens of powdered zinc sulfide struck by alpha particles. The scattering of alpha particles by various foils was used by Rutherford to establish the modern notion of the nuclear atom. Since 1947, photomultipliers or PIN diodes, which feed into amplifiers, have replaced the eye and have also proved to be very efficient in measuring the time of arrival of a particle, as well as its energy or energy loss. Scintillations are light pulses emitted in the electronic decays of excited atoms or molecules, the

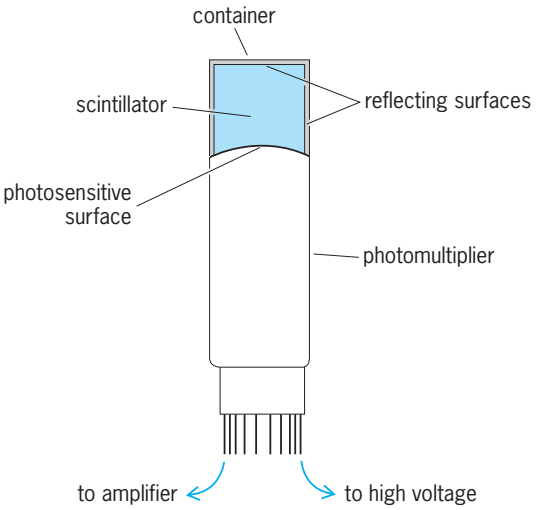


Diagram of a scintillation counter.

energy responsible for these excitations having been provided, directly or indirectly, by the detected particle. Scintillations are one example of luminescent behavior, and are related to the process called fluorescence. *See* FLUORESCENCE; LUMINESCENCE; PHOTOMULTIPLIER.

Operation of bulk counter. Following the innovation in 1947 of using bulk material, instead of screens of powdered luminescent materials, scintillation counters have been made of transparent crystalline materials, liquids, plastics, or glasses. In order to be an efficient detector, the bulk scintillating medium must be transparent to its own luminescent radiation, and since some detectors are quite extensive, covering meters in length, the transparency must be of a high order. One face of the scintillator is placed in optical contact with the photosensitive surface of the photomultiplier or PIN diode (see **illus.**). In order to direct as much as possible of the light flash to the photosensitive surface, reflecting material is placed between the scintillator and the inside surface of the container.

In many cases it is necessary to collect the light from a large area and transmit it to the small surface of a photomultiplier. In this case, a "light pipe" leads the light signal from the scintillator surface to the photomultiplier with only small loss. The best light guides and light fibers are made of glass, plastic, or quartz. It is also possible to use lenses and mirrors in conjunction with scintillators and photomultipliers. *See* OPTICAL FIBERS.

A charged particle, moving through the scintillator, loses energy and leaves a trail of ions and excited atoms and molecules. Rapid interatomic or intermolecular transfer of electronic excitation energy follows, leading eventually to a burst of luminescence characteristic of the scintillator material. When a particle stops in the scintillator, the integral of the resulting light output, called the scintillation response, provides a measure of the particle energy, and can be calibrated by reference to particle sources of known energy. Photomultipliers or PIN diodes may be operated so as to generate an output pulse of amplitude proportional to the scintillation response.

When a particle passes completely through a scintillator, the energy loss of the particle is measured. When a gamma ray converts to charged particles in a scintillator, its energy may also be determined. When the scintillator is made of dense material and of very large dimensions, the entire energy of a very energetic particle or gamma ray may be contained within the scintillator, and again the original energy may be measured. Such is the case for energetic electrons, positrons, or gamma rays which produce electromagnetic showers in the scintillator. Energy spectra can be determined in these various cases by using electronic equipment to convert amplitudes of the output pulses from the photomultiplier or PIN diode to digital form, for further processing by computers or pulse-height analyzers.

Characteristics. Scintillation counters have several characteristics which make them particularly useful as detectors of x-rays, gamma rays, neutrons, and other nuclear and high-energy particles.

Efficiency and size. Charged particles, x-rays, and low-energy gamma rays can be detected with efficiencies close to 100%. Small, thin detectors may be used to define a particle's position accurately. When detection of very energetic particles is desired, the scintillator can be made very large and massive, and viewed by many photomultipliers.

Speed. Useful scintillation materials emit light flashes which rise to peak intensity in less than 10^{-9} s and decay to half of this intensity in 10^{-9} s to 10^{-6} s. The rapid rise makes scintillators useful for fast timing measurements. Short decay times allow very fast counting while avoiding pileup, which is very characteristic of older and slower particle detectors such as the Geiger counter. The Cerenkov counter is similar to a scintillation counter, but depends only on the index of refraction of the transparent medium and the particles' velocity in the medium, and not on luminescence processes. *See* CERENKOV RADIATION.

Energy resolution. While the energy resolution attained by scintillation counters is useful, their performance in this respect is markedly inferior to that of semiconductor detectors such as silicon, germanium, or lithium-drifted germanium at low temperatures. However, semiconductor counters have been rather limited in size, so that they do not usually compete with the large inorganic scintillators needed in high-energy research. Often coincidence or anticoincidence combinations are made between semiconductor counters and scintillation counters. *See* JUNCTION DETECTOR.

Substances used. Scintillation counter materials are generally classified into crystal scintillators, glass scintillators, and organic scintillators. The organics include crystal, liquid, and plastic types.

Inorganic scintillators. Inorganic crystal scintillators are generally characterized by the presence of heavy elements. The most widely used inorganic scintillator is sodium iodide activated with a small amount of thallium salt. The usual designation of this material is NaI(Tl). Cesium iodide is also useful, and may be activated with thallium or sodium, or used in pure form. Other useful crystalline scintillators include calcium fluoride, barium fluoride, cadmium

tungstate, and bismuth germanate (BGO). NaI(Tl) and bismuth germanate are particularly useful for detecting gamma rays because of the presence of iodine and bismuth, respectively; the high atomic numbers of these elements (53 and 83, respectively) are important in both the photoelectric and pair production processes which result from the interaction of a gamma ray with the base material. *See* GAMMA-RAY DETECTORS.

Glass scintillators. Glass scintillators formed from lithium silicate and containing cerium as activator are used to detect thermal and low-energy (less than 0.1 MeV) neutrons. Glasses containing the isotope ^{6}Li are sensitive to neutrons through the charged particles released in the nuclear reaction $^6\mathrm{Li}(n,t)^4\mathrm{He}$, which has a high cross section for thermal neutrons. Glasses of similar chemical composition but containing lithium depleted in the ^{6}Li isotope are insensitive to neutrons and can be used to monitor backgrounds due to other types of radiation.

Organic scintillators. Liquid and plastic scintillators are the most extensively used forms of organic scintillator. Organic crystals such as anthracene and stilbene are also used, but the liquids and plastics are favored for the ease with which they may be formed in arbitrary shapes, large or small, and for the economy they offer in achieving large, sensitive detecting volumes. The timing precision achieved by the faster liquids and plastics is typically in the range of 10^{-9} to 10^{-10} s. The solvents which form the bulk constituents of liquid and plastic scintillators are often xylene, toluene, benzene or mineral oil (liquids), or polystyrene or polyvinyltoluene (plastics). Other constituents of the liquid or plastic solutions are compounds such as terphenyl, diphenyloxazole, and 1,4-bis-{2-(5-phenyloxazolyl)}-benzene, known as POPOP.

Liquid scintillators are used in volumes ranging from a few tenths of a cubic inch or less through several hundreds of cubic feet. The largest scintillators are used in high-energy physics to study neutrino interactions. Liquid scintillators also have the capacity to assimilate other liquids or substances and form homogeneous media, thereby providing an efficient and simple means for measuring the products of nuclear reaction or radioactive decay in those substances. Plastic scintillators are used in a variety of forms ranging from films a few micrometers thick, or fibers less than 0.01 in. (0.25 mm) in diameter, to large sheets or blocks up to about 15 ft^3 (0.5 m^3) in volume. *See* NEUTRINO.

Pulse-shape discrimination. The response of most liquid scintillators can be improved significantly if dissolved oxygen is either removed from the solution or displaced by nitrogen or an inert gas. This procedure also endows many liquid scintillators with the ability to pulse-shape-discriminate, that is, to exhibit scintillation decay properties which depend on the type of ionizing particle causing the scintillation. Pulse-shape discrimination is also a strong effect in anthracene and stilbene crystals, but not in plastic scintillators. Most organic scintillators contain a high proportion of hydrogen, and therefore detect fast neutrons (energy greater than 0.1 MeV) via the proton recoils generated by internal neutron-proton scattering. Pulse shape discrimination may be used to select these proton scintillations while rejecting others, especially those caused by gamma rays. Liquid and crystalline organic scintillators are therefore very effective in detecting and identifying neutrons, even when the background of gamma radiation is high. The response of organic scintillators to electrons is linear, that is, proportional to electron energy. However, their response to heavier particles is nonlinear; and for different particles of the same energy, the response is smaller, the heavier the particle.

Additions of compounds and samples. Appreciable quantities of compounds of a number of elements can be incorporated in liquid scintillators (in appropriate chemical form) without degrading the scintillation performance excessively. Some examples are lithium, boron, gadolinium, or cadmium compounds, which facilitate the detection of low-energy neutrons (less than about 0.1 MeV) through neutron capture in these elements. Organic liquid scintillators are extensively used in the medical, biological, and environmental fields to detect radioactive samples and tracer compounds labeled with the beta emitters ^{14}C or ^{3}H, or with other radionuclides. A carrier solution containing the sample or tracer compound is assimilated or dissolved in the liquid scintillator to form a uniform counting sample or scintillating medium with a volume of a few milliliters. Such counting samples are often processed in large batches by using sophisticated instrumentation which incorporates automatic sample changing and the capacity to carry out all the necessary data reduction automatically. *See* RADIOACTIVE TRACER.

Manufacture and handling. The handling of liquid scintillators requires extreme care, because many are highly inflammable and because small concentrations of common impurities can drastically impair the scintillation performance by quenching the energy-transfer mechanism of the scintillation process. Commercial manufacturers offer a wide variety of crystal, glass, liquid, and plastic scintillators, which may also be tailored to specific requirements. Liquid scintillators can be packaged in a variety of forms, for example, bottled in bulk form or encapsulated in ready-to-use glass or metal-and-glass containers. NaI(Tl) crystals must be hermetically isolated from the atmosphere because they are hygroscopic and deteriorate rapidly if exposed to moisture. They are often supplied optically coupled to a photomultiplier, in a sealed canister.

Applications. Fundamental discoveries in physics and astrophysics have been made with scintillators, and scintillation counters have found application in fields as diverse as nuclear physics, elementary particle physics, atomic physics, solid-state studies, chemistry, nutrition studies, biology, nuclear medicine, astrophysics and space physics, geology, nuclear reaction technology and monitoring, petroleum and mineral resources exploration, and oceanography. *See* PARTICLE DETECTOR.

Frank D. Brooks; Robert Hofstadter

Bibliography. J. B. Birks, *The Theory and Practice of Scintillation Counting*, 2d ed., 1967; F. D. Brooks, Development of organic scintillators, *Nucl. Instrum. Meth.*, 162:477–505, 1979; R. L. Heath, R. Hofstadter, and E. B. Hughes, Inorganic scintillators: A review of techniques and applications, *Nucl. Instrum. Meth.*, 162:431–476, 1979.

Scleractinia

An order of the subclass Zoantharia which comprises the true or stony corals (**Fig. 1***a*). These are solitary or colonial anthozoans which attach to a firm substrate. They are profuse in tropical and subtropical

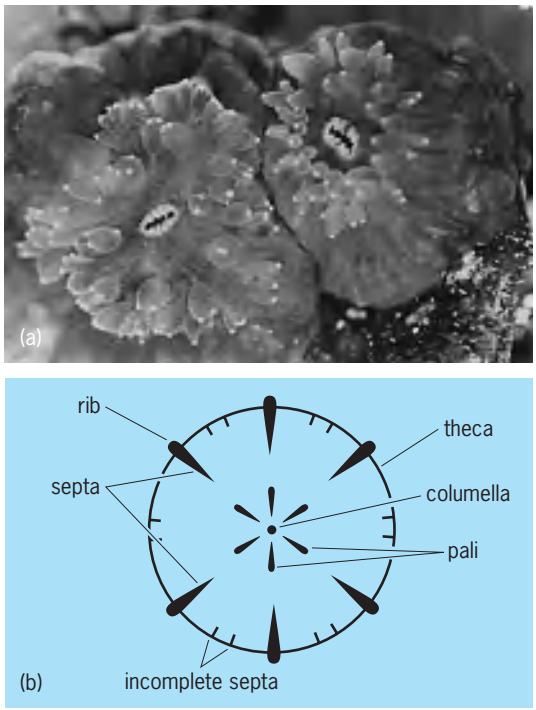


Fig. 1. Solitary corals. (*a*) Coral polyps, *Oulangia* sp. (*b*) Diagrammatic figure, cross section.

Fig. 2. Skeletons of aggregated colonies of *Pocillopora damicornis caespitosa* formed 15 days after fixing.

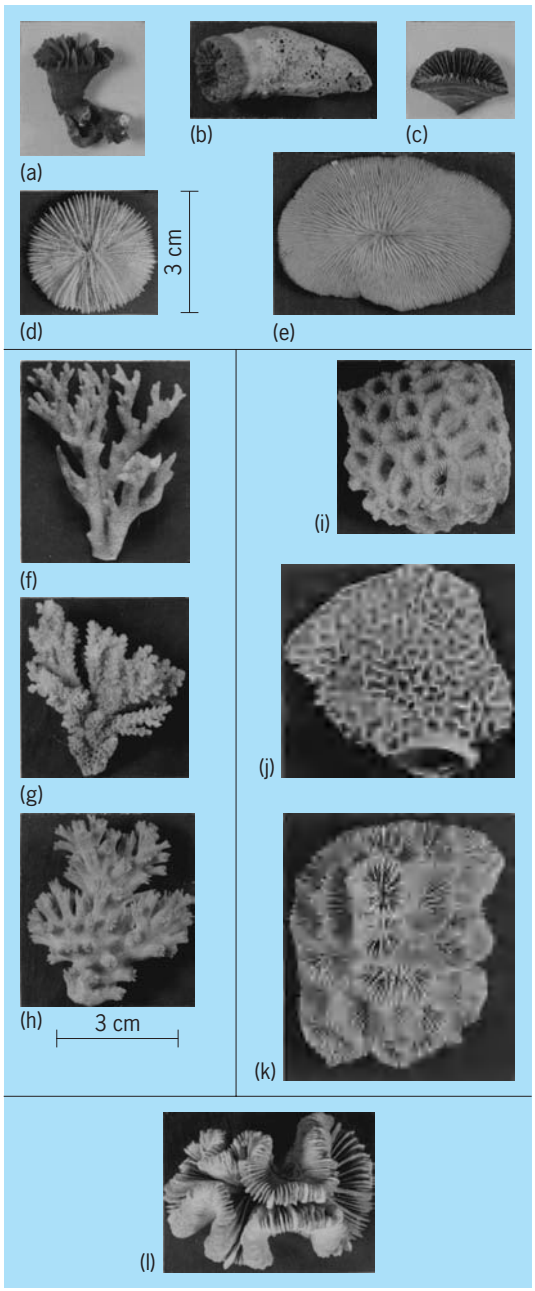


Fig. 3. Coral skeletons. (*a*) *Desmophyllum dianthus*. (*b*) *Balanophyllia gigas*. (*c*) *Flabellum distinctum*. (*d*) *Fungia actiniformia palauensis*. (*e*) *Fungia scutaria*. (*f*) *Pocillopora acuta*. (*g*) *Acropora hyacinthus*. (*h*) *Acrhelia horrescens*. (*i*) *Favia pallida*. (*j*) *Platygyra lamellina*. (*k*) *Acanthastrea echinata*. (*l*) *Trachyphyllia geoffroyi*.

waters and contribute to the formation of coral reefs or islands. Some species are free and unattached.

Most of the polyp is impregnated with a hard calcareous skeleton secreted from ectodermal calcioblasts. As soon as the planula settles, it secretes a thin skeletal basal plate, from which radiating vertical platelike partitions or septa arise; then characteristic thecal formation follows. The basal parts of the inner septal edges are often fused to form a columella (**Fig. 2**), which is sometimes ridged by vertical thin plates or pali. The outer septal edges, projecting

beyond the theca, are the costae or ribs. Such is the outline of the corallum. Various other structures develop, such as the endo- or exothecae which comprise the dissepiments. They are inside and outside the wall of the corallum, respectively (Fig. 1*b*).

Skeleton. The solitary corals form cylindrical, discoidal, or cuneiform skeletons (**Fig. 3***a–e*), whereas colonial skeletons are multifarious. *Stylophora*, *Seriatopora*, *Pocillopora*, and *Acropora* form branching skeletons (Fig. 3*f–h*), while the skeletal systems of *Favia*, *Favites*, *Porites*, *Coeloseris*, and *Goniastrea* are massive (Fig. 3*i–k*). In others, such as *Echinopora*, *Plerogyra*, *Lobophyllia*, and *Trachyphyllia* (Fig. 3*l*), all sorts of shapes are found.

Colony formation. The polyps increase rapidly by intra- or extratentacular budding, and the skeletons of polyps which settle in groups may fuse to form a colony (Fig. 2). The pyriform, ciliated planula swims with its aboral extremity, which is composed of an ectodermal sensory layer, directed anteriorly. Planulation occurs periodically in conformity with lunar phases in many tropical species. *See* ZOANTHARIA.

Kenjai Atoda

Fossils. Scleractinian corals possess robust skeletons, so they have a rich fossil record. Because they are restricted mainly to tropical belts, they help indicate the position of the continents throughtout the Mesozoic and Cenozoic periods. They are also important for understanding the evolution of corals and the origin and maintenance of reef diversity through time. Pleistocene corals shows persistent reef coral communities throughout the last several hundred thousand years. Environmental degradation has led to the dramatic alteration of living coral communities during the past several decades.

Geologic record. The first Scleractinia appeared during the Middle Triassic (**Fig. 4**). Corals may have had multiple origins at the same time that they evolved skeletons, or they may have evolved much earlier (during the Paleozoic) and multiple groups acquired skeletons at the same time in the Triassic. Recent molecular work points to the evolution of the Scleractinia well before they first appeared some 240 million years before present. If true, soft-bodied Scleractinia existed in Paleozoic seas with the skeletonized rugose and tabulate corals. The Permian extinction event wiped away the rugose and tabulate corals, paving the way for scleractinians to evolve skeletons.

Scleractinians played a prominent role in Triassic reefs, where they were highly zoned. The end-Triassic extinction event led to the collapse of the first Mesozoic reef ecosystem with its new scleractinian components. The middle Oxfordian (Late Jurassic) marked an important phase of reef expansion. Reefs became so widespread that large and extensive reef belts of the size of the modern Australian Great Barrier Reef were developed in western Europe and the Caucasus belt. Late Jurassic reefs were zoned into fore-reef and back-reef facies, and atolls were common. These reefs are now preserved in geological sections up to several hundred meters thick. The all-time global maximum of coral diversity occurred in the Late Jurassic with over 200 genera recorded.

The Cretaceous continued the Upper Jurassic pattern, but in the Barremian (Early Cretaceous) coral-mimicking rudist bivalves began to compete with and replace the scleractinians as the dominant reef biota. Corals generally occupied separate habitats at greater depths than earlier. Corals returned to dominance in the Late Cretaceous, following extinction of the rudistids. By the close of the Cretaceous, extensive coral reefs occurred worldwide. Few coral reefs are recorded from the Paleocene. But from the Eocene to Recent, scleractinian coral-dominated reef tracts occur in tropical belts.

Environmental perturbation. Corals and reefs are subject to a number of stresses: seasonal temperatures above 30°C (86°F) or below 20°C (68°F); sharp seasonal temperature fluctuations (for example, El Niño); high input of land-derived muds, clays, ash, and such; low salinity or hypersalinity; low light levels caused by suspended detritus; high nutrient input; rapid sea-level change; excessive tidal range and storm events; changes in local shelf topography; and biological agents such as disease, plankton blooms, and outbreak predator populations (such as crown-of-thorns starfish). Triassic coral reefs were disrupted by the loss of shallow-water habitat, and climate change at the end of the Cretaceous appears to have been the driving force in the gradual extinction of corals. Cenozoic reefs have been most affected by ocean current and temperature fluctuations induced by climatic change. Successive Plio-Pleistocene glaciations of the past 2–3 million years and the accompanying multiple sea-level fluctuations suggest that the Cenozoic was a relatively turbulent setting for reef development.

Distribution. The earliest scleractinians were not reef-dwelling but occupied the ancient Tethys Sea near the present-day Bavarian, Austrian, and Italian Alps, and Indo-China. Middle and Late Triassic corals were widespread throughout the Tethys region and around most of the equatorial Panthalassa Ocean rim (including India, Australia, New Guinea, and western North and South America) [**Fig. 5**]. In the Middle Jurassic, reef development proliferated in present-day Europe and the Mediterranean but remained poorly developed in Panthalassa. Late Jurassic reefs extended from Greenland and Sakhalin in the north to Chile and Argentina in the south. By this time, there were well-established coral provinces. An Asiatic province extended along the northern margin of Tethys, with subprovinces along the east Asian coast; a southern Tethys province extended to Panthalassa South America; and a northern Tethys province extended to Panthalassa North America. The paleobiogeographic pattern that developed was the precursor to the Tethys/Atlantic/eastern Pacific pattern of the Paleogene. During the Miocene, reef development was at a global maximum for the Cenozoic. The continents were close to their present position, and reef growth occurred throughout the tropical seas. Pleistocene and modern reef

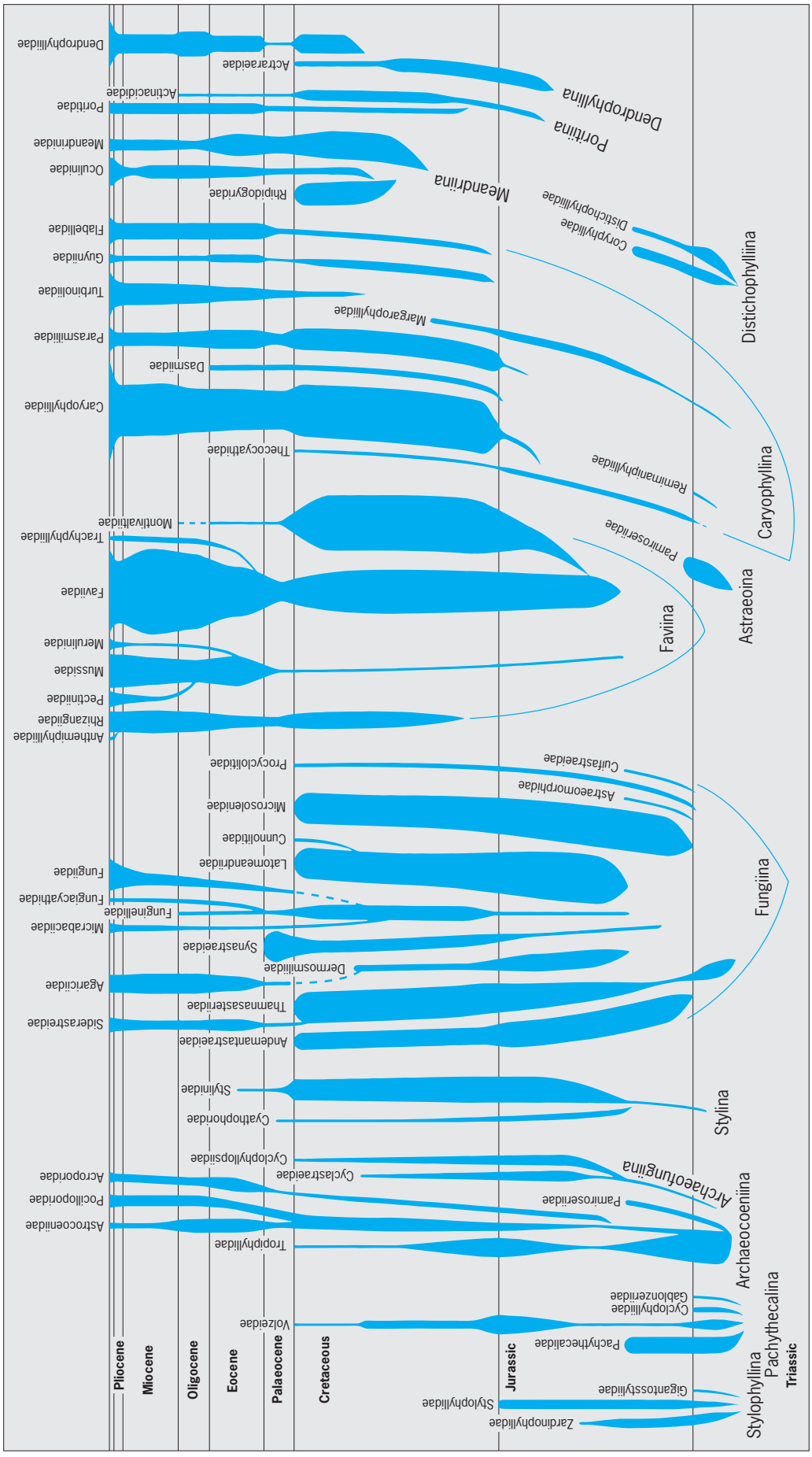


Fig. 4. Family-level phylogeny of the Scleractinia through geological time. Branch widths indicate numbers of genera per family for each geological interval. (*From J. E. N. Veron, Corals in Space and Time: The Biogeography and Evolution of the Scleractinia; copyright © 1995 J. E. N. Veron and the Australian Institute of Marine Science; used by permission of the publisher, Cornell University Press*)

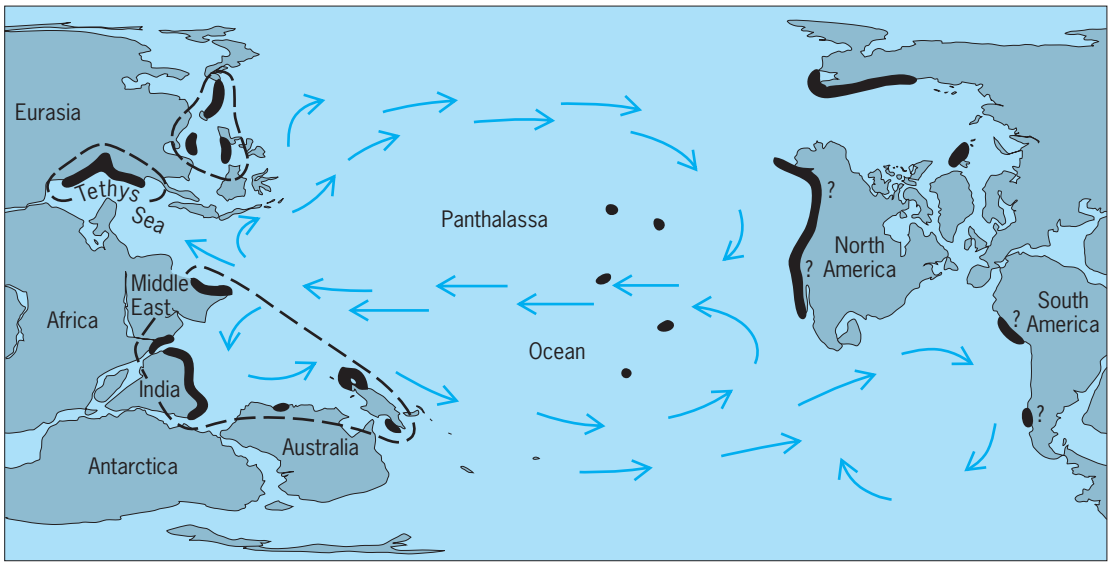


Fig. 5. Late Triassic distribution of corals and coral reefs. Broken lines encompass the major coral provinces. (*From J. E. N. Veron, Corals in Space and Time: The Biogeography and Evolution of the Scleractinia; copyright © 1995 by J. E. N. Veron and the Australian Institute of Marine Science; used with permission of the publisher, Cornell University Press*)

development is best developed between 30°N and 30°S latitude. John M. Pandolfi

Bibliography. P. Copper, Enigmas in Phanerozoic reef development, *Mem. Ass. Australasian Palaeontol.*, 8:371–385, 1989; J. A. Fagerstrom, *The Evolution of Reef Communities*, Wiley, New York, 1987; S. L. Romano and S. R. Palumbi, Evolution of scleractinian corals inferred from molecular systematics, *Science*, 271:640–642, 1996; J. E. N. Veron, *Corals in Space and Time: The Biogeography and Evolution of the Scleractinia*, University of New South Wales Press, Sydney, 1995.

Sclerenchyma

Single cells or aggregates of cells whose principal function is thought to be mechanical support of plants or plant parts. Sclerenchyma cells have thick secondary walls and may or may not remain alive when mature. They vary greatly in form and are of widespread occurrence in vascular plants. Two general types, sclereids and fibers, are widely recognized, but since these intergrade, the distinction is sometimes arbitrary.

Sclereids. These range from isodiametric to much elongated cells and may be branched. They may occur as isolated cells (idioblasts), groups of cells, or extensive tissues. Five major kinds of sclereids are described here: brachysclereids, macrosclereids, osteosclereids, astrosclereids, and trichosclereids.

Brachysclereids (stone cells) have the form of the parenchyma cells from which they are derived by secondary sclerosis (hardening). They have conspicuous, frequently ramiform (branched) pits in their thickened walls and are found in the shells of nuts, the pits of stone fruits, the bark of many trees, and the xylem and pitch of some plants. Nests of brachysclereids form the grit in the flesh of pears and quinces (**illus.** *a*).

Macrosclereids are rodlike cells formed from the protoderm (embryonic epidermis) of certain seed coats (illus. *b*). Their secondary walls are unevenly thickened.

Osteosclereids, or "bone" cells, are rodlike with swollen ends. They occur in seed coats and in some leaves.

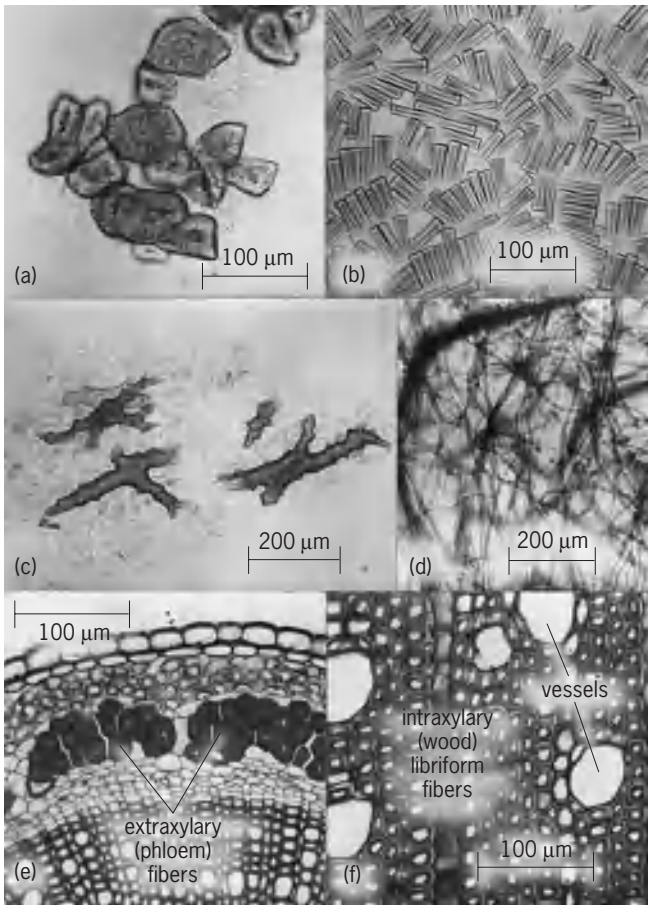
Astrosclereids tend to be radiately branched but are otherwise quite variable (illus. *c*). They occur in the leaves of many plants (*Trochodendron*, *Pseudotsuga*, and *Mouriria*) and in the petioles of *Camellia*. Astrosclereids sometimes have crystals embedded in their walls (*Castalia*).

Trichosclereids (illus. *d*) are long and slender, resembling fibers, with which they intergrade (leaves of *Monstera* and olive).

In general, foliar sclereids are idioblastic. They often arise from initials in the ground meristem. Branched forms extend themselves into intercellular spaces and even between cells. Sometimes they are associated with the ends of veinlets.

Fibers. Typically, fibers have tapering ends and are very long in proportion to their width. They usually occur as coherent strands of tissue and are rarely idioblastic. Individual fiber cells vary in length from a few millimeters to more than half a meter (*Boehmeria nivea*). Their thick walls are often sparsely pitted, and the pits usually appear simple. Fibers are widely distributed both in the primary and in the secondary body of vascular plants. Two broad types are generally recognized: intraxylary, or wood, fibers occur in the xylem (illus. *e*); extraxylary, or bast, fibers are found in the phloem and cortex of dicotyledons (illus. *f*) and in association with vascular bundles in the stems and leaves of monocotyledons.

Libriform (elongated, thick-walled) wood fibers (illus. *f*), which may constitute as much as 50% of an angiosperm wood, are so called because of their structural resemblance to phloem fibers. They often intergrade with tracheary (water-conducting)



Sclerenchyma. (*a*) Brachysclereids (stone cells) from fruit of pear. (*b*) Macrosclereids (rod cells) from macerated seed coat of bean. (*c*) Branched sclereids from petiole of *Camellia*. (*d*) Fiberlike sclereids in cleared leaf of olive. (*e*) Transection of a portion of stem of *Linum* showing the extraxylary (phloem) fibers. (*f*) Transection of wood of *Cornus* showing the intraxylary (wood) libriform fibers and vessels.

elements. Intermediate forms are called fiber tracheids. Septate wood fibers and fiber tracheids have transverse partitions which develop after the secondary wall is laid down (*Vitis* and *Hypericum*). Gelatinous or mucilaginous fibers have hygroscopic cell walls. They are found in many angiosperm woods such as oak (*Quercus*) and black locust (*Robinia*).

The fibers of commerce, excepting cotton and kapok (which consist of unicellular hairs), are bast fibers. The leaf fibers from monocotyledons, such as sisal (*Agave sisalina*), henequen (*A. fourcroydes*), and abaca (*Musa textilis*), are known as hard fibers. They are lignified and coarse and are used in the manufacture of cordage. The soft, or stem, fibers include jute (*Corchorus*), true hemp (*Cannabis*), ramie (*B. nivea*), and flax (*Linum usitatissimum*). Flax fibers, the raw material for linen, contain little or no lignin. *See* NATURAL FIBER; PHLOEM; PLANT ANATOMY; XYLEM.

Norman H. Boke

Sclerosponge

A class of sponges that lay down a compound skeleton comprising an external, basal mass of calcium carbonate, either aragonite or calcite, and internal siliceous spicules and protein fibers. The living tissue forms a thin layer over the basal calcareous skeleton and extends into its surface depressions; the organization of the tissue is similar to that of encrusting demosponges. Sclerosponges are common inhabitants of cryptic habitats on coral reefs in both the Caribbean and Indo-Pacific biogeographic regions. Their discovery has reopened questions of the affinities of certain fossil reef–inhabiting organisms, the stromatoporoids and the chaetetid and favositid tabulate "corals," all of which have been generally regarded as members of the phylum Cnidaria. *See* DEMOSPONGIAE.

Astrosclera. This sponge, first described in 1900 by J. Lister, is notable for its reticulate spherulitic skeleton of aragonite, in the interstices of which the living tissues occur. Siliceous spicules in the form of acanthostyles are also present, and some of these may be incorporated into the calcareous moiety of the skeleton.

Stromatoporoid affinities. *Astrosclera* resembles the stromatoporoids (1) in general aspect as a calcareous, mound-shaped organism varying in diameter from 0.4 to 6 in. (1 to 15 cm), (2) in the reticulate pattern of its fibrous aragonitic skeleton, (3) in the occurrence of astrorhizae (**Fig. 1**) at the surface of the skeleton, (4) in the absence of regularly repeated zooidal units in the skeleton, and (5) in the presence of an epitheca showing concentric growth lines on the lower side of the animal. In postulating a relationship between *Astrosclera* and stromatoporoids it must be remembered that a single relict species is being compared with a host of organisms comprising about 2000 described species that have existed for 4×10^8 years, during which time they have radiated into many divergent evolutionary lines. *Astrosclera* is an end point of one of these lines.

The reticulate skeleton of *Astrosclera* resembles the skeletons of some Paleozoic and most Mesozoic stromatoporoids in general aspect except that it lacks tabulae. In large Pacific specimens of *Astrosclera*, 6 in. (15 cm) across and 3 in. (8 cm) high, living tissue extends to a depth of only about 0.2 in. (5 mm) into the interstices of the skeleton; the basal skeletal mass is filled in by secondary deposits of aragonite. Instead of the intermittent secretion of tabular partitions to wall off the interior of the skeleton, in *Astrosclera* a continuous secretion of aragonite occurs internal to the living tissue.

In Mesozoic stromatoporoids the skeletal microstructure is fibrous in nature, with crystalline units oriented at right angles or obliquely to linear centers of calcification. A fibrous microstructure is uncommon among Paleozoic stromatoporoids, in which compact and cellular types predominate. These compact and cellular microstructures may result from diagenetic processes, however, as may the occurrence of calcite in fossil forms. A fibrous aragonitic skeleton is probably the primary condition in stromatoporoids. In the spherulitic microstructure of *Astrosclera* the aragonitic crystalline units are oriented around a point (**Fig. 2**), whereas in

stromatoporoids they are formed around linear centers of calcification.

The most striking similarity between *Astrosclera* and stromatoporoids is the presence of astrorhizae on the surface of the skeletal mass. The astrorhizae of *Astrosclera* and Mesozoic stromatoporoids are comparable in minute detail. In both, into the base of the surface grooves that shelter exhalant channels leading to a central oscule, there open oblique skeletal tubes carrying smaller exhalant channels that join the main surficial ones. The walls of these tubes look like tabulae lying across the base of the grooves. Vertical skeletal tubes join the central oscular region from below, bringing exhalant channels to the surface from deeper areas. Vertical series of astrorhizae, each lying on a previous growing surface and joined by vertical tabulate tubes, characterize many Paleozoic stromatoporoids with seasonal growth patterns; such features are absent from the tropical, continuously growing *Astrosclera*.

The presence of siliceous spicules in *Astrosclera* finds no certain counterpart among stromatoporoids. N. D. Newell reported spicules in a Pennsylvanian stromatoporoid that he named *Parallelopora mira*, but C. W. Stearn doubts Newell's placement of this form. W. D. Hartman and T. F. Goreau have pointed out questionable pseudomorphs of spicules in Mesozoic stromatoporoids. Since northern Pacific populations of *Astrosclera* lack siliceous spicules, the thrust of this argument against a stromatoporoid affinity for this sponge loses some of its force.

Astrosclera may be regarded as a living descendant of the stromatoporoids, and this group may be classified as an order of the class Sclerospongiae. Stearn, however, prefers to consider the stromatoporoids as a subphylum of the phylum Porifera bearing no direct relationship with sclerosponges. Some authors, such as J. St. Jean, Jr., classify the late Paleozoic and Mesozoic stromatoporoids in a distinct order, the Sphaeractinoidea. The high probability of the occurrence of a similar fibrous aragonitic skeleton in "true" stromatoporoids as well as in "sphaeractinoid" forms largely obviates the necessity for a separation of the time series into different orders, however.

Ceratoporella. This sponge, first dredged off Cuba in 1878, was rediscovered by Goreau on Jamaican coral reefs in 1965. *Ceratoporella* secretes a massive skeleton of aragonite that is covered in life by a thin veneer of living tissue. The surface of the calcareous skeleton is marked by closely spaced calicles (**Fig. 3**) that share common walls and vary from 0.008 to 0.02 in. (0.2 to 0.5 mm) in lumen diameter. Each calicle continues to the base of the skeletal mass but is filled in secondarily by aragonite as the sponge grows upward so that only the distal 0.04 to 0.05 in. (1.0 to 1.2 mm) is filled with living tissue. The surface of the skeleton is marked by stellate impressions that indicate the position of the radiating exhalant canal systems. In addition to possessing a calcareous skeleton, *Ceratoporella* secretes siliceous spicules in the form of acanthostyles. These become embedded in the calcareous skeletal mass as it grows upward, and there they begin to erode away.

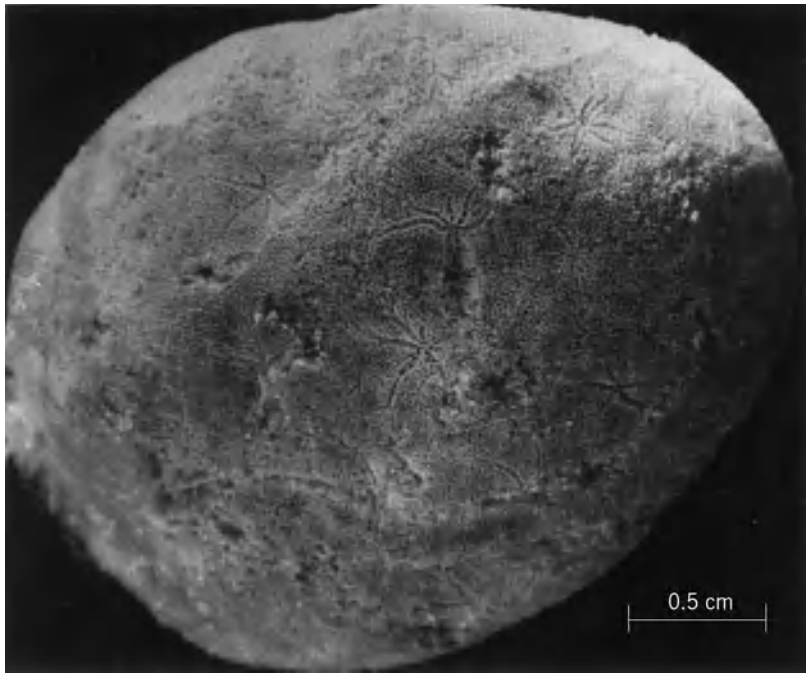


Fig. 1. Skeleton of *Astrosclera*, with astrorhizal patterns on surface.

Three genera related to *Ceratoporella* also occur on Caribbean reefs. In these the surface of the basal calcareous skeleton is marked by branched or lamellate upright processes, and the living tissues grow down into the irregularly shaped depressions between them. A species of one of these genera, *Stromatospongia*, is found on reefs of western Pacific islands.

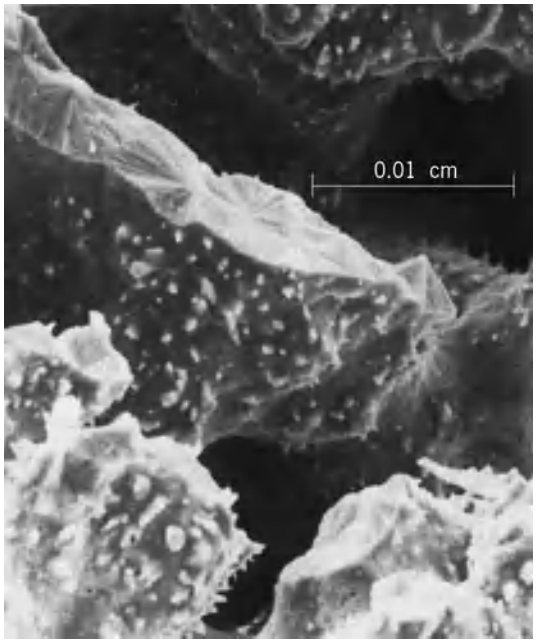


Fig. 2. Skeletal reticulum of *Astrosclera* showing spherulitic arrangement of crystalline units.

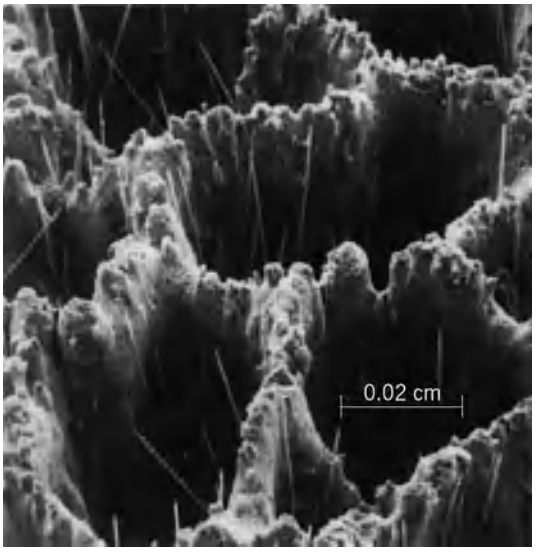


Fig. 3. Surface calicles of *Ceratoporella* showing protruding siliceous spicules. (*From W. D. Hartman and T. F. Goreau, Trans. Conn. Acad. Arts Sci., 44:131–148, 1972*)

Chaetetid affinities. The calcareous skeleton of *Ceratoporella* is remarkably similar to that of the group of fossil coralline animals known as chaetetids, ranging from Ordovician to Cretaceous times. Chaetetids, like *Ceratoporella*, have a calcareous skeleton composed of contiguous, aseptate, cylindrical tubes with shared walls between adjacent units. Another striking resemblance between the two groups is the mode of asexual reproduction by longitudinal fission, accomplished by the ingrowth of longitudinal "pseudosepta" to form a pair of daughter calicles. Further, the microstructure of the walls of the calicles in both is trabecular in nature, and the internal diameter of the calicles of *Ceratoporella* lies within the range of those found among chaetetids. In both groups the regular arrangement of contiguous longitudinal calicles breaks down in some species to a meandroid or irregular pattern.

Differences between *Ceratoporella* and its relatives on the one hand and chaetetids on the other include the incorporation of siliceous spicules into the calcareous skeleton of the former. No comparable occurrence is known among the chaetetids. Moreover, in *Ceratoporella*, the base of the skeleton is separated from the peripheral growing part by a continuous secretion of calcium carbonate rather than a periodic laying down of tabulae, as is characteristic of chaetetids. Since the characters shared by the two groups outweigh the differences, Hartman and Goreau have proposed placing *Ceratoporella* and its relatives in an order Ceratoporellida along with the order Chaetetida as related groups in the class Sclerospongiae. Fossil species of ceratoporellids are known from Triassic strata.

It seems likely that the rapid expansion of the scleractinian corals on Mesozoic and Cenozoic reefs brought the stromatoporoids and chaetetids into eclipse, with the surviving members relegated to cryptic reef habitats except at depths below those inhabited by hermatypic corals.

Acanthochaetetes wellsi. Unlike other sclerosponges described earlier, this Pacific sponge has a basal skeleton of calcite composed of contiguous vertical tabulate calicles ornamented within by vertical rows or irregular clumps of spines (**Fig. 4**). Living tissue extends to depths of 0.05–0.08 in. (1.2–2.0 mm) in the calicles and to the level of the outermost tabula in each. Ostia, located in the floor of shallow, curved grooves, open into the units of sponge tissue enclosed in the calicles; oscules occur at the center of slightly raised mounds and are converged upon by stellate configurations of exhalant channels that are impressed in the calcitic skeleton below as astrorhizae. Siliceous spicules of two types, tylostyles and modified spirasters, occur in the living tissue but are seldom incorporated into the calcitic skeleton.

Extinct species of *Acanthochaetetes* occur in Jurassic and Cretaceous strata. Hartman and Goreau have placed the family Acanthochaetetidae in a new order of sclerosponges, the Tabulospongida.

Favositid affinities. Certain characteristics of the calcareous skeleton of *Acanthochaetetes*, such as the presence of spines on the walls of the tabulate calicles and the method of asexual reproduction by means of budding walls between calicles, are strikingly similar to characteristics of the Favositida, an

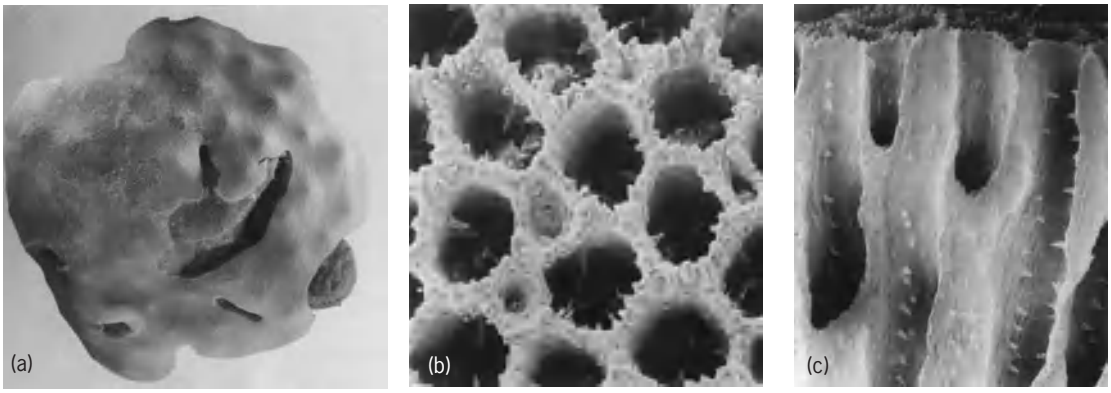


Fig. 4. *Acanthochaetetes wellsi*. (*a*) Skeleton with astrorhizal patterns on surface. (*b*) Calicles viewed from above; two developing asexual buds are seen in lower left. (*c*) Longitudinal fracture of calcitic skeleton showing rows of spines in calicles.

order of Paleozoic tabulate corals. However, the calicles of species of the favositids tend to bear vertical rows of pores that permit physical connections between the living tissue of adjacent calicles, while only submicroscopic pores, one-hundredth the size of those of favositids, occur in *Acanthochaetetes*. Further, whether the calcium carbonate of the calicle walls of favositids was aragonitic or calcitic is still disputed by paleobiologists, and astrorhizal patterns on the surface of favositid skeletons are unknown. Thus, despite some remarkable similarities between the Favositida and the Tabulospongida, the postulate that there was a phylogenetic relationship between the two groups, and its corollary that the Favositida were sponges instead of cnidarians, remains unresolved. *See* SPHAERACTINOIDEA; STROMATOPOROIDEA; TABULATA.

Willard D. Hartman

Bibliography. W. D. Hartman and T. F. Goreau, *Postilla*, 167:1–21, 1975; W. D. Hartman and T. F. Goreau, *Symp. Zool. Soc. London*, 25:205–243, 1970; W. D. Hartman and T. F. Goreau, *Trans. Conn. Acad. Arts Sci.*, 44:131–148, 1972; C. W. Stearn, *Lethaia*, 5:269–288, 1972.

Scorpaeniformes

Fishes of the order Scorpaeniformes, called the mailcheeked fishes, are distinguished by a unique character, the suborbital stay. The stay is an extension of the hypertrophied third infraorbital bone, which crosses the cheek obliquely and usually is firmly attached to the preoperculum. The head is usually spiny, and the body may be spiny or have bony plates, or both, or be smooth and appear naked. Almost all scorpaenids have prominent pectoral fins, either broad and fanlike or winglike, with the lower rays deeply incised. The order consists of about 1271 species and 266 genera in 25 families.

Order Scorpaeniformes
- Suborder Dactylopteroidei (flying gurnards)
- Scorpaenoidei (scorpionfishes)
- Platycephaloidei (searobins and flatheads)
- Anoplopomatoidei (sablefishes)
- Hexagrammoidei (greenlings)
- Normanichthyiodei (no common name)
- Cottoidei
  - Superfamily Cottoidea (sculpins and poachers)
  - Cyclopteroidea (lumpfishes and snailfishes)

Dactylopteroidei. This suborder consists of one family and seven species. The flying gurnards are tropical fishes of the Indo-Pacific and Atlantic oceans. They resemble searobins in having large, colorful pectoral fins with the inner rays free. "Flying gurnards" is a misnomer; rarely, if ever, do these fishes leave the ocean floor where they "walk" by alternately moving their pelvic fins. Large pectoral fins notwithstanding, there is no evidence that flying gurnards ever break the surface of the water and glide. However, *gurnard*, meaning "grunting," is appropriate, because these fishes produce stridulations by using the hyomandibular bone.

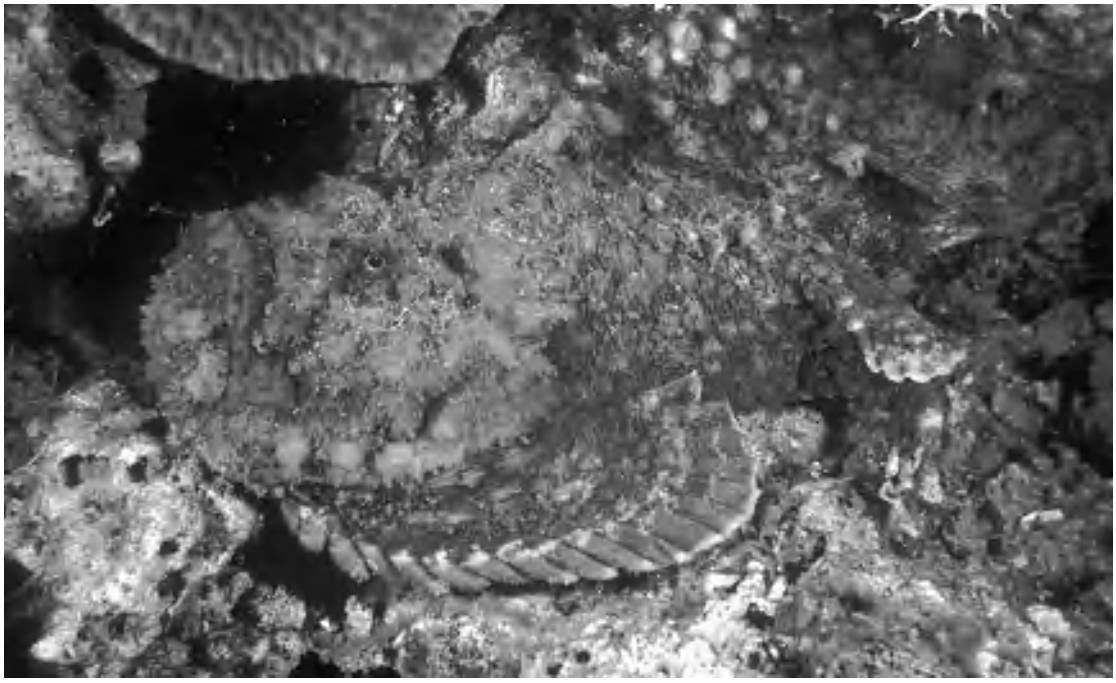
Scorpaenoidei. Thirty-five percent of all scorpaeniform fishes are in the suborder Scorpaenoidei. Of its six families, the principal one is Scorpaenidae (scorpionfishes), which accounts for 80% of the 444 species in the suborder. These fishes are represented in all of the tropical and temperate seas of the world and are the most generalized scorpaeniforms.

Scorpaenidae. The family Scorpaenidae contains the world's most venomous fishes. Scorpaenids are identified by a compressed body; a head usually with conspicuous ridges and spines; an opercle and preopercle armed with spines; scales that, when present, are usually ctenoid; a usually single dorsal fin; well-developed pectoral fins, fanlike and rounded, rarely one to three free rays; venomous glands at the base of the dorsal, anal, and pelvic spines; internal fertilization, with some species oviparous and others ovoviviparous. Most scorpionfishes are bottom dwellers that sit motionless among coral, rock, or grass, and, depending on their color patterns for concealment, wait for a venturesome prey to come within striking distance.

Among the venomous fishes is the beautiful red lionfish (*Pterois volitans*). Its sting causes a number of debilitating symptoms in humans which could lead to death if not properly and promptly treated. The lionfish is an Indo-Pacific native but now is known in the western North Atlantic from Florida to New York where it has proliferated, undoubtedly from released aquarium specimens. A popular subject for underwater photographers, the lionfish should be approached with caution, as it is quite aggressive when threatened. However, far more dangerous are the stonefishes (for example, *Synanceia*). They lie quietly while partially buried in the substrate and are so well camouflaged that they appear to be a rock or piece of coral (see **illustration**). Their spines are like hypodermic needles, ready to deliver the most venomous neurotoxin (ichthyocanthotoxin = toxin from fish spines) known in the fish kingdom. Punctures by these fishes bring on agonizing pain and could lead to a quick death.

Platycephaloidei. These fishes have an elongate body and large head, usually with spines and ridges. The head varies from a steep anterior to a strongly depressed profile. The dorsal fin is in two separate parts, the first with 6 to 12 stiff spines well separated from the soft rays. The lower two or three pectoral rays are enlarged and free in many species. All species are benthic, with some inshore and others occupying depths to 1500 m [5000 ft] (*Hoplichthys*, for example). Species in the families Triglidae and Peristediidae occupy all tropical and temperate seas; the other species are essentially limited to the Indo-Pacific.

Triglidae. The triglids (searobins) are characterized by a head typically encased in bony plates and armed with numerous stout spines, and with no barbel on the lower jaw. The body is scaly, with paired rostral



Stonefish (*Synanceia verrucosa*). (*Photo © John E. Randall*)

projections giving the snout a bilobed appearance in dorsal view. The lower three pectoral rays are enlarged, flexible, and free from each other and the rest of the fin, thus appearing like little legs. The free rays function as tactile organs as the triglids creep over the bottom in search of food, which consists mostly of crustaceans and mollusks. Triglids are good sound producers, both in and out of the water. The sound is produced by certain muscles vibrating the air bladder. There are ten genera and 70 species in this family.

Peristediidae. Armored searobins and triglids share some trenchant characters but differ in the former having the entire body encased in spine-bearing plates, two free pectoral rays, large barbels on the lower jaw, and usually conspicuously long and narrow rostral projections. There are four genera and 30 species in this family, all found in deep water.

Anoplopomatoidei. The sablefishes have little affinity with the other scorpaeniforms. The only uniting character, the suborbital stay, may have evolved independently. The head is without spines, ridges, or cirri, and the spinous dorsal fin is well separated from the soft dorsal. There is one family with two genera and two species ranging from Japan, across the Bering Sea, to California.

Hexagrammoidei. Greenlings have cirri on the head but lack ridges and spines. The dorsal fin is single but with a notch between the spinous portion and the soft rays. There are five genera and 11 species, all endemic to the North Pacific.

Normanichthyiodei. The head lacks spines and the body is covered with ctenoid scales. Previously in the suborder Cottoidei, its relationships remain very uncertain. There is only one species in this suborder, which is known to occur off Peru and Chile.

Cottoidei. The cottoidei are the largest group of scorpaeniform fishes, accounting for 49% of the species. The physiognomy of cottoids is quite variable, with overall shapes from elongate to globose. The head is usually large and lacking spines; the body either is naked (or appearing naked) or covered by scales or bony plates. The dorsal fin is either single or double; the pelvic fin has one spine and two to five rays, or may be completely absent. The pectoral fin is either continuous or the lower rays are free as in triglids. Cottoidei is recognized as two monophyletic lineages, the superfamilies Cottoidea and Cyclopteroidea.

Cottoidea. This superfamily consists of nine families. The five families described below account for 96% of the Cottoidea.

Cottidae (sculpins) are the largest family, with 70 genera and 300 species accounting for 75% of Cottoidea species. A lateral process of the hyomandibular bone is a unifying character of the family. Also, the skin appears naked, although the body may be covered with prickles and have a few ctenoid scales in the lateral line. The eyes are large and directed upward. There are two prominent preopercular spines, no spine in the anal fin, one spine and two or three soft rays in the pelvic fin, and large and fanlike pectoral fins. An air bladder is absent. Sculpins are essentially Holarctic in distribution. The marine species reach their greatest diversity in the coastal waters of the North Pacific. Freshwater species occur in North America and northern Eurasia, including a dozen or so species endemic to Lake Baikal. Four genera and 32 described species occur in North America, *Cottus* being the principal genus.

Agonidae (poachers) are mostly quite elongate and resemble armored searobins in having the body

covered with bony plates. Their antitropical distribution is unusual for scorpaeniforms, ranging in the southern hemisphere to the tip of South America and in the northern hemisphere in the Atlantic and Pacific oceans to near the Arctic Circle, with the greatest diversity in the North Pacific. Their vertical range is from shallow inshore waters to depths of over 1000 m (3,300 ft).

Psychrolutidae (fathead sculpins) are robust fishes with a large round head and a usually continuous dorsal fin. Some have well-ossified head bones with spines, while others have poorly ossified head bones, no spines, and a wide interorbital distance. These scorpaenids are found in both hemispheres of the Indo-Pacific and Atlantic oceans, at depths ranging from shallow inshore waters to depths of 2800 m (9,200 ft).

Comephoridae (Baikal oilfishes) and Abyssocottidae are freshwater families with elongate bodies, a large head, large terminal mouth, and a spinous dorsal fin separate from a soft dorsal fun. The two species of Comephoridae, which are endemic to Lake Baikal, lack body scales and prickles, have no pelvic fin, and possess very long pectoral fins. The Abyssocottidae, consisting of 20 species primarily from Lake Baikal, have pelvic fins and relatively short pectorals. The two families collectively contribute about half of the lake's known fish fauna.

Cyclopteroidea. In species of this superfamily the pelvic fins, when present, are thoracic in position and usually modified to form a suction-cup-like disk. The pectoral fins are broad and tucked under the operculum; the gill openings are rather restricted, and the lateral line is usually absent. Two families, Cyclopteridae and Liparidae, are the most bizarre scorpaeniforms.

Cyclopteridae (lumpfishes) possess a globose body and are covered with thick skin, which is usually tubercular but may be smooth or spinous. The dorsal spines are flexible and often hidden beneath skin, and the soft dorsal and anal fins are small and never confluent with the caudal fin. By means of the ventral sucking disc, these fishes can attach themselves to rock, which they may resemble. Seven genera and 29 species occupy the cold marine waters of the northern hemisphere.

Liparidae (snailfishes) are tadpole-shaped fishes with soft, smooth, jellylike skin. The spines of the dorsal fin are weak and flexible; the soft dorsal and anal fins are long and often confluent with the caudal fin. The pelvic fins, when present, form a sucking disk. The 19 genera and about 195 species are represented over the widest geographical range of all scorpaeniforms, from Arctic to Antarctic seas. Their vertical range is from tidal pools to depths of over 7000 m (23,000 ft). *See* ACTINOPTERYGII; OSTEICHTHYES; PERCIFORMES. Herbert Boschung

Bibliography. J. S. Nelson, *Fishes of the World*, 3rd ed, 1994; B. B. Washington, W. N. Eschmeyer, and K. M. Howe, Scorpaeniformes: Relationships, in H. G. Moser et al. (eds.), *Ontogeny and Systematics of Fishes*, pp. 438–447, Spec. Publ. no. 1, American Society of Ichthyology and Herpetology, 1984.

Scorpiones

An order of the Arachnida characterized by chelate pedipalps and chelicerae (pincerlike appendages), pectines (feathery chemo- and mechanoreceptors used to survey the texture of the ground surface and detect pheromones), and a narrow, flexible postabdomen bearing a venomous segment (telson) with a terminal sting (aculeus).

Morphology. The scorpion body is divided into a cephalothorax (prosoma), covered by an unsegmented carapace, and a segmented abdomen (opisthosoma) (**Figs. 1** and **2**). The opisthosoma is differentiated into an anterior preabdomen (mesosoma) and a postabdomen (metasoma), which, together with the telson, constitutes the "tail" or cauda. The cephalothorax bears the chelicerae, pedipalps, and four pairs of walking legs. The preabdomen contains seven segments, the postabdomen five. One pair of small, simple median eyes (ocelli) and, depending on the species, 2–5 (usually 3) pairs of anterolateral eyes are situated on the carapace. Troglobitic (cave-dwelling) scorpions lack all or some eyes (usually the median pair). Each of the seven preabdominal segments is covered by a tergite, the ventral segments by a sternite. The genital aperture, covered by the genital operculum, opens ventrally on the first segment. A small, quadrate basal piece articulates laterally with the pectines on the second segment. The lateral areas of sternites III–VI possess four pairs of oblique, oval slits, the openings (spiracles) of the internal respiratory organs (book lungs).

Classification. Scorpions are derived from amphibious ancestors that lived in the Silurian, more than 400 million years ago. Paleozoic scorpions closely resemble modern descendants in anatomical details, except that some were considerably larger. *Brontoscorpio anglicus* measured approximately 1 m (3.3 ft) in length. Scorpions were formerly considered the sister group of other arachnids because they closely resemble extinct marine eurypterids (the sister group of arachnids) but recent data suggest that scorpions are embedded in the arachnid lineage, and merely retain primitive features.

Approximately 1490 extant species of scorpions, placed in 170 genera, and between 13 and 20 families (depending on the authority), have been described worldwide. They vary in color from translucent through tan or brown to black, or combinations thereof, and in size from 10 mm to 21 cm (0.4–8.3 in.).

Distribution. Scorpions occur on all continents except Antarctica, but are most abundant and diverse in tropical and subtropical regions, especially in desert and semidesert habitats. They also occur in savannas and grasslands, in deciduous, coniferous, and tropical rainforests, on high mountain slopes (above 5500 m, 18,000 ft, elevation) in the Alps, Himalayas, and Andes; in some of the deepest caves (nearly 1000 m, 3300 ft, below the surface), and in the intertidal zone. Although most scorpions are terrestrial, some are arboreal. Favored habitats include burrows (up to 1 m, or 3.3 ft, deep, to escape hot diurnal

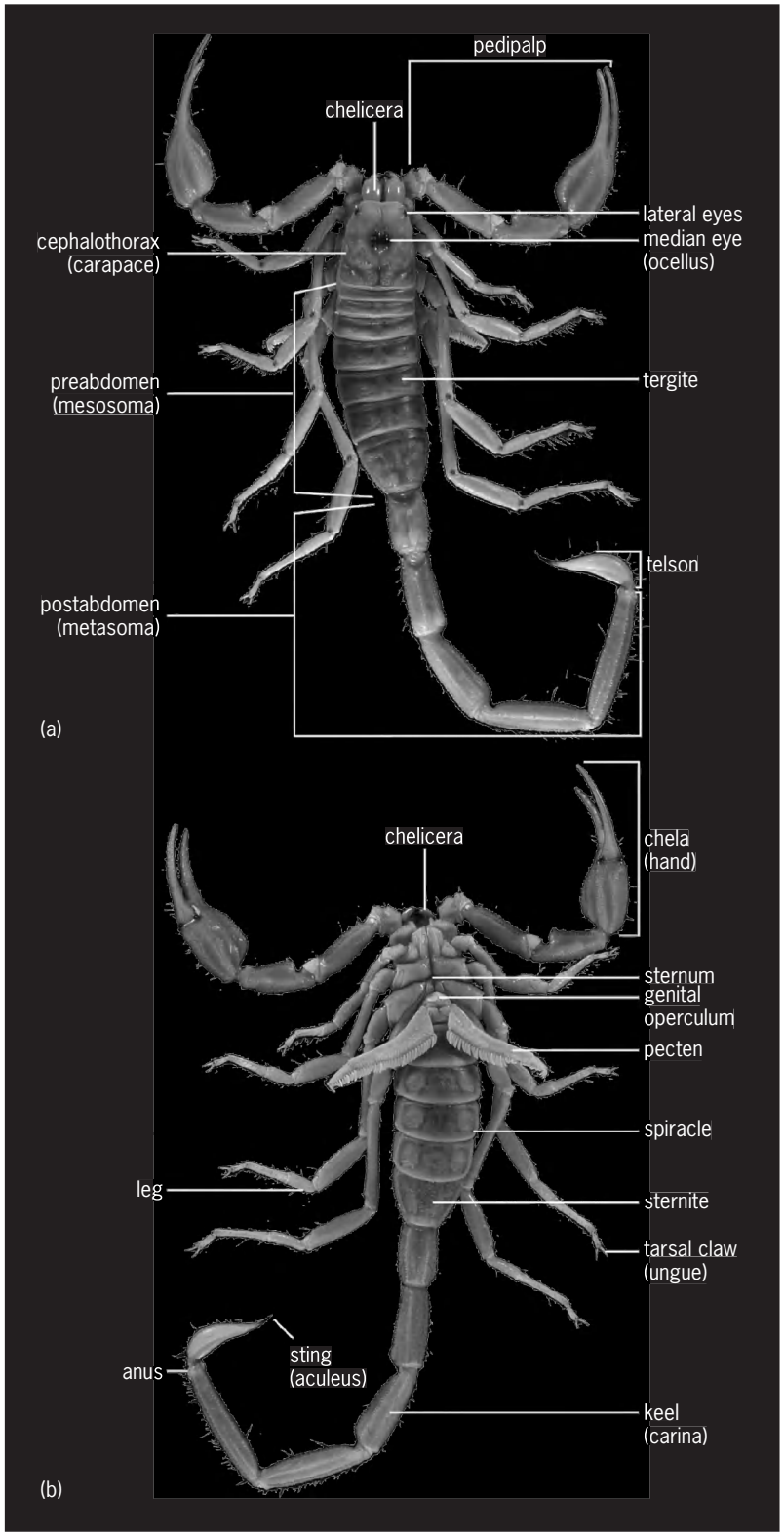


Fig. 1. Scorpion, *Smeringurus mesaensis* (Stahnke, 1957), adult male. (*a*) Dorsal view. (*b*) Ventral view. (*Photos by Randy Mercurio and Steve Thurston; copyright © American Museum of Natural History*)

temperatures), spaces under tree bark, logs, stones, and in rock crevices. Some species adapt well to human environments. Scorpions are often abundant in suitable habitats. Densities of 1 per square meter were reported for the Middle Eastern *Leiurus quinquestriatus* and of 8–12 per square meter for the intertidal *Serradigitus littoralis* from Baja California.

Ecology. Scorpions are primarily nocturnal. All fluoresce under long-wave ultraviolet light, facilitating their collection and observation at night. Their simple eyes detect luminosity, but little else. Prey are detected with slit sense organs in the tarsi, sensory setae (trichobothria) on the pedipalps, and the pectines, and are attacked with the chelate pedipalps or venomous sting. Scorpions with slender pedipalps are prone to sting their prey, those with robust pedipalps to crush prey mechanically, reserving the sting for large or strong prey. All use the pedipalps to manipulate prey, tearing pieces off with the chelicerae to be digested in a preoral cavity before being sucked into the gut.

Scorpions are important consumers in some communities. *Scorpio maurus* was reported to eat an annual average of 11% of the Israeli isopod population; *Urodacus yaschenkoi* consumes 7.9 kg/ha (7.8 lb/acre) of invertebrate prey in Australia. Cannibalism and predation by other scorpion species may be the most important sources of scorpion mortality, but other invertebrate predators (such as centipedes) and vertebrates are also important predators. Mortality is highest immediately after birth, lower for individuals of intermediate age, and high for adults (65%, 30%, and 60% per year, respectively, for the Australian *Urodacus manicatus*). Mortality is particularly high among males due to increased mobility during the breeding season and cannibalism by females. Skewed adult sex ratios of 1.2–1.4:1 are typical. Social behavior occurs rarely in species of *Heterometrus*, *Opisthacanthus*, and *Pandinus*, in which family groups with overlapping generations cooperate to construct and occupy communal burrows, inhabited by individuals of various ages.

Reproduction and development. Reproduction in scorpions is indirect. Intromission occurs via a spermatophore. The male attaches the spermatophore to the substrate while grasping the female by the pedipalps or chelicerae during a mating dance or "promenade-aux-deux." The species-specific spermatophore catapults the sperm mass into the female gonopore when a lever is touched (a lock-and-key mechanism). All scorpions are viviparous (livebearing). Embryos develop in the reproductive tract and receive nourishment from maternal tissues.

Fig. 2. Scorpion, *Hadrurus arizonensis* (Ewing, 1928), adult male. (*Photo by Randy Mercurio; copyright © American Museum of Natural History*)

Some are parthenogenetic. Scorpions have low reproductive rates in comparison to other terrestrial arthropods and are among the most long-lived. Gestation times are long (several months to more than a year) and litter sizes small (1–105). Young are fairly large at birth and helpless (altricial), clinging to the mother for the first few molts before dispersing. Time to sexual maturity varies from 2 to 8 years, depending on the species. Average longevity is 4–5 years, but larger species may live 25–30 years. Scorpions do not molt as adults.

Small litter sizes, long generation times, and low survivorship among sexually immature females contribute to a low rate of population increase for most scorpions. Many are habitat specific, range-restricted, and sensitive to environmental degradation, exacerbating their risk of extinction due to human activities, including habitat destruction and harvesting for the souvenir and exotic pet trades.

Toxicity. Scorpion venoms contain multiple low-molecular-weight proteinaceous neurotoxins that block sodium and potassium channels, preventing the transmission of nerve impulses across synapses. Scorpion envenomation represents a significant cause of morbidity and mortality in some regions (such as Mexico, North Africa, and the Middle East). Some 100,000 scorpion stings occur annually in Mexico, killing as many as 800 people (mostly young children and the elderly). The figures may be even higher in North Africa and the Middle East. However, most scorpions are harmless. The sting may be painful, but not dangerous. About 25 species, all in the family Buthidae, are considered medically important worldwide. Most of these occur in the New World genera *Centruroides* and *Tityus*, and the Old World genera *Androctonus*, *Buthus*, *Leiurus*, *Mesobuthus* and *Parabuthus*. *Centruroides exilicauda*, from Arizona, California, and New Mexico, is the only species known to be lethal in the United States. The venom of these scorpions has proved fatal to healthy children up to 16 years of age and to adults suffering from hypertension and general debility.

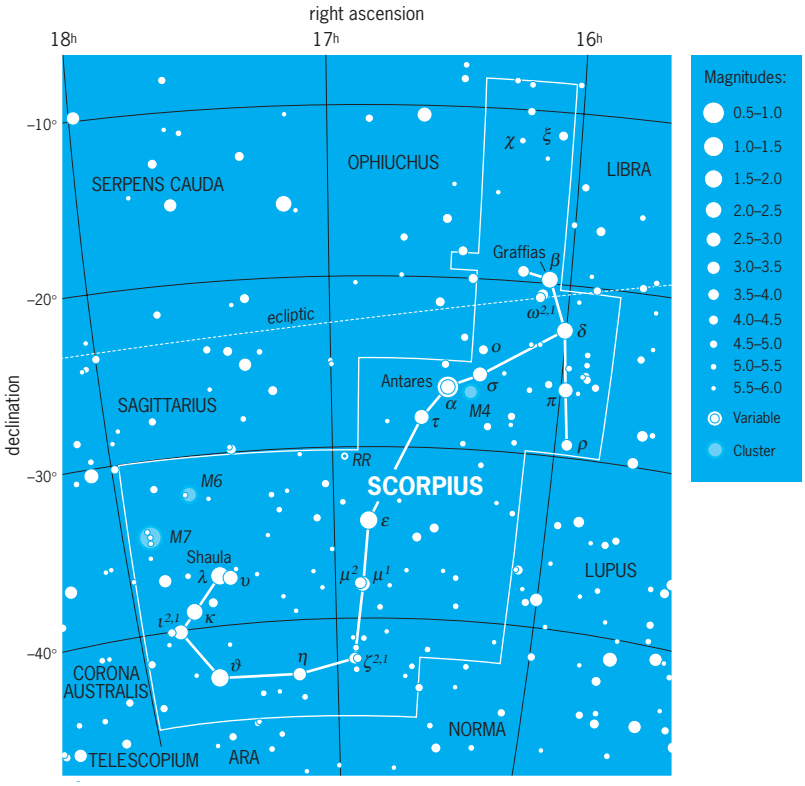
Morphology. Scorpions are prominent in mythology and folklore. A classical zodiacal constellation is named after scorpions. The toxicity and fearsome appearance of scorpions contribute to a fascination that has always and continues to surround them. *See* ARACHNIDA.

Lorenzo Prendini

Bibliography. P. Brownell and G. Polis, *Scorpion Biology and Research*, Oxford University Press, 2001; V. Fet et al., *Catalog of the Scorpions of the World (1758–1998)*, New York Entomological Society, 2000; V. Fet and P. A. Selden, *Scorpions 2001: In Memoriam Gary A. Polis*, British Arachnological Society, Burnham Beeches, Bucks, UK, 2001; G. A. Polis, *The Biology of Scorpions*, Stanford University Press, 1990.

Scorpius

The Scorpion, a large zodiacal constellation (see **illustration**). It probably represents the scorpion in Greek mythology sent by Apollo to sting and kill



Modern boundaries of the constellation Scorpius, the Scorpion. The celestial equator is 0° of declination, which corresponds to celestial latitude. Right ascension corresponds to celestial longitude, with each hour of right ascension representing 15° of arc. Apparent brightness of stars is shown with dot sizes to illustrate the magnitude scale, where the brightest stars in the sky are 0th magnitude or brighter and the faintest stars that can be seen with the unaided eye at a dark site are 6th magnitude. (*Wil Tirion*)

Orion, in order to protect Apollo's sister Artemis's virginity. Now Scorpius and Orion are separated by half the celestial sphere from each other, though Scorpius eternally pursues Orion. *See* ORION; ZODIAC.

The constellation loops through the sky like a giant fish-hook. The bright supergiant star Antares is the 15th brightest star in the sky. Its name means Rival of Mars, since both are bright and reddish in color, and Mars sometimes passes close to Antares. *See* ANTARES.

Scorpius lies along the Milky Way and contains many beautiful star clusters and star clouds. The globular cluster M4 and the open clusters M6 and M7 are especially noticeable. *See* MESSIER CATALOG; STAR CLOUDS; STAR CLUSTERS.

The modern boundaries of the 88 constellations, including this one, were defined by the International Astronomical Union in 1928. *See also* CONSTELLATION.

Jay M. Pasachoff

Scramjet

An air-breathing jet engine that relies for propulsion on the compressing or ramming effect on air taken into the engine inlet at supersonic speeds, normally when the aircraft is traveling at speeds above Mach 4. The term is derived from "supersonic combustion ramjet." Scramjet technology now in development is expected to make global rapid travel and

affordable orbit access a reality in coming years. A breakthrough was realized in the spring of 2003 when a flight-weight scramjet engine was run at 4.5 and 6.5 times the speed of sound (Mach 4.5 and 6.5) using conventional hydrocarbon fuel. More significantly, this represents the first flight-weight, hydrocarbon-fueled scramjet engine. This test was conducted by the Pratt & Whitney Space Propulsion Hydrocarbon Scramjet Engine Technology (HySET) team, in partnership with and sponsored by the U.S. Air Force Research Laboratories (AFRL) Propulsion Directorate. *See* MACH NUMBER.

Scramjet operation. A ramjet is a jet engine that relies on the compressing or ramming effect on the air taken into the inlet while the aircraft is in motion. The geometry of the engine compresses the air rather than a piston, as in an internal combustion engine, or a compressor, as in a gas turbine. In a ramjet, the compressed air is slowed to subsonic speeds (less than Mach 1) as it is routed into the combustion chamber, where it is mixed with fuel, burned, and expanded through a nozzle to generate thrust.

A scramjet normally operates at speeds above Mach 4. To get a missile or aircraft to Mach 4, a rocket or jet engine is typically used for the initial acceleration. As the vehicle is traveling through the atmosphere on its way to Mach 4, the ignition condition for the scramjet, air is being compressed at hypersonic (greater than Mach 5) and supersonic (Mach 1–5) speeds on the lower surface of the aircraft and enters the engine at high pressure (**Fig. 1**). From the scramjet inlet, the air flows at supersonic speed into the combustor, where fuel is mixed with the air, burned, and expanded to make thrust. The mixing, burning, and expansion process takes place in less than 0.001 second. Engineers who developed this technology have referred to scramjet combustion as attempting to "light a match in a hurricane."

Advantages. In their operating speed regime, scramjets offer several advantages over conventional jet and rocket engines. Although not as efficient as a gas turbine engine, a scramjet has the ability to operate at much faster speeds. While burning hydrocarbon fuel, a scramjet engine can operate at speeds up to Mach 8. Using hydrogen as the fuel, a scramjet engine can operate at speeds in excess of Mach 12. Because of its air-breathing nature, the performance (specific impulse) of a scramjet is significantly higher than that of the best-performing rocket engines. One big advantage of a scramjet engine is the minimal number of moving parts required for operation. Since it relies on the shape of its flow path to compress the air by dynamic means, a scramjet engine does not require the use of high-pressure pumps, compressors, or turbines as do gas turbine or rocket engines. While reducing moving parts increases engine reliability and durability, benefits such as high speed, relative fuel efficiency, and range are the reasons that scramjets will revolutionize the aerospace industry.

The development of scramjet propulsion technology will enable affordable and reusable space transportation systems with operation much like a commercial airliner. Scramjet-powered access-to-space systems that will take off and land horizontally like commercial airplanes are being conceptualized today. These systems will be designed to achieve orbit while employing relatively low acceleration profiles, potentially opening access to orbital flight to a much larger section of the population.

Scramjets will be used in conjunction with either gas turbine engines or rockets in combined cycle

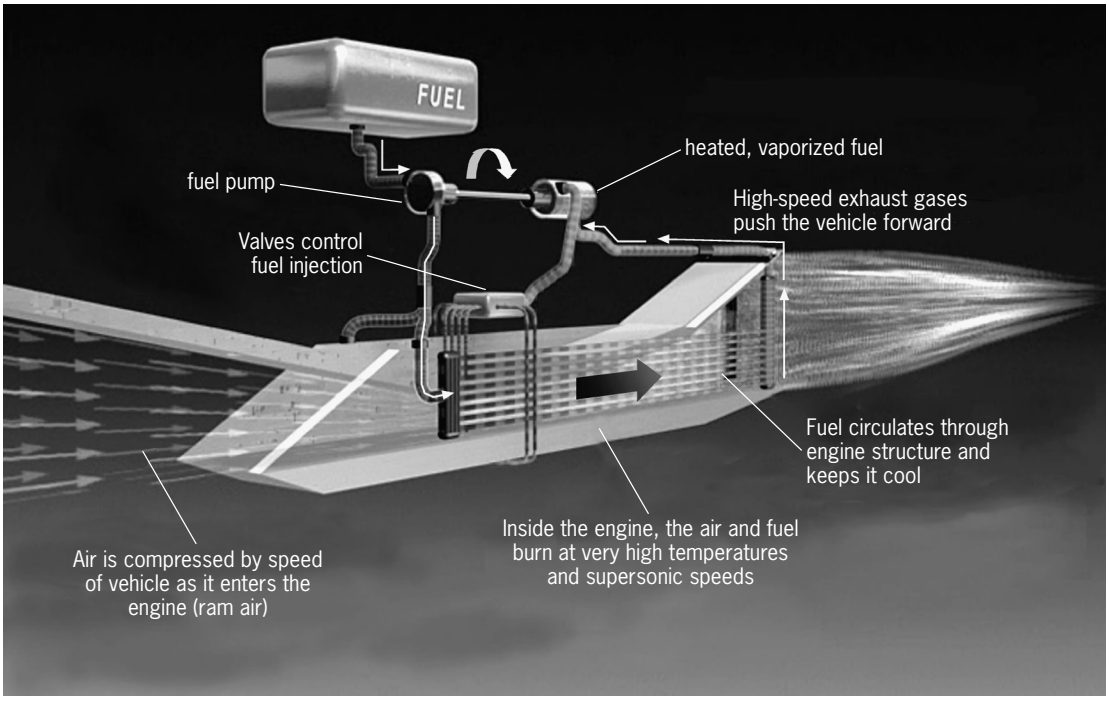


Fig. 1. Operation of a scramjet.

engines (CCEs). In a turbine-based combined cycle (TBCC) engine, a turbine engine will accelerate the aircraft from takeoff to a scramjet takeover speed (Mach 3–4), at which point the scramjet will be ignited and the turbine will be shut down. In a rocket-based combined cycle (RBCC), small rockets embedded in the flow path of the scramjet will first be used as ejectors to accelerate the vehicle to ramjet and then to scramjet operation. In both cases, the CCEs will be used to accelerate the space vehicle to escape velocity while flying through the atmosphere using the oxygen in the air. As the vehicle gains altitude and escapes the Earth's atmosphere, propulsion will transition to rockets since air will not be available to the scramjets. *See* ROCKET PROPULSION; TURBINE PROPULSION.

Future air-breathing access-to-space systems could increase the payload fraction by two-to-five times that of conventional rocket-powered systems. Since the scramjet engine takes its oxidizer from the air, a scramjet-powered vehicle does not require an oxidizer tank or a supply control system. This reduction in oxidizer volume and weight can be translated into smaller vehicles and reduced launch costs.

Initial technology developments for reusable applications will concentrate on developing variable geometry to optimize scramjet performance and extend the Mach number envelope. Since a hypersonic vehicle will spend prolonged periods traveling at high speed in the atmosphere, its surface will require significantly more thermal protection than the expendable launch systems in use today. Structural and thermal durability will be increased to allow multimission capability. Additionally, scientists and engineers are defining scalability factors that must be taken into consideration when designing the larger engines required for space applications.

Development. The U.S. Air Force established the Hypersonic Technology (HyTech) Program in 1995 after the National Aero-Space Plane's development was terminated. In 1996, Pratt & Whitney won a contract for HySET to demonstrate the performance and structural durability of a hydrocarbon-fueled scramjet engine. As a result of the HySET Program's success the X43-C, a joint USAF/NASA program, has emerged. The goal of this program is to flight-test a three-module derivative of the HySET engine simulating elements of a space mission. In parallel and as risk reduction for the X43-C Program, the Air Force is also considering flying a propulsion flight demonstration program: the Endothermically Fueled Scramjet Engine Flight Demonstrator (EFSEFD) Program. This flight demonstration will use a single HySET engine derivative module and will focus on the military application of the scramjet.

Building block approach. Once the hypersonic vehicle mission requirements were translated into scramjet engine component requirements, the HySET team generated the computational fluid dynamic codes. These codes define combustion products and optimize the fueling locations and concentrations required for top performance. Their results were used to design the hardware, allowing optimization of the engine design prior to test.

Fig. 2. Performance Test Engine (PTE).

From subscale testing, the team updated codes and defined the Performance Test Engine (PTE). This full-scale test of a heat sink engine validated the computer codes and demonstrated engine performance and operability. The PTE is a heavy-weight, heat-sink engine fabricated in solid copper (**Fig. 2**). The use of conventional jet fuel provided a major challenge to the team. Prior to this program, most developmental scramjets used hydrogen, which is much easier to mix and burn than jet fuel. In addition, the team developed a special fuel-conditioning system to simulate the fuel temperatures that would be generated by a fuel-cooled engine in flight. *See* AIRCRAFT FUEL.

The PTE was operated at Mach 4.5 and 6.5 consistent with scramjet takeover and cruise conditions for a missile application. The team tested and examined 95 test points, demonstrating excellent operability. The engine achieved 110% of its performance test objectives at Mach 4.5 conditions and 100% of its test objectives at Mach 6.5. Based on the performance and operability of the engine during test, the team went on to design and fabricate a more sophisticated full-scale, fuel-cooled engine. *See* AIRCRAFT FUEL.

As part of the building block approach, the team built on lessons learned, and used test results from the PTE to calibrate and improve analytical tools leading to the evolution of the Ground Demonstration Engine (GDE). With every test result, the analytical tools used for the design were calibrated, resulting not only in the ability to understand the behavior of the hardware at test but also in the definition of a design set of tools to be employed in subsequent iterations. Along with the development of the scramjet's operability and performance, the HySET team used the building block approach for developing the thermal and structural technologies. Multiple thermostructural components with increasingly larger sizes were designed, built, and tested in

scramjet operating conditions. In 2001, this culminated in the creation of the first hydrocarbon-fueled scramjet engine, GDE-1, weighing less than 150 lb (59 kg) and using readily available nickel-based alloys for its fabrication.

Operation. In the GDE-1 engine, the fuel must enter the combustor as a vapor in order to facilitate mixing and burning with the airflow. This is achieved through regenerative cooling of the hardware using a revolutionary endothermic cooling approach. The liquid fuel enters the front of the engine and is routed through the structure to the rear of the engine. The fuel cools the metallic structure as it passes through very small longitudinal passages. The passages in the structure are coated with a catalyst. While the fuel is absorbing the heat from the structure, the fuel temperature and pressure rise, causing the catalyst to "crack" the fuel in an endothermic reaction; that is, the fuel breaks down into gaseous lighter hydrocarbon components such as hydrogen, ethylene, and methane. This process gasifies the fuel for easier burning while allowing the fuel to absorb a significantly larger amount of heat from the scramjet structure than otherwise would be possible. The process is akin to a flying fuel refinery. At the rear of the engine structure, the cracked fuel is collected and metered into the scramjet combustor and burned. The combustion process expands through a nozzle, where it produces positive net thrust.

Testing. In the spring of 2003, the GDE-1 was tested initially at low speed (Mach 4.5) and then at Mach 6.5 (**Fig. 3**). As in the prior component tests, the results were used to calibrate and validate the design tools. By July 2003, GDE-1 had completed about 60 tests, successfully validating the thermostructural, performance, and operability predictions. The results of GDE-1 testing have proven that with today's technology scramjets fueled by hydrocarbon fuels can generate the thrust required to support hypersonic flight.

The next scramjet engine, GDE-2, was also fuel-cooled and flight-weight as was GDE-1. However, this was designed as a full system test to introduce control hardware and software so that the engine could run as a complete closed-loop system. Fabrication of the GDE-2 was completed in the spring of 2005.

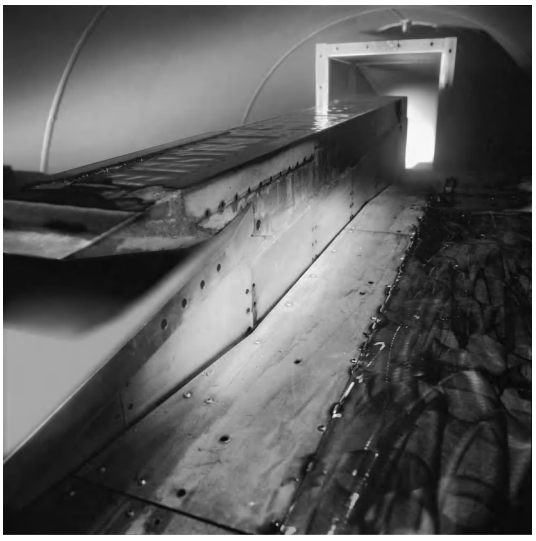


Fig. 3. GDE-1 test at Mach 4.5.

Outlook. Building on the lessons learned during PTE, GDE-1, and GDE-2 ground testing, the HySET team's goal is flight demonstration of this technology. The X43-C and EFSEFD programs will provide the first opportunities to demonstrate the operation of a hydrocarbon-fueled scramjet engine in flight. The lessons learned during these flight experiments will pave the way for hypersonic flight. *See* HYPERSONIC FLIGHT.

Joaquin Castro

Bibliography. E. T. Curran and S. N. B. Murthy (eds.), *Developments in High-speed Vehicle Propulsion Systems*, vol. 165, AIAA, 1996; E. T. Curran and S. N. B. Murthy (eds.), *High-speed Flight Propulsion Systems*, vol. 137, AIAA, 2000; E. T. Curran and S. N. B. Murthy (eds.), *Scramjet Propulsion*, vol. 189, AIAA, 2000.

Scrapie

A transmissible, invariably fatal disease of adult sheep and goats characterized by degeneration of the central nervous system. It is one of the transmissible spongiform encephalopathies, a disease family which includes bovine spongiform encephalopathy (BSE) in cattle (commonly known as mad cow disease), chronic wasting disease in deer, and Kuru and variant Creutzfeldt-Jakob disease in humans. Scrapie was first discovered in European sheep in the early eighteenth century and is now present in many sheep populations worldwide, excluding Australia and New Zealand. *See* NERVOUS SYSTEM DISORDERS.

Clinical signs. Scrapie has a long incubation period. It is widely believed that most animals are infected at or shortly after birth, with about 90% of sheep developing signs of disease at 2–5 years of age. Scrapie affects both sexes and is insidious and progressive in its onset, with a clinical course that lasts weeks or months. The clinical signs can be highly variable between flocks, depending on the strain of agent and breed of sheep, but often begin with hyperexcitability or dullness, and proceed to pruritus (itching), ataxia (incoordination), trembling, and loss of condition. Pruritus is accompanied by scratching or scraping (hence the name), leading to wool loss (**Fig. 1**). There is no treatment for scrapie, and affected animals will, if not euthanized, become recumbent and die.

Etiology. Scrapie is caused by a transmissible agent, which can be found in cell-free filtrates of the brain and other organs from scrapie-affected sheep. Importantly, inocula remain infectious in the absence of detectable nucleic acid, but lose infectivity after treatment with methods that denature protein. The prion (proteinaceous infectious particle) hypothesis, for which Stanley Prusiner was awarded the 1997 Nobel Prize in Physiology or Medicine, proposes that the etiologic agent of scrapie and other transmissible spongiform encephalopathies is

the host protein, PrP, which can exist in a nonpathological form, PrP^{C} or PrP_{Sen}, and a pathological form, PrP^{Sc} or PrP_{Res}, with the two forms differing in conformation only. According to the prion hypothesis, PrP^{Sc} converts PrP^{C} into more PrP^{Sc}, and this cascade reaction leads, ultimately, to disease pathology. However, PrP^{Sc} could simply be a marker for infection, rather than the infectious agent itself, as transmissible spongiform encephalopathy infections have been reported in the absence of detectable PrP^{Sc}. *See* PRION DISEASE.

Pathology. Infection is believed to occur mostly by the oral route, with entry to the body through the ileal Peyer's patch (a collection of lymphoid follicles in the ileum of the small intestine), and dissemination through the lymphatic/vascular system to tissues of the lymphoreticular system, including tonsil, spleen, retropharyngeal, and mesenteric lymph nodes. With modern immunohistochemical methods, the presence of PrP^{Sc} can be detected in the lymphoreticular system just months after infection, primarily within the follicular dendritic cells of primary and secondary B cell follicles. Infection spreads from the lymphoreticular system through the autonomic nervous system to the central nervous system, where PrP^{Sc} accumulates primarily in glial cells (astrocytes) and neurons. Within the brain there is astrocytic gliosis (production of astrocytes in area of neurodegeneration), the accumulation of deposits of PrP^{Sc}, vacuolation, and the characteristic spongiform degeneration (**Fig. 2**). *See* CELLULAR IMMUNOLOGY; LYMPHATIC SYSTEM.

It is possible that some infections are transmitted via nonoral routes. For example, scrapie can be efficiently transmitted by scarification (skin or tissue incision) and via the conjunctiva (mucous membrane covering the eyeball and lining the eyelids).

Transmission. There is very little evidence for vertical transmission (direct transmission from infected mother to offspring in utero) of any transmissible spongiform encephalopathy, including scrapie. By contrast, there is substantial evidence for horizontal (direct lateral; among related or unrelated individuals) transmission of scrapie between sheep housed together, although the precise mechanism is not clear. However, it is widely believed that many scrapie infections are acquired by lambs at or shortly after birth. Infectivity and PrP^{Sc} can be readily detected in the placenta and fetal membranes of infected ewes, and these tissues are believed to play a central role in transmission. One model for horizontal transmission is that an infected ewe gives birth to lambs that are themselves uninfected but may be accompanied by an infected placenta and other birth tissues/fluids. The newly born lambs and others that are in close association (within the same lambing pen, perhaps), if genetically susceptible to scrapie, then become infected by ingestion of infected material.

There are reports in the scientific literature of sheep, sourced originally from scrapie-free areas, acquiring scrapie after being kept on farmland that previously held infected stock. For example, in Iceland

Fig. 1. Poll Dorset sheep with severe clinical signs of scrapie.

in the 1940s all sheep within the scrapie-affected north were slaughtered to help eradicate another sheep disease, maedi visna. Restocking occurred 1 to 3 years later with lambs sourced from the scrapie-free south. Reportedly, scrapie reappeared in farms that were previously scrapie-affected, but not in farms that were previously disease-free. These observations have led to the concept of the environmental persistence of the scrapie agent and the possibility of indirect lateral transmission between sheep mediated by a contaminated environment. How sheep actually acquire the infection from the environment is unknown.

The route of scrapie transmission between goats is also not known, but is probably similar to that of sheep.

Genetics. Whether a sheep can acquire scrapie infection or, following infection, proceed to clinical disease is under the control of the gene that encodes its PrP protein, called the *PrP* gene. Five polymorphic variants (alleles or haplotypes) of the *PrP* gene that exhibit this linkage to scrapie are commonly found in sheep. The variants encode the following amino acid sequences in PrP: alanine (A) at codon 136, arginine (R) at codon 154, and arginine (R) at codon 171 ($A_{136}R_{154}R_{171}$, or ARR for short), and, continuing this notation, ARQ, AHQ, ARH, and VRQ (in which Q is glutamine, H is histidine, and V is valine).

Each sheep inherits one haplotype from each of its parents, giving it 1 of 15 possible *PrP* genotypes often written as, for example, ARQ/VRQ or ARR/AHQ. The relationship between the haplotypes/genotypes and susceptibility to scrapie is complex. Briefly, ARR and AHQ confer resistance to classical scrapie, while ARQ and VRQ confer susceptibility. The ARR/ARR genotype of sheep is the most resistant to scrapie, while the VRQ/VRQ genotype is the most susceptible. The effect of the ARH haplotype

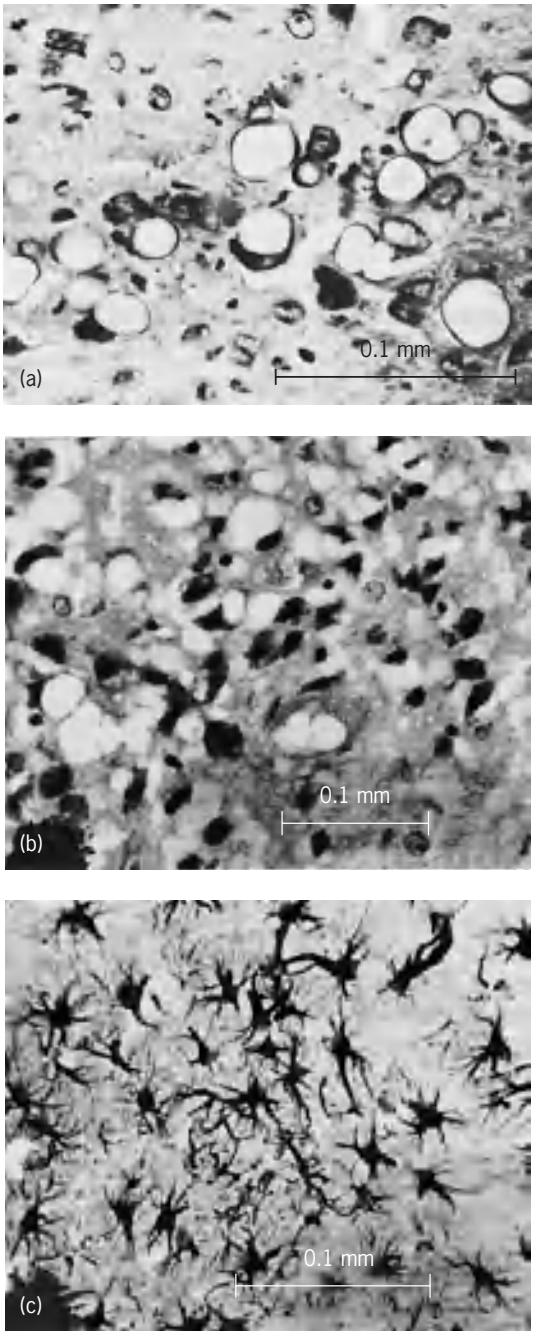


Fig. 2. Photomicrographs of the brain from scrapie-affected animals showing (*a*) vacuoles in neurons of the medulla, (*b*) spongy degeneration, (*c*) astrocytic hypertrophy (astrocytic gliosis in mice)

varies, depending on the other haplotype present. Small changes in genotype can have a remarkably large effect on susceptibility to scrapie. For example, British sheep of the ARQ/VRQ genotype are 300 times more likely to be reported with scrapie than sheep of the AHQ/VRQ genotype, despite differing only in arginine or histidine being encoded by codon 154 of one of the two haplotypes. The molecular mechanism that links *PrP* genotype to resistance/susceptibility is not understood.

Several strains of scrapie exist, and some exhibit patterns different from those described above. In many countries, including the United States, the VRQ haplotype is absent in the major sheep breeds, and sheep of the ARQ/ARQ genotype are most susceptible to the local scrapie strains. Sheep of the AHQ/AHQ genotype are the most susceptible to a scrapie strain discovered in Norway in 1998 (Nor98), while VRQ/VRQ sheep appear to be resistant to it. Novel diagnostic tests for scrapie have detected sheep with "atypical" scrapie infections (so-called because of an unusual distribution of PrP^{Sc} in the brain), including some of the ARR/ARR genotype which have hitherto been considered entirely or almost entirely resistant to scrapie. Under experimental conditions bovine spongiform encephalopathy can be transmitted to sheep and behaves as a novel scrapie strain; sheep of the ARQ/ARQ genotype are most susceptible, and sheep of the ARR/ARR genotype can succumb to infection and clinical disease if inoculated intracerebrally. *See* ALLELE; GENE; GENE ACTION; GENETIC CODE.

Control. Many approaches to scrapie control have been tried in the past, such as cleaning and disinfection of premises and culling of affected animals, entire populations of age-cohorts, genetically related animals, and even whole-flock slaughter. Such methods have generally not brought long-term success. The recent discovery of the strong association between scrapie and *PrP* genotype has, however, raised the possibility of controlling the disease at a national level by selective breeding. Genetic control programs are now under way in many countries, including the United States and the entire European Union. These programs are based on large-scale genetic testing, followed by culling of the most genetically susceptible sheep, and promoting the use of ARR/ARR sheep for breeding. The extreme resistance of sheep of the ARR/ARR genotype to most strains of transmissible spongiform encephalopathy suggests a high probability of success for these programs, although the recent discovery of "atypical" scrapie infections in ARR/ARR sheep is an important concern. *See* INFECTIOUS DISEASE; VETERINARY MEDICINE.

M. Baylis

Bibliography. H. B. Parry, *Scrapie Disease in Sheep: Historical, Clinical, Epidemiological, Pathological and Practical Aspects of the Natural Disease*, Academic Press, Oxford, 1983; S. B. Prusiner, *Prion Biology and Diseases*, 2d ed., Cold Spring Harbor Laboratory Press, 2004; R. M. Ridley and H. F. Baker, *Fatal Protein. The Story of CJD, BSE, and Other Prion Diseases*, Oxford University Press, 1998.

Screening

A mechanical method of separating a mixture of solid particles into fractions by size. The mixture to be separated, called the feed, is passed over a screen surface containing openings of definite size. Particles smaller than the openings fall through the screen and are collected as undersize. Particles larger than the openings slide off the screen and are caught as oversize. A single screen separates the feed into only two fractions. Two or more screens may be operated in series to give additional fractions. Screening

occasionally is done wet, but most commonly it is done dry.

Industrial screens may be constructed of metal bars, perforated or slotted metal plates, woven wire cloth, or bolting cloth. The openings are usually square but may be circular or rectangular. In rectangular openings the separation is controlled by the smaller dimension.

When the opening in a screen is larger than 1 in. (2.5 cm), the actual opening size in inches is specified. When the opening is 1 in. or less, the screen size is designated as the number of openings per linear inch; that is, a 20-mesh screen has 20 openings per inch. The actual size of an opening is less than that corresponding to the mesh number by the thickness of the metal between openings. Mesh sizes range from about 4 in. (10 cm) to 400-mesh, but screens finer than 100- or 150-mesh are seldom used industrially. Particles smaller than this are usually separated by other methods. *See* MECHANICAL CLASSIFICATION; SEDIMENTATION (INDUSTRY).

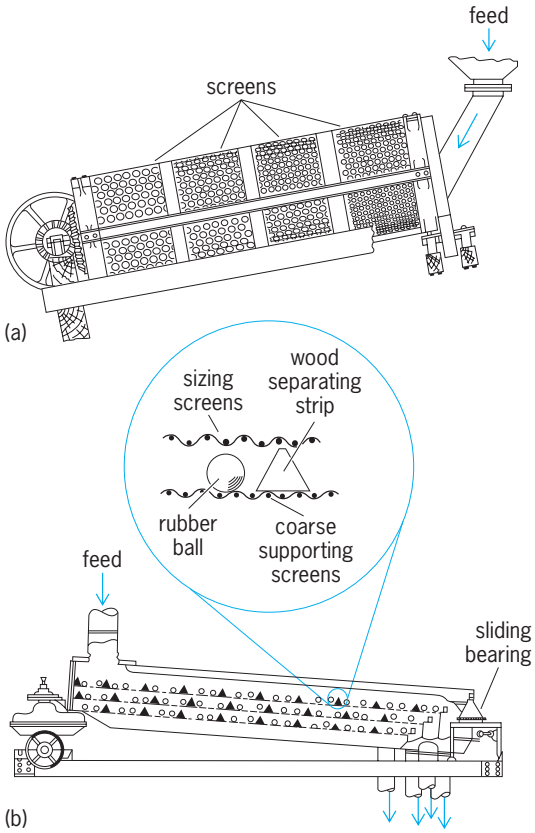
Testing sieves are used to measure the size distribution of mixtures of particles and to specify the performance of commercial screens and other equipment. They are made of woven wire screening. Mesh and wire size of each screen are standardized. The usual range is from 1- to 200-mesh. In use, a set of standard screens is arranged serially in a stack, with the finest mesh on the bottom and the coarsest on top. An analysis is conducted by placing a weighed sample on the top screen and shaking the stack mechanically for a definite time. Particles passing the bottom screen are caught in a pan. The particles remaining on each screen are removed and weighed, and the weights of the individual increments are converted to percentages of the total sample.

In most screens the particles are pulled through the openings by gravity only. Large particles are heavy enough to pass through the openings easily, but intermediate and smaller particles are too light to pass through the screen unaided. Most screens are agitated to speed the particles through the openings. Common methods are to revolve a cylindrical screen about an axis slightly inclined to the horizontal or to shake, gyrate, or vibrate a flat screen.

Two effects tend to restrict the fall of small particles through a screen: blinding of the screen by wedging of particles into the openings and sticking of individual particles to each other and to the screen. The motion of the screen reduces blinding, and sometimes positive means are provided to drive wedged particles back through the openings. Sticking is severe if the particles are damp, and may be reduced by drying the feed.

Two common screens are shown in the **illustration**. The trommel, a revolving screen, is a combination of four screens in series. This unit gives five products. The gyrating screen uses bouncing balls under the sizing screens to dislodge particles caught in the meshes. The balls are supported on coarse screens which offer no resistance to the materials passing the sizing screens. This screen gives four products.

Ideally, a screen should effect a sharp separation between undersize and oversize, and the largest particle in the undersize should be just smaller than the smallest particles in the oversize. An actual screen does not do this, and appreciable portions of both fractions have the same size range. Even testing screens shows this overlap between adjoining fractions. *See* MECHANICAL SEPARATION TECHNIQUES.



Types of screens. (*a*) Trommel. (*b*) Gyratory.

Warren L. McCabe

Bibliography. A. Grandison and M. Lewis, *Separation Processes*, 1993; W. L. McCabe, J. C. Smith, and P. Hariott, *Unit Operations in Chemical Engineering*, 5th ed., 1993.

Screening smoke

Smoke clouds used for screening purposes. Traditionally, smoke is defined as a visible aerosol (that is, a suspension of solid or liquid particles in a gas) resulting from incomplete combustion. Military screening smokes are a special case. They are generally composed of liquid droplets with a mean particle size around 1 micrometer, and produce a bright white cloud of long duration. The effectiveness of these aerosols results from their size and composition. The individual aerosol particles efficiently scatter light in all directions. This scattering, known as Mie scattering, reduces the intensity of any light passing through the cloud and produces considerable glare, the white cloud effect, to reduce the target-to-background contrast. *See* SCATTERING OF ELECTROMAGNETIC RADIATION; SMOKE.

In warfare, screening smokes have three major functions: blinding, covering, and deception.

Blinding smokes are placed on an enemy's position to cause confusion and reduce enemy effectiveness. Covering smokes are usually placed between the enemy and a combat unit or on the unit's position to deny the enemy knowledge of the unit's location and actions. Deception, that is, putting a smoke screen in a location of no immediate interest, has been used throughout history to misdirect enemy interests and to gain surprise.

Screening smoke compositions have remained unchanged since the mid-1940s. However, methods of delivery have been improved to keep pace with the rapid rise in the use of high-speed, sophisticated weaponry.

There are four well-established methods of smoke production: burning phosphorus in air, burning pyrotechnic compositions, vaporization and recondensation of oils, and dispersion of reactive liquids.

Phosphorus smokes. The most widely used and effective screening smokes are formed by burning elemental phosphorus in air. The reacting phosphorus rapidly forms a variety of phosphorus oxides [reactions (1) and (2)], which absorb atmospheric

$$P_4 + 5O_2 \rightarrow 2P_2O_5 \quad (1)$$

$$P_2O_5 + 3H_2O \rightarrow 2(H_3PO_4) \quad (2)$$

moisture to form aerosols of dilute phosphoric acid. Phosphorus smokes may be delivered as white phosphorus, plasticized white phosphorus, or red phosphorus. Various munition configurations have been developed to control the burning rate of the phosphorus oxidation reaction and thus eliminate the problem of thermal pluming found in early white phosphorus munitions (see **illus.**). Phosphorus smoke clouds are very efficient. They have a yield factor (that is, the ratio of the mass of smoke produced to the mass of starting material) of between 3 to 7 depending on the relative humidity. These high yields result from the phosphorus oxides being hygroscopic and thus absorbing significant amounts of atmospheric water. Phosphorus smokes are delivered from artillery projectiles, grenades, mortar rounds, and rockets. *See* PHOSPHORUS.

Thermal pluming from a white phosphorus munition. (*U.S. Army*)

Pyrotechnic compositions. Aerosols formed by burning pyrotechnic compositions generally are composed of hydrated chloride salts. Most NATO countries use a composition known as HC, a mixture of zinc oxide, hexachloroethane, and various burning rate modifiers, which produces a hydrated zinc chloride aerosol. Other nations use an alternate composition, often called the Yershov composition, which contains ammonium chloride plus fuels and modifiers and produces an ammonium chloride aerosol. Both formulations have a yield factor ranging from 0.9 to 2, depending on relative humidity, and both produce aerosols with similar optical properties. The principal military munitions containing these mixtures are grenades and smoke pots. *See* PYROTECHNICS.

Oil smokes. Oil smokes are used when large areas must be screened for long periods. These smokes are produced by vaporizing an oil and allowing it to cool in air, where it condenses into fine aerosol droplets that form a stable cloud. In general, they are much less expensive than other methods for producing screens. Oil smokes are produced by spraying diesel oil on the exhaust manifold of a vehicle and allowing the exhaust gas pressure to propel the cloud away from the vehicle, or by using fog oil, a specially produced smoke oil, with mechanical smoke generators. The optical properties of the smoke produced by the two oils are identical; however, fog oil clouds last much longer because of their lower vapor pressure. During World War II, the Allies used simultaneous operation of many fog oil generators to screen several square miles of terrain, port areas, and staging zones from air observation. Some of these screens lasted for over 30 days.

Oil smokes are also useful for nonmilitary purposes. Farmers have used oil smoke for many years to prevent the freezing of fruit and other crops. The blanket of smoke reduces the loss of heat by radiation from ground and plant surfaces. Oil smoke is also used as a carrier for insecticides and fungicides in agriculture and forestry.

Reactive liquid smokes. The most common reactive smokes are chlorosulfonic acid, which produces sulfuric acid smoke, and titanium tetrachloride, which produces a dilute titanium hydroxide and hydrochloric acid smoke mixture. Both have optical properties similar to phosphorus smokes. These reactive liquids are obsolete in the U.S. Army inventory for several reasons: they are highly corrosive and dangerous; smoke is produced by hydrolysis reactions, thus making them ineffective in extremely dry conditions; and their main method of employment, spray from tanks on the wings of slow-flying aircraft, is no longer militarily acceptable.

Joseph J. Vervier

Bibliography. A. Baily and S. G. Murray, *Explosives, Propellants, and Pyrotechnics*, 1989; J. A. Compton, *Military Chemical and Biological Agents: Chemical and Toxicological Properties*, 1988; W. C. Hinds, *Aerosol Technology; Properties, Behavior, and Measurement of Airborne Particles*, 2d ed., 1998.

Screw

A cylindrical body with a helical groove cut into its surface. For practical purposes a screw may be considered to be a wedge wound in the form of a helix so that the input motion is a rotation while the output remains translation. The screw is to the wedge much the same as the wheel and axle is to the lever in that it permits the exertion of force through a greatly increased distance. *See* WEDGE.

Figure 1 shows a frictionless screw with square threads mounted on a ball thrust bearing being used to raise a load Q. The load and other forces may be considered to be concentrated at an effective radius called the pitch radius. The force diagram is similar to the diagram for a wedge, except that wedge angle θ is replaced by screw lead angle λ, which is the angle between the thread of the helix and a plane perpendicular to the axis of rotation. Subscript t is added to indicate that force F is applied tangent to a circle, giving Eq. (1). Because F_t acts at the pitch radius R, the torque on the screw is given by Eq. (2).

$$F_t = Q \tan \lambda \tag{1}$$

$$T = F_t R = QR \tan \lambda \tag{2}$$

When a practical screw is considered, friction becomes important, and the torque on a screw with square threads is given by Eq. (3), and the efficiency

$$T = QR \frac{\tan \lambda + \mu}{1 - \mu \tan \lambda} \tag{3}$$

of a screw with square threads by Eq. (4), where μ is the coefficient of friction.

$$\eta = \tan \lambda \frac{1 - \mu \tan \lambda}{\tan \lambda + \mu} \tag{4}$$

If a screw jack (**Fig. 2**) is used to raise a heavy object, such as a house or machine, it is normally desirable for the screw to be self-locking, that is, for the screw to not rotate and lower the load when the torque is removed with the load remaining on the jack. *See* SCREW JACK.

For a screw to be self-locking, the coefficient of friction must be greater than the tangent of the lead angle: $\mu > \tan \lambda$. *See* SCREW FASTENER.

The screw is by far the most useful form of inclined plane or wedge and finds application in the bolts and nuts used to fasten parts together; in lead and feed screws used to advance cutting tools or parts in machine tools; in screw jacks used to lift such objects as automobiles, houses, and heavy machinery; in screw-type conveyors (**Fig. 3**) used to move bulk materials; and in propellers for airplanes and ships. *See* SCREW THREADS; SIMPLE MACHINE.

Richard M. Phelan

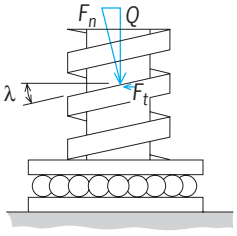


Fig. 1. Diagram of screw. See text for symbols.

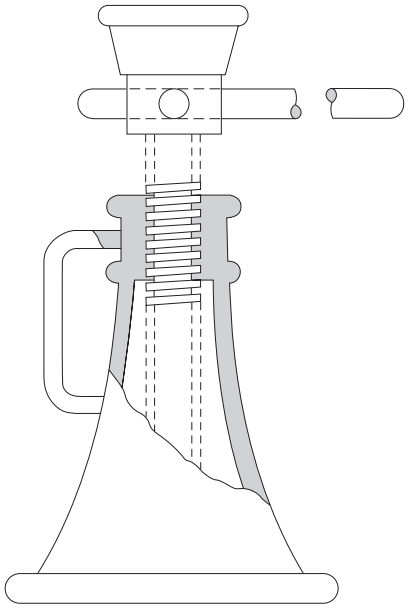


Fig. 2. Screw jack, used to raise heavy objects.

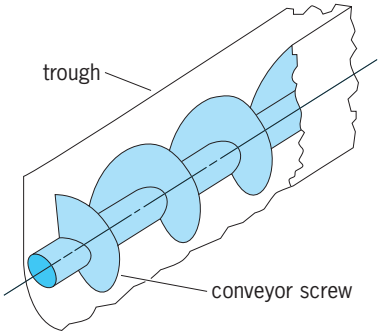


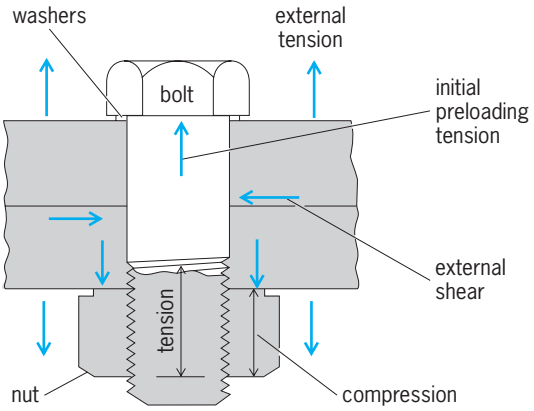
Fig. 3. Screw conveyor, used to move bulk objects.

Bibliography. J. Phillips, *Freedom in Machinery*, vol. 1: *Introducing Screw Theory*, 1984, vol. 2: *Screw Theory Exemplified*, 1990.

Screw fastener

A threaded machine part used to join parts of a machine or structure. Screw fasteners are used when a connection that can be disassembled and reconnected and that must resist tension and shear is required. A nut and bolt is a common screw fastener. Bolt material is chosen to have an extended stress-strain characteristic free from a pronounced yield point. Nut material is chosen for slight plastic flow.

The nut is tightened on the bolt to produce a preload tension in the bolt (see **illus.**). This preload has several advantageous effects. It places the bolt under sufficient tension so that during vibration the relative stress change is slight with consequent improved fatigue resistance and locking of the nut.



Forces on a bolt fastener under preloading.

Preloading also increases the friction between bearing surfaces of the joined members so that shear loads are carried by the friction forces rather than by the bolt.

The tightened nut is under compression; the bolt is under tension. The deformation that accompanies these forces tends to place the entire preload on the thread nearest the bearing surface. Thus, concentration of loading is counteracted if the nut is slightly plastic so as to set under load. This yielding may be achieved by choice of material or by special shape. If a soft gasket is used in the assembly, preloading is less effective; the bolt may carry the full tension and shear loads. *See* BOLT; JOINT (STRUCTURES); NUT (ENGINEERING); SCREW THREADS. Frank H. Rockett

Bibliography. American Society for Testing and Materials, *Section 15*, *General Products*, *Chemical Specialties*, *and End Use Products*, vol. 15.08: *Fasteners*, 1989; R. O. Parmley, *Standard Handbook of Fastening and Joining*, 3d ed., 1997; J. E. Shigley and C. R. Mischke, *Fastening*, *Joining*, *and Connecting: A Mechanical Designers' Handbook*, 1989; J. E. Shigley and J. J. Vicker, *Theory of Machines and Mechanisms*, 2d ed., 1995.

Screw jack

A mechanism for lifting and supporting loads, usually of large size. A screw jack mechanism consists of a thrust collar and a nut which rides on a bolt; the threads between the nut and bolt normally have a square shape. A standard form of screw jack has a heavy metal base with a central threaded hole into which fits a bolt capable of rotation under a collar thrusting against the load. Screw jacks are also used for positioning mechanical parts on machine tools in order to carry out manufacturing processes. These jacks are small in size and have standard V threads. Load and stress calculations may be performed on screw jacks in the same manner as in power screws and screw fastenings. *See* SCREW; SIMPLE MACHINE. James J. Ryan

Bibliography. N. Chironis, *Mechanisms and Mechanical Devices Sourcebook*, 2d ed., 1996; V. M. Faires and R. M. Keown, *Mechanism*, 5th ed., 1960, reprint 1980; S. Molian, *Mechanism Design*: *An Introductory Text,* 2d ed., 1997.

Screw threads

Continuous helical ribs on a cylindrical shank. Screw threads are used principally for fastening, adjusting, and transmitting power. To perform these specific functions, various thread forms have been developed. A thread on the outside of a cylinder or cone is an external (male) thread; a thread on the inside of a member is an internal (female) thread (**Fig. 1**). *See* SCREW.

Types of thread. A thread may be either right-hand or left-hand. A right-hand thread on an external member advances into an internal thread when turned clockwise; a left-hand thread advances when turned counterclockwise. If a single helical groove is cut or formed on a cylinder, it is called a single-thread screw. Should the helix angle be increased sufficiently for a second thread to be cut between the grooves of the first thread, a double thread will be formed on the screw. Double, triple, and even quadruple threads are used whenever a rapid advance is desired, as on fountain pens and valves. The helices on a double thread start 180° apart, while those on a triple begin 120° apart. A multiple thread produces a rapid advance without resort to a coarse thread.

Pitch and major diameter designate a thread. Lead is the distance advanced parallel to the axis when the screw is turned one revolution. For a single thread, lead is equal to the pitch; for a double thread, lead is twice the pitch. For a straight thread, the pitch diameter is the diameter of an imaginary coaxial cylinder that would cut the thread forms at a height where the width of the thread and groove would be equal.

Thread forms have been developed to satisfy particular requirements (**Fig. 2**). Those employed on fasteners and couplings and those used for making adjustments are generally of the modified 60° V type (unified and ISO metric). Where strength is required for the transmission of power and motion, a thread having faces that are more nearly perpendicular to the axis is preferred, such as the modified square and the acme. These threads, with their strong thread sections, transmit power nearly parallel to the axis of

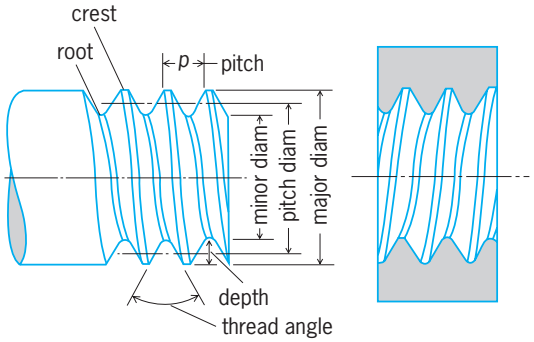


Fig. 1. Screw thread nomenclature.

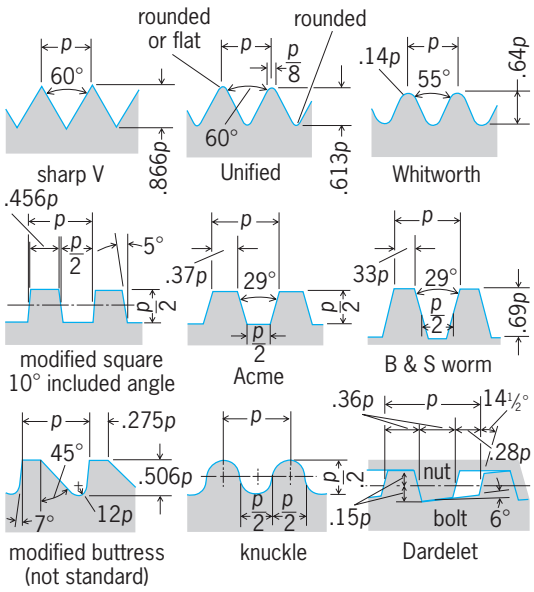


Fig. 2. Screw threads, for particular needs. (*After W. J. Luzadder, Fundamentals of Engineering, 8th ed., Prentice-Hall, 1977*)

the screw. The sharp V, formerly found on set screws, is now rarely used because of the difficulty of cutting sharp roots and crests in quantity production.

Thread forms. The unified thread form accompanied by the unified thread standards has in the past fulfilled the necessary requirements for interchangeability of threaded products between the United States, Great Britain, and Canada. This modified thread form represented two compromises between British Whitworth and United States National thread forms. The British accepted the 60° thread angle of the American thread, and the Americans accepted a rounded root and crest similar to the Whitworth form. The United States National unified form for external threads has a rounded root and may have either a rounded or flat crest. The unified thread contours provide advantages over both the earlier forms. The design provided for greater fatigue strength and permitted longer wear of cutting tools. The unified thread is a general-purpose thread that is particularly suited for fasteners.

Threads on bolts and screws may be formed by rolling, cutting, or extruding and rolling. Cut threads frequently produce a tighter fit than do rolled threads because of irregularities and burrs. To facilitate assembly, cut threads should have a finished point. Rolled threads ordinarily can be manufactured at less cost, but the range of sizes of bolts that can be produced is limited.

The Whitworth 55° thread, similar in many respects to the United States standard form, is the standard V form used in Great Britain. The unified form, identical with the United States standard form, is also standard in Great Britain.

A modified square thread, with an angle of 10° between the sides, is sometimes used for jacks and vises where strength is needed for the transmission of power in the direction of the thread axis.

When originally formulated, the acme thread was intended as an alternate to the square thread and other thread forms that were being used to produce transverse motion on machines. Used for a variety of purposes, the acme thread is easier to produce than the square thread, and its design permits the use of a split nut to provide on-off engagement. A modification of the acme is the 29° stub acme. Because the basic depth of the thread is reduced to 0.30 of the pitch, the thread section is strong and well suited to power applications where space limitations make a shallow thread desirable. A still further modification of the acme is the 60° stub having a 60° angle between the sides of the thread. This thread has a basic depth of 0.433 of the pitch. The Brown and Sharp worm thread, a modified form of the acme thread, is used for transmitting power to a worm wheel.

The buttress or breech-block thread is designed for transmitting an exceptionally high pressure in one direction only. In its original form the pressure flank was perpendicular to the thread axis, and the trailing flank sloped at 45°. To simplify the cutting of the face, modern practice is to give the pressure flank a 7° slope. This form of thread is applicable for assembled tubular members because of the small radial thrust. Buttress threads are used on breech mechanisms of large guns and for airplane propeller hubs.

The knuckle rolled thread is found on sheet-metal shells of lamp bases, fuse plugs, and in electric sockets. The thread form, consisting of circular segments forming crests and roots tangent to each other, has been standardized. The thread is usually rolled, but it may be molded or cast. The Dardelet thread is self-locking in assembly.

SI (ISO) metric thread system. The ISO (International Organization for Standardization) thread form (**Fig. 3**) has essentially the same basic profile as the unified thread and, when the changeover to the metric system has been completed in the United States, it will supplant the unified form. However, before the metric thread system is fully accepted in the United States, minor modifications of thread characteristics may be made to satisfy industry.

The principal differences between the SI metric thread system and the unified inch system are in the basic sizes, the magnitude and application of

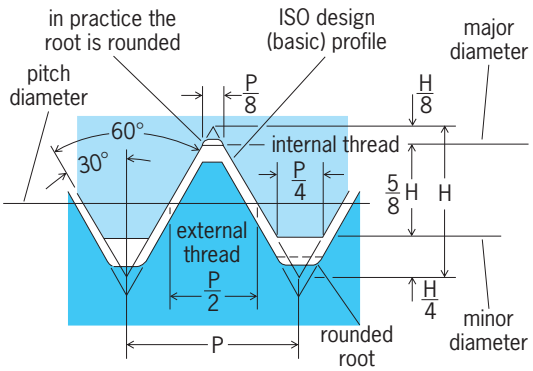


Fig. 3. ISO metric internal and external thread design profiles. (*After W. J. Luzadder, Fundamentals of Engineering Drawing, 8th ed., Prentice-Hall, 1981*)

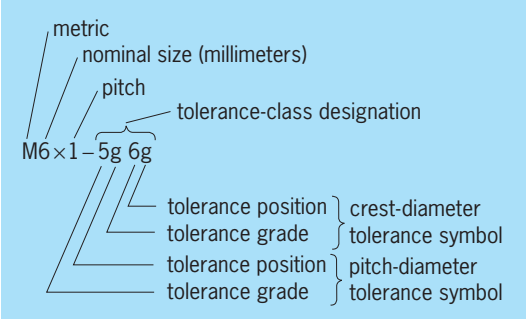


Fig. 4. Metric thread system designations.

allowances and tolerances, and the method for designating and specifying threads.

The basic designation (**Fig. 4**) consists of the letter M followed by the nominal size (basic major diameter in millimeters) and the pitch. The nominal size and pitch are separated by the sign × as in M8 × 1 (ISO designation) and M20 × 1.5.

For coarse series threads, the indication of the pitch may be omitted. The complete designation of an ISO metric thread includes the basic designation followed by the specification for tolerance class. The tolerance class designation includes the symbol for the pitch-diameter tolerance followed by the symbol for the crest-diameter tolerance. Each of these symbols consists of a number indicating the tolerance position. A dash separates the tolerance-class designation from the basic designation. When the pitch-diameter and crest-diameter tolerance symbols are identical, the symbol is given only once.

The ISO standard provides a number of tolerance grades that reflect various magnitudes. Basically, three metric tolerance grades are recommended: 4, 6, and 8. Grade 6 is commonly used for general-purpose threads. This is the closest ISO grade to the unified 2A and 2B fits. The numbers of the tolerance grades reflect the size of the tolerance. Tolerances below grade 6 are smaller and are specified for fine-quality requirements. Tolerances above grade 6 are larger and are recommended for coarse-quality and long lengths of engagement.

Tolerance positions establish the maximum material limits of the pitch and crest diameters for both internal and external threads. The series of tolerance-position symbols to reflect varying amounts of allowance are, for external threads, e = large allowance, g = small allowance, h = no allowance; and for internal threads, G = small allowance, H = no allowance.

A desired fit between mating threads may be specified by giving the internal thread tolerance-class designation followed by the external thread tolerance class. The two designations are separated by a slash, as in M6 × 1 − 6H/6g.

ISO metric thread series are selections of diameter-pitch combinations that are distinguished from each other by the different pitch applied to a specific diameter (M10 × 1.5 coarse series or M10 × 1.25 fine series). Depending upon design requirements, the thread for a particular application might be selected from the coarse-thread series, fine-thread series, or one of the contant-pitch series.

Unified and standard thread series. Unified and American National screw thread standards for screws, bolts, nuts, and other threaded parts consist of six series of threads that cover nonstandard combinations of diameter and pitch. Each series of standard threads has a specific number of threads per inch for a particular diameter. In general, the unified screw threads are limited to three series: coarse (UNC), fine (UNF), and extrafine (UNEF). A 1/4-in.-diameter (6-mm) thread in the UNC series has 20 threads/in., while in the UNF series it has 28.

In the unified and American National standards, the coarse thread series (UNC and NC) is recommended for general industrial use on threaded machine parts and for screws and bolts. Because a coarse thread has fewer threads per inch than a fine thread of the same diameter, the helix angle is greater, and the thread travels farther in one turn. A coarse thread is preferred when rapidity and ease of assembly are desired, and where strength and clamping power are not of prime importance.

The fine thread series (UNF and NF) is recommended for general use in the automotive and aircraft field for threads subject to strong vibration. In general, a fine thread should be used only where close adjustment, extra strength, or increased resistance to loosening is a factor.

The extrafine thread series (UNEF and NEF) is used principally in aeronautical structures where an extremely shallow thread is needed for thin-walled material and where a maximum practicable number of threads is required for a given length.

The 8-thread series (8N) with 8 threads/in. for all diameters is used on bolts for high-pressure flanges, cylinder-head studs, and other fasteners against pressure, where it is required that an initial tension be set up so that the joint will not open when steam or other pressure is applied. This series has come into general use for many types of engineering work. It is sometimes used as a substitute for the coarse thread series for diameters greater than 1 in. (2.5 cm).

The 12-thread series (12UN or 12N) is a uniform pitch series that is widely used in industry for thin nuts on shafts and sleeves. This series is considered to be a continuation of the fine-thread series for diameters greater than $1^1/_2$ in. (3.8 cm).

The 16-thread series (16UN or 16N) is a uniform pitch series for applications that require a very fine thread, as in threaded adjusting collars and bearing retaining nuts. This series is used as a continuation of the extrafine thread series for diameters greater than 2 in. (5 cm).

The manufacturing tolerance and allowance permitted distinguish one class of thread from another under the unified thread system. Because allowance is an intentional difference between correlated size dimensions of mating threads, and tolerance is the difference between limits of size, a thread class may be considered to control the looseness or tightness between mating threaded parts. The American National standard for unified and National screw

threads provides classes 1A, 2A, and 3A for external threads, classes 1B, 2B, and 3B for internal threads, and classes 2 and 3 for both internal and external threads. Classes 1A and 1B are for ordnance and other special uses. Classes 2A and 2B are the recognized standards for the bulk of screw-thread work and for the normal production of threads on screws, bolts, and nuts. Classes 3A and 3B are for applications where smaller tolerances than those afforded by class 2A and 2B are justified and where closeness of fit between mating threads is important. Class 3A has no allowance.

The screw-thread fit needed for a specific application can be obtained by combining suitable classes of thread. For example, a class 2A external thread might be used with a class 3B internal thread to meet particular requirements.

Designation of unified, square, and acme threads. Threads are designated by standardized notes (**Fig. 5**). Under the unified system, threads are specified by giving in order the nominal diameter, number of threads per inch, the initial letters such as UNC or UNF that identify the series, and class of thread (1A, 2A, and 3A or 1B, 2B, and 3B). Threads are considered to be right-hand and single unless otherwise noted; therefore the specification for a left-hand thread must have the letters LH included. To indicate multiplicity, the word DOUBLE, TRIPLE, and so on must follow the class symbol.

Thread fastener uses. Most threaded fasteners are a threaded cylindrical rod with some form of head on one end. Various types of threaded fasteners are available; some, such as bolts, cap screws, and machine screws, have been standardized; others are special designs. The use of removable threaded fasteners is necessary on machines and structures for holding together those parts that must be frequently disassembled and reassembled. Because standard threaded fasteners are mass produced at relatively low cost and are uniform and interchangeable, they are used whenever possible.

Fasteners are identified by names that are descriptive of either their form or application: set screw, shoulder screw, self-tapping screw, thumb screw,

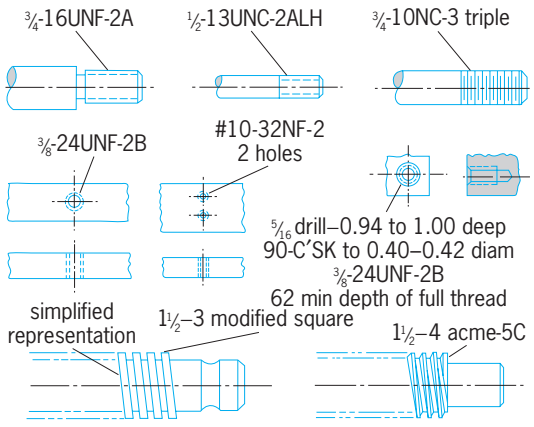


Fig. 5. Thread identification by standardized symbols. (*After W. J. Luzadder, Fundamentals of Engineering Drawing, 8th ed., Prentice-Hall, 1981*)

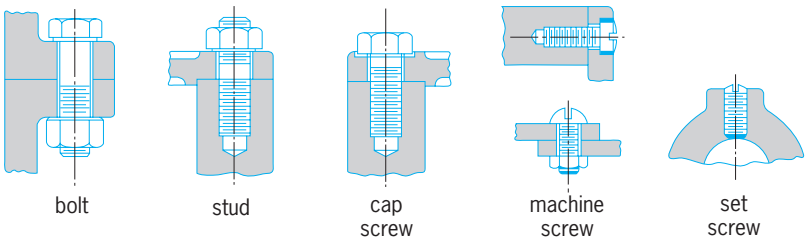


Fig. 6. Common types of fasteners. (*After T. E. French and C. J. Vierck, Engineering Drawing and Graphic Technology, 12th ed., McGraw-Hill, 1978*)

and eyebolt. Of the many forms available, five types meet most requirements for threaded fasteners and are used for the bulk of production work: bolt, stud, cap screw, machine screw, and set screw (**Fig. 6**). Bolts and screws can be obtained with varied heads and points (**Fig. 7**).

Fastener materials. Metallic materials and certain plastics are employed as thread fasteners. Fasteners most commonly used in commercial practice are made of alloys of aluminum, copper (brass and bronze), and steel. For special applications (when required to be heat or corrosion resistant, for example), they are made from a wide variety of materials—metallic or nonmetallic—suitable to the operating environment. Stainless steels, titanium alloys, high-nickel alloys, and nylon are among such materials.

A bolt is generally used for drawing two parts together. Having a threaded end and an integral head formed in manufacture, it is passed through aligned clearance holes in the two parts and a nut is applied.

A stud is a rod threaded on both ends. Studs are used for parts that must be removed frequently and for applications where bolts would be impractical. They are first screwed more or less permanently into one part before the removable member with corresponding clearance holes is placed into position. Nuts are used on the projecting ends to draw the parts together.

Cap screws (bare or plated) are widely used in machine tools and to assemble parts in all types of machinery and equipment. The most common heads are hexagonal, flat, round, fillister, and socket (Fig. 7 and **Fig. 8**). Torque is applied to the fastener by means of a turning device that has a tip designed to mate with a conforming contour or socket in the fastener head; the latter may take the form of a straight slot, hexagonal or fluted socket, crossed Phillips head, or any number of special configurations generally called "tamper proof."

When mating parts are assembled, the cap screws pass through clear holes in one member and screw into threaded holes in the other. The head, an integral part of the fastener along with the thread, holds the parts together. In the automotive industry, the hexagon-head cap screw in combination with a nut is often used as an automotive hexagon-head bolt.

Machine screws, which are similar to cap screws and fulfill the same purpose, are employed principally in the numbered diameter sizes on small work having thin sections or with a machine nut to function as a small bolt.

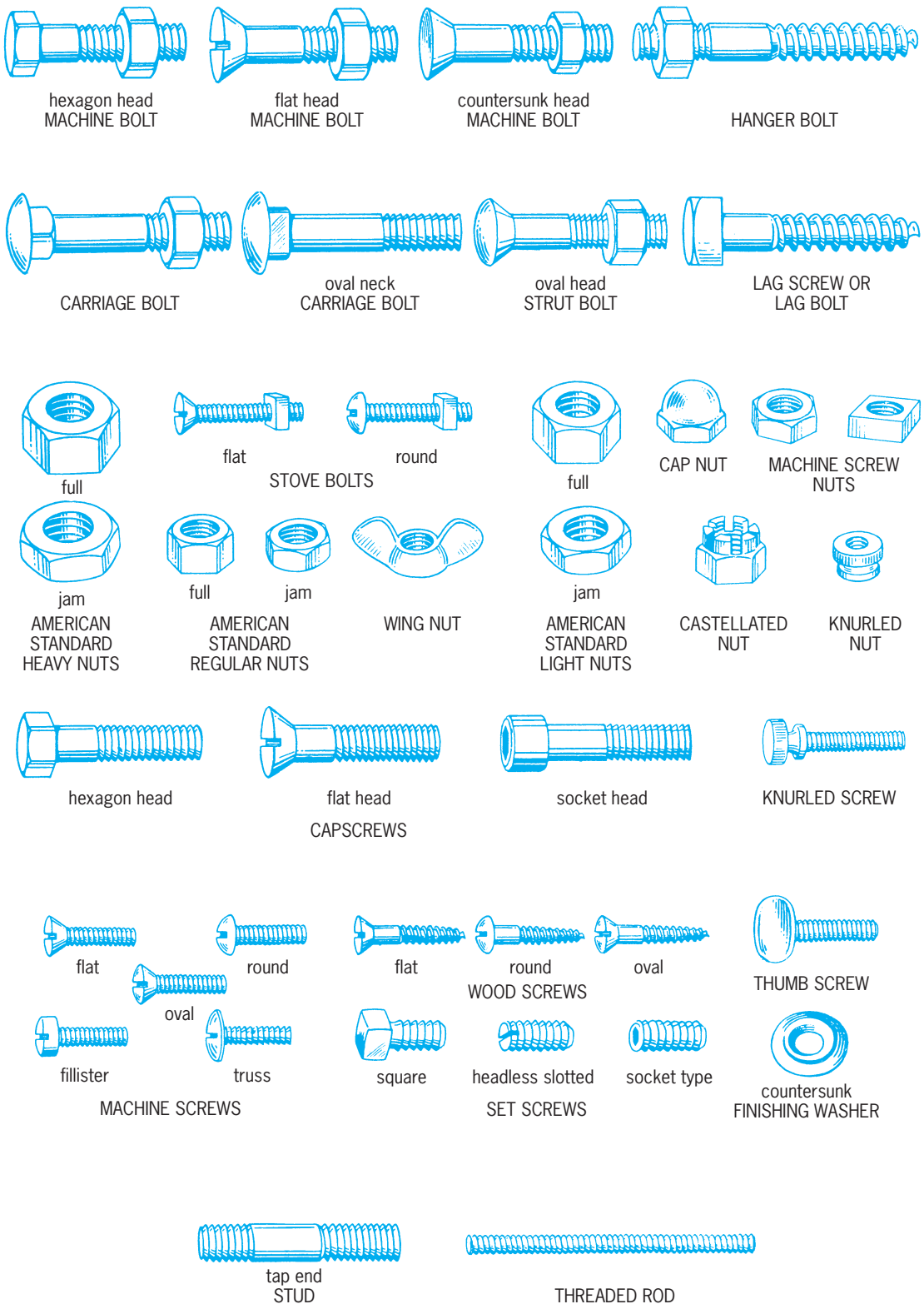


Fig. 7. Threaded fasteners are made in a wide variety of types and used for various purposes. Some of the most common are shown here. Screws may be slotted or hollow-headed. Hollow-headed screws may have a socket or Phillips head or any of a number of special contours for patented turning systems. (*Reynolds Metals Co.*)

Set screws made of hardened steel are used to hold parts in a position relative to one another. Normally, their purpose is to prevent rotary motion between two parts, such as would occur in the case of a rotating pulley and shaft combination. Set screws can be purchased with any one of six types of points in combination with any style of head (Fig. 7 and **Fig. 9**). In the application of set screws, a flat surface is formed on the shaft to provide not only a seat for the screw point, but also clearance space for any burr

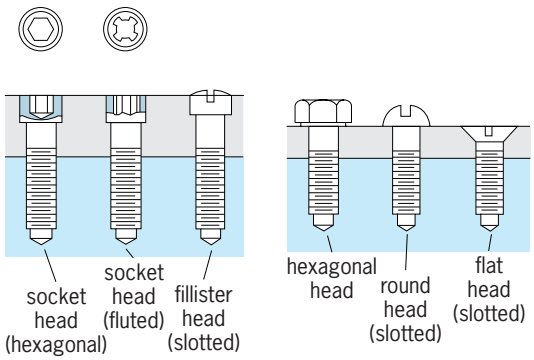


Fig. 8. Cap screws. (*After W. J. Luzadder, Fundamentals of Engineering Drawing, 8th ed., Prentice-Hall, 1981*)

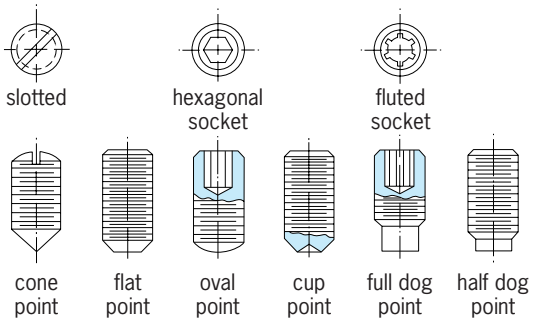


Fig. 9. Set screws. (*After W. J. Luzadder, Fundamentals of Engineering Drawing, 8th ed., Prentice-Hall, 1981*)

which may be raised when the screw is tightened. Without that clearance, the shaft is removed from the assembly with great difficulty, and may result in significant damage to the shaft as well as the bearing.

Wood screws, lag screws, and hanger bolts are used in wood (Fig. 7). Wood screws may have either slotted or recessed heads.

Self-tapping screws (**Fig. 10**) have a specially hardened thread that makes it possible for the screws to form their own internal thread in sheet metal and soft materials when driven into a hole that has been drilled, punched, or punched and reamed. The use of self-tapping screws eliminates costly tapping operations and saves time in assembling parts.

Metric fasteners (ISO). Existing standards for the production of threaded metric fasteners in the United States are in general agreement with ISO standards, with minor differences. When these differences have been resolved, final standards bearing the approval of both ISO and ANSI (American National Standards Institute) will become available for worldwide use. These approved metric fastener standards will assure international interchangeability. Being more restrictive in the choices of sizes and styles than the standards of the past, the new metric standards will make possible reduced inventories and permit bulk pricing advantages.

Metric threaded fasteners are available in only one series of diameter-pitch combinations that range from M1.6 × 0.35 to M100 × 6. The thread pitches are between those of the coarse thread and fine thread series of the present unified (inch) threads. Except for metric socket screws, all threaded fastener products have one tolerance grade, 6g for bolts, machine screws, and so forth, and 6H for nuts. The 6g/6H threads closely match the unified 2A/2B threads for fasteners, in that there is an allowance on the external thread and no allowance on the internal thread.

The standard lengths for short fasteners are 8, 12, 14, 16, and 20 mm. Fastener lengths from 20 to 100 mm increase in length by 5-mm increments. Metric fasteners are specified by giving in sequence the nominal size in millimeters, thread pitch, nominal length, product name (including type of head), material, and protective finish, if needed. Examples are M24 × 3 × 50 stainless steel hex bolt; M14 × 2 × 80 hex cap screw cadmium plated; M8 × 1.25 × 14 hex socket flat point set screw.

Pipe thread. Standard (Briggs) taper pipe threads (**Fig. 11**) are cut (on a cone) to a taper of $^1/_{16}$ in./in. to ensure a tight joint. Although a normal connection employs a taper external and a taper internal thread, an American National standard straight pipe thread having the same pitch angle and depth of thread as the corresponding taper pipe thread is used for pressure-tight joints for couplings, pressure-tight joints for fuel- and oil-line fittings, and for loose-fitting and free-fitting mechanical joints. Assemblies made with taper external threads and straight internal threads are frequently preferred to assemblies employing all taper threads; the assumption is made that relatively soft or ductile metals will adjust to the taper external pipe thread. A modified pipe thread, the United States standard Dryseal pipe thread (taper

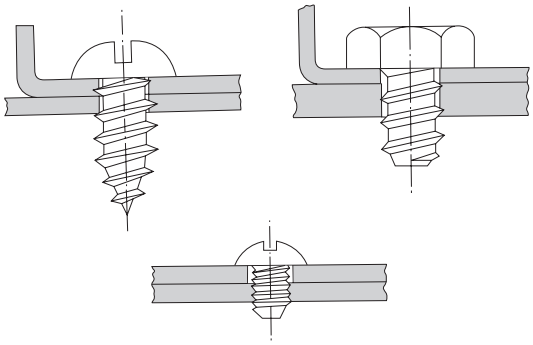


Fig. 10. Three applications of self-tapping screws.

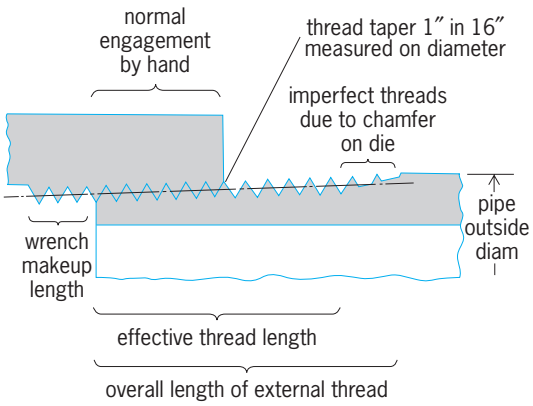


Fig. 11. Standard (Briggs) taper pipe thread.

and internal straight), is used for pressure-tight connections that are to be assembled without lubricant or sealer. Except for a difference in the truncation at the roots and crests, the general form and dimensions of this thread are the same as those of the standard taper pipe thread. The principal uses for the Dryseal thread are in refrigerant pipes, automotive and aircraft fuel-line fittings, and for gas and chemical shells (ordnance). For railing joints, where a rigid mechanical thread joint is needed, external and internal taper threads are used. Basically, the external thread is the same as the standard taper pipe thread, except that the length of the thread is shortened at the end to permit use of the larger-diameter portion of the thread. Warren J. Luzadder

Bibliography. T. E. French, C. J. Vierck, and R. J. Foster, *Engineering Drawing and Graphic Technology*, 14th ed., 1993; Industrial Fasteners Institute, *Metric Fastener Standards*, 1986; W. J. Luzasser and J. M. Duff, *Fundamentals of Engineering Drawing*, 11th ed., 1992; R. O. Parmley, *Standard Handbook of Fastening and Joining*, 3d ed., 1997; J. E. Shigley and C. R. Mischke, *Fastening, Joining, and Connecting: A Mechanical Designer's Handbook*, 1989.

Scrophulariales

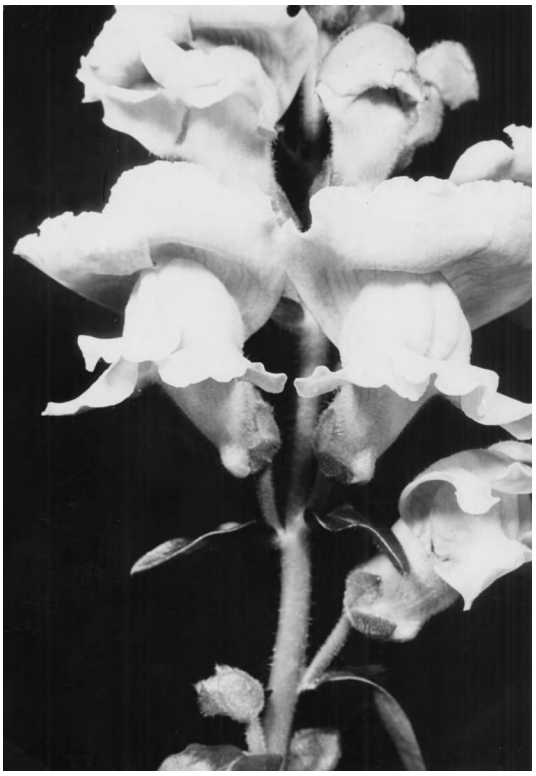
An order of flowering plants, division Magnoliophyta (Angiospermae), in the subclass Asteridae of the class Magnoliopsida (dicotyledons). The order consists of 12 families and more than 11,000 species. The largest families are the Scrophulariaceae (about 4000 species), Acanthaceae (about 2500 species), Gesneriaceae (about 2500 species), Bignoniaceae (about 800 species), and Oleaceae (about 600 species). The Scrophulariales are Asteridae, with a usually superior ovary and generally either with an irregular corolla or with fewer stamens than corolla lobes, or commonly both. They uniformly lack stipules.

The Scrophulariaceae are characterized by their usually herbaceous habit; irregular flowers with usually only two or four functional stamens; axile placentation; and dry, dehiscent fruits with more or less numerous seeds that have endosperm. Such well-known ornamentals as snapdragon (*Antirrhinum majus*; see **illus.**), foxglove (*Digitalis purpurea*), and calceolaria belong to the Scrophulariaceae.

The Acanthaceae are distinguished by their chiefly herbaceous habit, mostly simple and opposite leaves, lack of endosperm, and especially the enlarged and specialized funiculus, which is commonly developed into a jaculator that expels the seeds at maturity. The family includes several fairly common ornamentals, such as species of *Acanthus*, *Aphelandra*, *Beloperone* (shrimp plant), and *Strobilanthes*.

Some other well-known members of the Scrophulariales are the African violet (*Saintpaulia ionantha*, in the Gesneriaceae), catalpa (Bignoniaceae), and lilac (*Syringa vulgaris*, in the Oleaceae). *See* ASTERIDAE; FLOWER; MAGNOLIOPHYTA; MAGNOLIOPSIDA; OLIVE; ORNAMENTAL PLANTS; PLANT KINGDOM.

Arthur Cronquist; T. M. Barkley



Flowers of the common snapdragon (*Antirrhinum majus*), a typical member of the family Scrophulariaceae and order Scrophulariales. (*J. Douglas Spillane*, *National Audubon Society*)

Scyphozoa

A class of the phylum Cnidaria traditionally containing four major groups—Coronatae, Rhizostomeae, Semaeostomeae, and Stauromedusae—all marine or estuarine. Most occur in two forms, the polyp, or scyphistoma, and the medusa, or scyphomedusa. However, some (the Stauromedusae) are polyplike and sessile throughout their lives, while others (such as the semaeostomes *Pelagia* and *Stellamedusa* and the coronate *Periphylla periphylla*) are always pelagic and lack the sessile scyphistoma stage. Among the Cnidaria, the Scyphozoa are characterized by having well-developed medusae of large size and fairly well organized scyphistomae of small size. The Scyphozoa are most closely related to the other medusae-containing classes—Cubozoa and Hydrozoa—within the Cnidaria, all of which evolved from a polyp-only ancestor currently represented by members of Class Anthozoa. Recent analyses of DNA and morphology indicate that the Stauromedusae group may represent an intermediate evolutionary stage between Class Anthozoa and Medusozoa (which would encompass Cubozoa, Hydrozoa, Coronatae, Rhizostomeae, and Semaeostomeae) and may, in fact, represent a distinct class of cnidarian (Class Staurozoa).

Morphology. Scyphistomae are usually solitary and small, measuring much less than an inch (only a

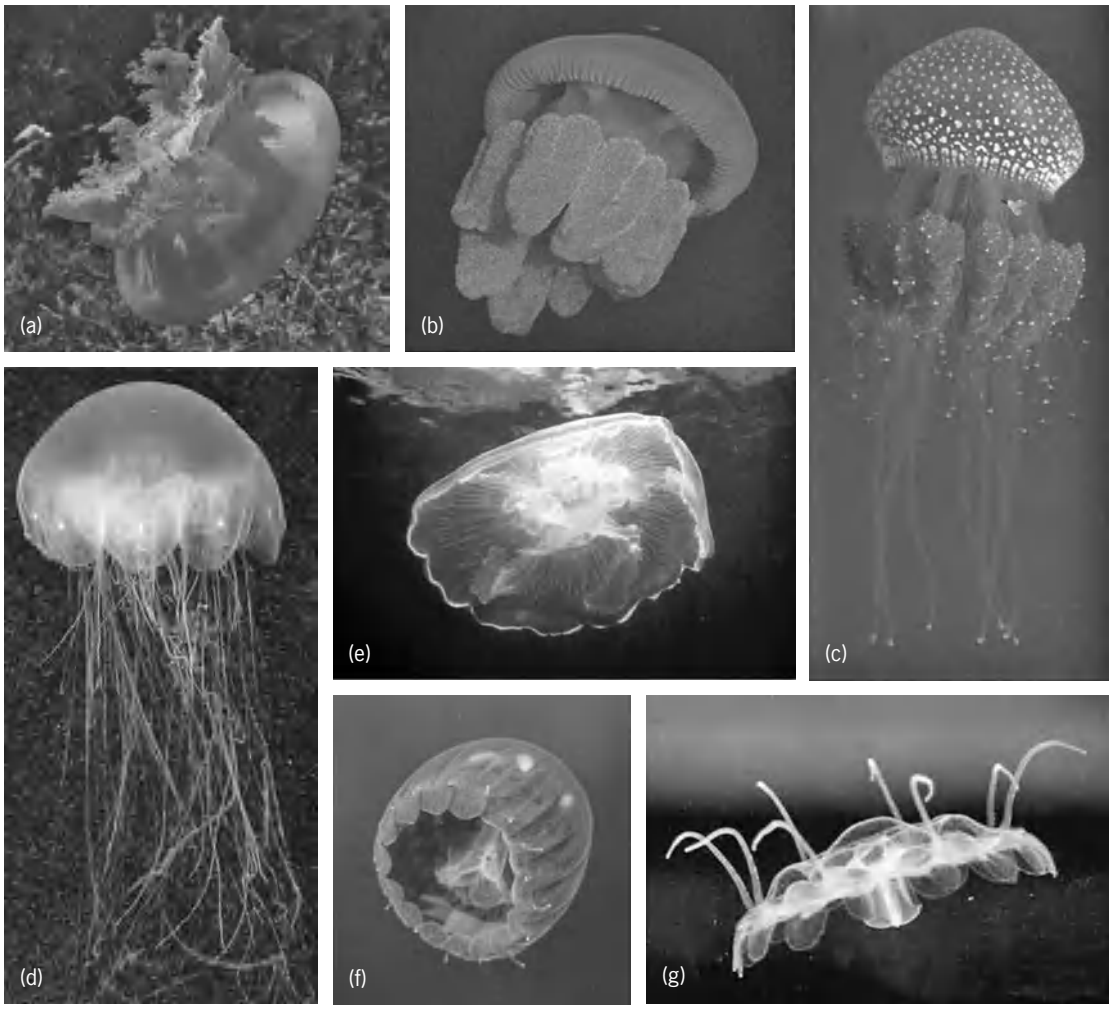


Fig. 1. Some common genera of scyphomedusae. Order Rhizostomeae. (*a*) *Cassiopea* (*photo by L. E. Martin*), (*b*) *Catostylus* (*photo by K. Pitt*), (c) *Phyllorhiza* (*photo by K. Pitt*). Order Semaeostomeae: (*d*) *Cyanea* (*photo by G. Edgar, reproduced with permission from Edgar, 2000*), (e) *Aurelia*. Order Coronatae: (*f*) *Linuche* (*photo by L. E. Martin*), (*g*) *Nausithoe* (*photo by A. Migotto*).

few millimeters) in height. A minority (including *Stephanoscyphus*, the polyp of *Nausithoe*) form colonies up to 4 to 10 in. (100–250 mm) high. They are sessile. Each solitary scyphistoma is composed of three major parts—pedal disk, stalk, and calyx—and is shaped, overall, something like a wine glass or vase. The narrow pedal disk has many gland cells that secrete a sticky fluid which attaches the scyphistoma to hard substrate. The stalk, along which run fine muscular strands enabling the polyp to contract, leads away from the substrate, then broadens into the calyx. The calyx is crowned by the oral disk with a circlet of tentacles around its perimeter and a four-sided mouth at its center. The mouth opens into the stomodeum, which leads to the stomach. The axes running from the four corners of the mouth are called perradii, and halfway between the perradii are the interraddii. At each interradius of the polyp there is an infundibulum, a canal which leads down into the stalk. The polypoid Stauromedusae are peculiar in having many of the features (for example, rhopalia, or sensory structures, and gonads) that are found only in the free-living medusae of Semaeostomeae, Rhizostomeae, and Coronatae.

Scyphomedusae, when fully grown, are generally large, sometimes attaining a diameter of 3 ft (1 m) and weighing over 300 lb (140 kg). Most scyphomedusae are natatory (adapted for swimming). The scyphomedusa is composed of the umbrella (or bell), which is the major propulsive structure, generally appended with oral arms. Differences in the shape, structure, ornamentation, and color of the bell and oral arms are the main features used to distinguish major groups and species of Scyphozoa (**Figs. 1** and **2**). The convex upper surface of the umbrella is called the exumbrella; the concave inner surface, the subumbrella. Muscles are distributed mainly in the subumbrella and the most important muscles are the ring muscles, which contract the umbrella for locomotion of the medusa. The umbrella may have various shapes, being disk-, cone-, or domelike. In the Coronatae, the central portion is partially separated from the skirtlike peripheral regions (the bell margin) by the coronal groove, from which the order gets its name. In all groups of scyphomedusae the margin of the umbrella is divided into many lappets (Fig. 2). Rhopalia, of

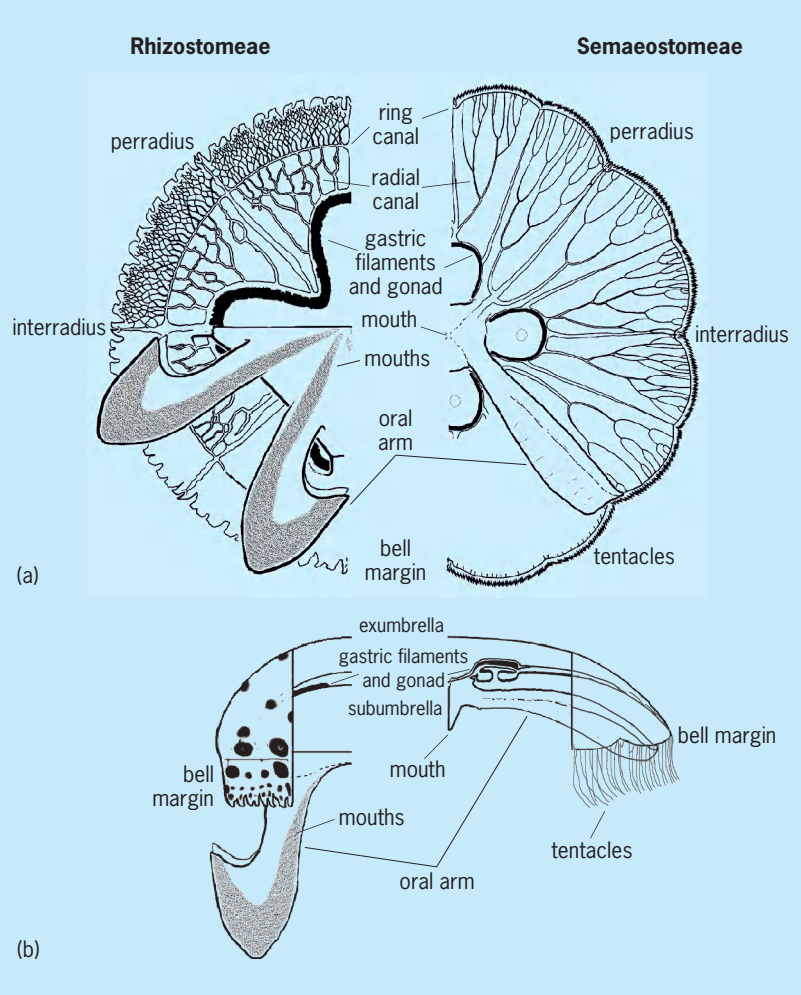


Fig. 2. Some of the major anatomical features of rhizostome (left) and semaeostome (right) medusae. (*a*) View of the subumbrella emphasizing features of the oral arms (lower portions) and the bell (cutaway, upper portions). (*b*) Side view of external and, with cutaway, internal features. The shape and size of these and other anatomical features can very considerably and are used to distinguish among taxa from class-level down to species.

which there are usually eight, are found between the lappets. Each rhopalium is composed of a statocyst or vesicle, containing a round inorganic concretion, and an ocellus, a structure composed of a lens, retina cells, and pigment cells. Tentacles also are found on the bell margin, or marginal regions of the subumbrella, and vary in number from eight to several hundred with the exception of the Rhizostomeae and some deep-water semaeostomes (for example *Deepstaria, Stellamedusa*, and *Tiburonia*) which have no tentacles. Tentacles, when present, have numerous cnidocytes, or "nettle cells," and longitudinal muscles which enable the tentacles to be extended and contracted to catch prey. The mouth, situated in the center of the subumbrella, is surrounded by four simple lips in some forms (such as Coronatae), but generally there are four well-developed oral arms (eight in Rhizostomeae). The oral arms may be used to catch and manipulate prey, to pass food from the tentacles to the mouth, and/or to brood larvae. The oral arms are most obvious in the Rhizostomeae, often being as substantial as the umbrella. In Rhizostomeae, the oral arms provide a large surface area covered by many small suctorial mouths that lead by way of canals to the gut, thus replacing the single larger central mouth adjacent to the stomach in semaeostome and coronate scyphomedusae. The stomach is often large and composed of four interradial pouches forming a cross-shape within the bell, particularly in the Rhizostomeae. Each pouch contains several rows of gastric filaments, involved in extracellular digestion, as well as the gonads.

Development and reproduction. The gonads develop from endoderm cells just above the gastric filaments. Medusae are generally gonochoristic, that is, have separate male and female individuals. The ripe ova, which are liberated from the broken gastrodermal wall of the ovary, enter the stomach cavity and then are discharged from the mouth into seawater. Fertilization may occur in the female gonad, in the stomach cavity, on the oral arms, or in open seawater. Cleavage is mainly radial, and the new endoderm is formed by ingression or epiboly. The egg develops into a ciliated planula that swims freely, except in Stauromedusae, in which the planula lacks cilia and creeps on the substratum. The planula attaches to seaweed or rocks and changes into a scyphistoma, which includes adding tentacles.

Scyphistomae reproduce asexually by various modes, including budding and pedal cysts, giving rise to other autonomous scyphistomae (or members of a colony). Under appropriate conditions, a scyphistoma will metamorphose into a strobila and, via an asexual process known as strobilation, which involves transverse differentiation and fission of the polyp, produce nascent scyphomedusae called ephyrae. An ephyra undergoes further metamorphosis into a young medusa. Thus most scyphozoans have a bipartite (two-stage) life history, alternating between sessile polyp and free-living medusa forms (**Fig. 3**). There are, however, exceptions. For example, all stauromedusae combine many features of the medusa into a polypoid stage (they have no free-living medusa phase); and the planulae of the semaeostome genus, *Pelagia*, and coronate species *Periphylla periphylla* develop directly into ephyrae.

Physiology. The major sensory structures are the rhopalia; ocelli permit medusae to sense and respond to changes in light intensity, and statocysts enable them to swim oriented relative to a vertical axis defined by the direction of gravity. The medusae may also respond to pressure, which changes with depth. In contrast, scyphistomae, which lack rhopalia, are largely indifferent to light.

Scyphomedusae and scyphistomae are sensitive to touch, often contracting the body (which alters the swimming direction of medusae) or tentacles. Chemical senses enable scyphozoans to respond to the presence of, swim toward, and catch food, although they have no structures dedicated to gustatory or olfactory functions. They are sensitive to organic acids, several fatty acids, and lipids but not to carbohydrates.

The Scyphozoa are ectothermic and their metabolic rate is influenced strongly by changes in the temperature of their environment. This is most evident in the rate of muscular contractions (pulses) of the bell, which slow dramatically when scyphomedusae are placed in colder water. However, species of scyphomedusae that live naturally at different temperatures may pulse at similar rates, indicating physiological adaptations to different environments. Scyphozoans are osmoconformers (that is, their body fluids change directly with a change in the concentrations of dissolved ions in their environment).

Biodiversity and biogeography. Important differences in morphology, DNA, and proteins used to identify jellyfish indicate that there are several hundreds of species of Scyphozoa. Most are typically found near the coast because scyphistomae are benthic. Exceptions are *Pelagia* and other wholly pelagic forms, including many of the mid- and deep-water scyphomedusae.

Order Rhizostomeae contains the most species; about 95 are currently recognized. Rhizostomes are particularly interesting because they are the most recently derived and also the most diverse scyphozoans, indicating comparatively recent rapid diversification. The rhizostomes are predominantly found in tropical regions, especially the Indo-West Pacific center of marine biodiversity. Order Semaeostomeae contains fewer species (approximately 75) which are more widely distributed, with representatives in polar, deep-sea, open-ocean, tropical, and temperate coastal regions. Coronatae is the least diverse of the three groups of scyphomedusae, with only approximately 50 species recognized, and is typically regarded as a deep-sea group although it does have a number of shallow-water representatives in both the tropics and temperate regions. There are only approximately 35 known species of stauromedusae, which are distributed predominantly in cold temperate and polar regions of northern and southern hemispheres.

Applied ecology. Scyphozoa are important members of marine ecosystems. Mostly thought of as predators (of zooplankton and sometimes fish), they also absorb dissolved organic matter and form symbioses with zooxanthellae. The number and size of medusae can increase very rapidly, a phenomenon known as a "jellyfish bloom." Blooms can disrupt human activities—blocking seawater intakes of desalination and power plants, overloading and tearing fishing trawls, and outcompeting or eating larvae of commercial fishes—and have therefore attracted considerable attention. Although rapid change in population size is a natural consequence of the jellyfish life-cycle, the concern (which there is little evidence for or against at this time) is that blooms are becoming more frequent and more intense possibly as a result of environmental degradation or invasive species. Some Semaeostomeae and Rhizostomeae are injurious to humans because of their cnidocytes. On the other hand, some medusae protect young fish and crustaceans which seek shelter among their oral arms. Some Rhizostomeae are eaten by several cultures of southeast Asia. *See* CNIDARIA.

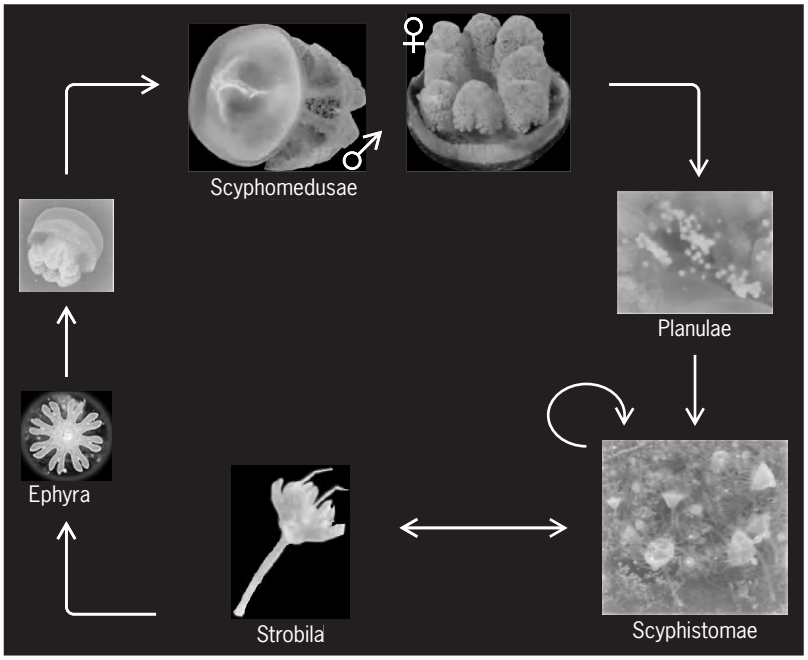


Fig. 3. Typical life history of species of Scyphozoa. Life stages not drawn to scale.

Michael N. Dawson

Bibliography. M. N. Arai, *A Functional Biology of Scyphozoa*, Chapman and Hall, London, 1997; T. Heeger, *Quallen: Gefährliche Schönheiten*, Wissenschaftliche, Stuttgart, 1998; C. E. Mills, Jellyfish blooms: Are populations increasing globally in response to changing ocean conditions?, *Hydrobiologia*, 451:55–68, 2001; F. S. Russell, *The Medusae of the British Isles. II Pelagic Scyphozoa with a Supplement to the First Volume on Hydromedusae*, Cambridge University Press, 1970; D. Wrobel and C. Mills, *Pacific Coast Pelagic Invertebrates—A Guide to the Common Gelatinous Animals*, Sea Challengers and the Monterey Bay Aquarium, 1998.

Sea anemone

Any polyp of the nearly 1000 anthozoan coelenterates belonging to the order Actiniaria. They occur intertidally and subtidally in marine and estuarine habitats, attached to solid substrates or burrowing into soft sediments. No fresh-water or truly planktonic species are known, though one genus, *Minyas*, secretes a float within its base and hangs suspended at the surface of the seas. Anemones may be very small, 0.04 in. (1 mm) long, to large individuals more than 3 ft (1 m) in length or diameter. *See* ACTINIARIA.

Each polyp is cylindrical with an adhesive base or a burrowing structure (physa) at its aboral end. The mouth occurs at the free end of the hollow body (column) and is generally surrounded by simple, unbranched tentacles (see **illus.**). Many species attach foreign material to their columns with

Sea anemone, a polyp belonging to the order Actiniaria.

varied adhesive structures. Food (fish, worms, mollusks, crustaceans, and dead organisms) is captured by the tentacles. Since sea anemones are sedentary, they depend on currents and wave action to carry their food to them.

Most anemones are capable of only limited locomotion. Attached forms creep on their basal disk, and a few move by somersaulting. Burrowing forms move through the substrate with the action of the physa. A few unrelated species swim by either writhing the body or beating the tentacles. Swimming is an escape response usually initiated by the attack of a predator such as a starfish or nudibranch.

Sea anemones carry out such movements as bending, elongating, and shortening by using muscles whose antagonist is the water within the body cavity and hollow tentacles. Water enters the body via special ciliated grooves in the actinopharynx, the siphonoglyphs. With all muscles relaxed, the cilia can create sufficient water pressure to reinflate the anemone. Elongation results from the contraction of circular muscles. Bending and shortening are caused by contraction of longitudinal muscles of the body wall, tentacular walls, and mesenteries. During normal movements, the volume of water contained by an anemone is held constant while the shape changes. The contained water is referred to as a hydrostatic skeleton. Water is released through the mouth and pores in the tips of the tentacles or body wall when the anemone is disturbed and contracts fully.

Sea anemones enter into a number of interesting symbiotic partnerships. Some are host to single-celled marine algae, which grow within the cells of the anemone. The algae provide organic materials which aid in the nutrition of the anemone. Other anemones live on gastropod shells inhabited by hermit crabs. Such anemones have special behavior which leads them to climb on gastropod shells, usually by somersaulting. Some hermit crabs can induce this behavior by stroking the anemone. Certain true crabs collect anemones and place them on their carapace and legs. With both crabs and hermit crabs, the presence of anemones tends to protect them from such predators as octopuses or other crabs. In the tropics, some anemones play host to anemone fishes which live around and among the tentacles. These fish, using a special behavior, cover themselves with mucus from the anemone and so are chemically disguised from the anemone. If the mucus is removed, the fish will be attacked by nematocysts and eaten by the host. Some anemones are very specific in their choice of substrate, occurring only on a certain species of seaweed or gorgonian.

See ANTHOZOA; COELENTERATA.

Cadet Hand

Sea breeze

A diurnal, thermally driven circulation in which a surface convergence zone often exists between airstreams having over-water versus over-land histories. The sea breeze is one of the most frequently occurring small-scale (mesoscale) weather systems. It results from the unequal sensible heat flux of the lower atmosphere over adjacent solar-heated land and water masses. Owing to the large thermal inertia of a water body, during daytime the air temperature changes little over the water while over land the air mass warms. Occurring during periods of fair skies and generally weak large-scale winds, the sea breeze is recognizable by a wind shift to onshore, generally several hours after sunrise. On many tropical coastlines the sea breeze is an almost daily occurrence. It also occurs with regularity during the warm season along midlatitude coastlines and even occasionally on Arctic shores. Especially during periods of very light winds, similar though sometimes weaker wind systems occur over the shores of larger lakes and even wide rivers and estuaries (lake breezes, river breezes). At night, colder air from the land often will move offshore as a land breeze. Typically the land breeze circulation is much weaker and shallower than its daytime counterpart.

Of all mesoscale phenomena, the sea breeze has been among the most studied, undoubtedly a result of its geographically fixed nature, its frequent occurrence, its ease of recognition using conventional observations, the concentration of observers in coastal zones, and its importance to local weather and climate. The sea breeze, among the simpler of the atmospheric circulation systems, has been frequently simulated by numerical models, expanding knowledge into areas where observations are lacking.

Structure. The occurrence and strength of the sea breeze is controlled by a variety of factors, including land-sea surface temperature differences; latitude and day of the year; the synoptic wind and its orientation with respect to the shoreline; the thermal stability of the lower atmosphere; surface solar radiation as affected by haze, smoke, and stratiform and convective cloudiness; and the geometry of the shoreline and the complexity of the surrounding terrain.

The classic sea breeze is represented as a closed circulation with a well-defined onshore flow layer, a distinct sea breeze front, (which can be very sharply defined, on the order of hundreds of meters in some cases), and a more diffuse return flow layer aloft in which air returns over the water and subsides in a

broad zone offshore. Sea breeze inflow layer depths vary from about 100 m (330 ft) to over 1000 m (3300 ft). Inland frontal penetration can vary from less than 1 km (0.6 mi) to over 100 km (60 mi), with propagation speeds ranging from nearly stationary to greater than 5 m/s (10 mi/h).

The offshore extent of the circulation is less well known, though it typically extends about twice as far offshore as the marine air has penetrated inland. Peak wind speeds are usually less than 10 m/s (20 mi/h). The overlying return flow layer depth is generally twice that of the inflow but is sometimes difficult to differentiate from the synoptic flow. Observational evidence has suggested that organized frontal zone upward motions can be as strong as 2–3 m/s (about 4–6 mi/h). Much less is known about the broader and weaker subsidence regions offshore. The initial sea breeze onshore flow is often nearly normal to the coastline, but because of the Coriolis effect the wind direction often tends to become nearly parallel to the shore by late afternoon. *See* CORIOLIS ACCELERATION; FRONT; WIND.

Impacts. The effects of the sea breeze are many, including significantly altering the direction and speed of the coastal zone winds; influencing low-level stratiform and cumuliform clouds; initiating, suppressing, and modifying precipitating convective storms; recirculating and trapping pollutants released into or becoming entrained into the circulation; perturbing regional mixing depths; and creating strong nearshore temperature, moisture, and refractive index gradients.

Cumulus suppression in the subsidence regions of the sea breeze cell offshore has been noted in satellite imagery. Especially in the tropics the sea breeze has been found to trigger the general development of thunderstorms. Sea breeze thunderstorms contribute approximately 40% of Florida's rainfall and probably more in many other coastal regions. Sea breeze circulations are major factors in exacerbating photochemical ozone pollution in cities as diverse as Los Angeles, Houston, Chicago, Oslo, Athens, and Tokyo. *See* ATMOSPHERIC GENERAL CIRCULATION; MESOMETEOROLOGY; METEOROLOGY; THUNDERSTORM.

Walter A. Lyons

Bibliography. B. W. Atkinson, *Meso-scale Atmospheric Circulations*, 1981; S. A. Hsu, *Coastal Meteorology*; 1988; T. F. Malone (ed.), *Compendium of Meteorology*, 1951; R. A. Pielke, *Mesoscale Meteorological Modeling*, 2d ed., 2001.

Sea-floor imaging

The process whereby mapping technologies are used to produce highly detailed images of the sea floor. Ever since large ships began sailing the oceans, there has been a need to know what the sea floor looks like. At first this need was simply to avoid dangerous shoals upon which a ship could wreck. In the mid-1800s, the need increased with the first laying of transocean communication cables. World War II and the initiation of submarine warfare elevated the need to a matter of national defense, and mapping of the world's ocean floors was greatly accelerated. The vast body of information accrued by this effort went far beyond aiding militaries. It also led to the theory of plate tectonics, the backbone of modern geology that explains volcanoes, earthquakes, mountain chains, and many other Earth features. At present, sea-floor mapping is also aiding a variety of societal endeavors. High-resolution images of the sea floor are being used to locate and manage marine resources such as fisheries and oil and gas reserves, identify offshore faults and the potential for coastal damage due to earthquakes, and map out and monitor marine pollution. The greatly improved sea-floor imaging is providing information on what processes are affecting the sea floor, where these processes occur, and how they interact. *See* MARINE GEOLOGY.

Mapping technologies. New, detailed images of the sea floor are revealing terrains as remarkable as those being radioed back to Earth by planetary probes. Modern technologies have revolutionized sea-floor mapping. The most commonly used are the four technologies described below, each of which provides a unique viewpoint of the sea floor.

Satellite altimetry. The technology that provides the broadest perspective but the lowest resolution is satellite altimetry (**Fig. 1***a*). A laser altimeter is mounted on a satellite and, in combination with land-based radars that track the satellite's altitude, is used to measure variations in sea-surface elevation to within 2 in. (5 cm). Removing elevation changes due to waves and currents, sea-surface height can vary up to 660 ft (200 m). These variations are caused by minute differences in the Earth's gravity field, which in turn result from heterogeneities in the Earth's mass. These heterogeneities are often associated with sea-floor topography. For example, the gravitational attraction of a massive undersea volcano pulls water toward it, producing a local bulge in the ocean surface. If the volcano is 6600 ft (2000 m) high and has an average radius of 12.5 mi (20 km), the sea surface above will typically form a bulge 6.6 ft (2 m) high. By using a mathematical function that equates sea-surface height to bottom elevations, global areas of the sea floor can be mapped within a matter of weeks. However, this approach has limitations. Sea-floor features less than 6–9 mi (10–15 km) in length are generally not massive enough to deflect the ocean surface, and thus go undetected. Furthermore, sea-floor density also affects the gravity field; and where different-density rocks are found, such as along the margins of continents, the correlation between Earth's gravity field and sea-floor topography breaks down. *See* ALTIMETER.

Side-scan sonar. This provides a higher-resolution view of the sea floor, but one which is a bit different from other technologies. In general, a side-scan sonar consists of two sonar units attached to the sides of a sled tethered to the back of a ship (Fig. 1*b*). Each sonar emits a burst of sound that insonifies a long, narrow corridor of the sea floor extending away from the sled. Sound reflections from the corridor that

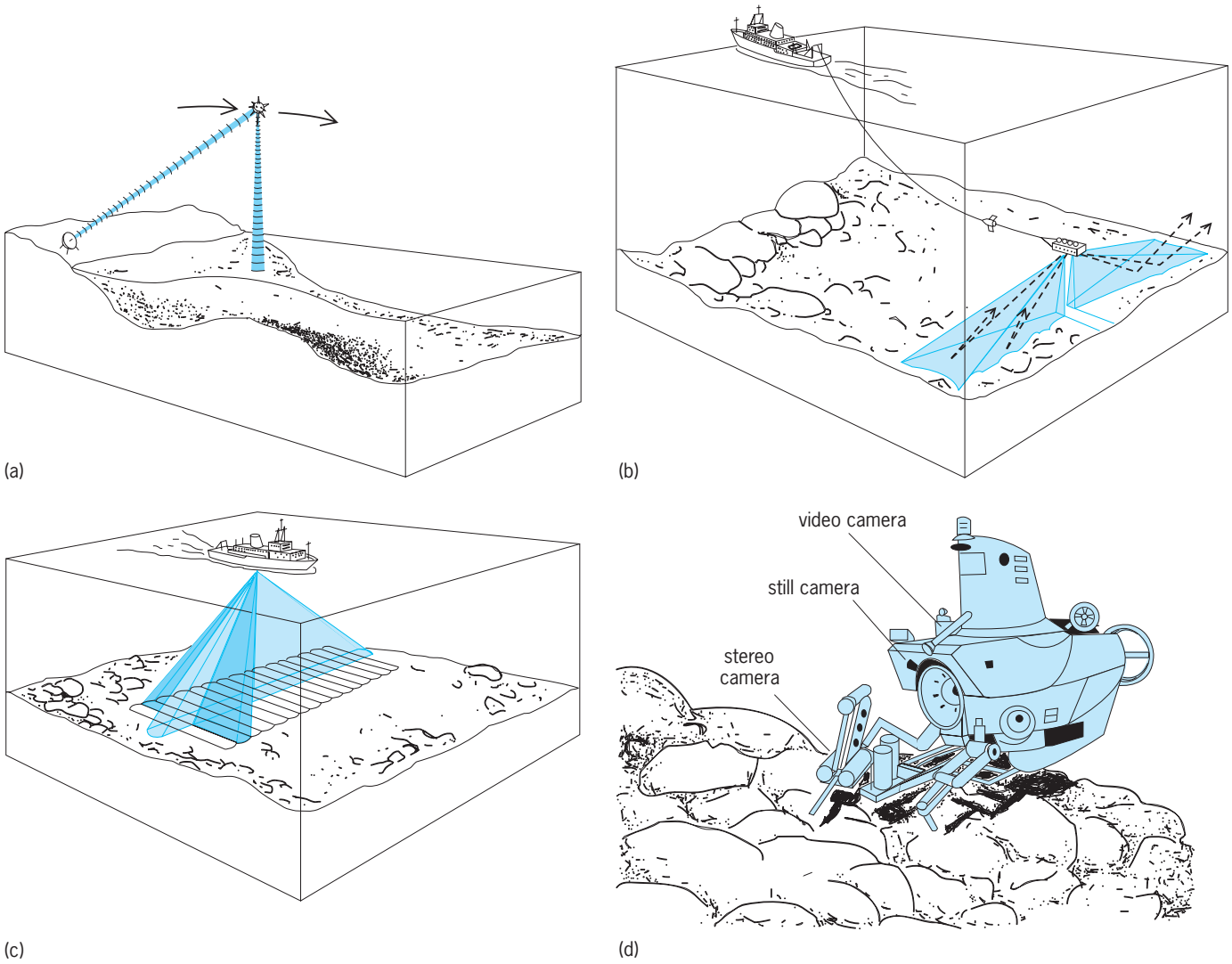


Fig. 1. Technologies that have revolutionized sea-floor mapping: (*a*) satellite altimetry; (*b*) side-scan sonar; (*c*) multibeam sonar; (*d*) submersibles. *(Courtesy of Roberto Osti)*

echo back to the sled are then recorded by the sonar in their arrival sequence, with echoes from points farther away arriving successively later. The sonars repeat this sequence of "talking" and listening every few seconds as the sled is pulled through the water so that consecutive recordings build up a continuous swath of sea-floor reflections, which provide information about the texture of the sea floor. For example, if the sea floor is flat and smooth, none of the sound from the sonars is reflected back to the instrument, just as a mirror does not reflect back light shown on it at an angle. But if the sea floor has a roughness, the sound will be reflected in multiple directions; and similar to light reflecting off broken glass, some of the reflections return to the sonar. By equating the amplitude of the returning echoes to different levels of gray, and the time that the echoes arrive back at the sled to distance, an image is obtained of sea-floor textures that looks like a black and white photograph. The textures in the image reflect a combination of sea-floor morphology and bottom sediment type; but like a single photograph, they cannot be used to determine sea-floor depths. *See* ECHO SOUNDER; SONAR; UNDERWATER SOUND.

Multibeam sonar. The best technology for mapping sea-floor depths or bathymetry is multibeam sonar (Fig. 1*c*). These systems employ a series of sound sources and listening devices that are mounted on the hull of a survey ship. As with side-scan sonar, every few seconds the sound sources emit a burst that insonifies a long, slim strip of the sea floor aligned perpendicular to the ship's direction. The listening devices then begin recording sounds from within a fan of narrow sea-floor corridors that are aligned parallel to the ship and that cross the insonified strip. Thus, sound reflections received at the listening devices are only from relatively small sea-floor regions where the insonified strip and listening corridors intersect. The timing of these reflections yields a profile of sea-floor depth, and successive profiles collected as the ship moves build up a continuous swath of sea-floor bathymetry along the ship's track. By running the survey the same way that one mows

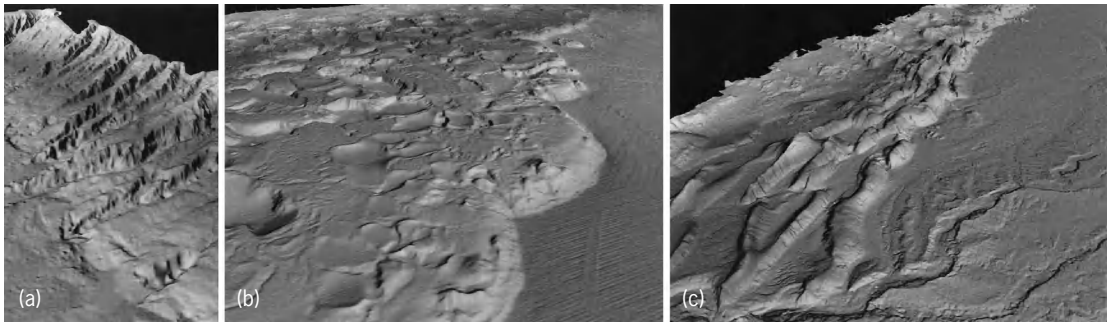


Fig. 2. In the original images, colors are used to portray sea-floor depth. Here, white represents the shallowest depths and dark gray the deepest; land is represented by black. (*a*) New Jersey continental slope. (*b*) Louisiana continental slope. (*c*) Oregon continental slope.

a lawn, adjacent swaths are collected parallel to one another to produce a complete sea-floor map of an area. *See* ACOUSTIC SIGNAL PROCESSING.

Direct visual imaging. The most modern swath mapping systems now collect both bathymetry and side-scan sonar imagery. But the most accurate and detailed view of the sea floor is provided by direct visual imaging through bottom cameras, submersibles (Fig. 1*d*), remotely operated vehicles, or if the waters are not too deep, scuba diving. Of these, bottom cameras were the first widely used means of getting an upclose view of the sea floor. However, because light is scattered and absorbed in waters greater than about 33 ft (10 m) deep, the sea-floor area that bottom cameras can image is no more than a few meters. This limitation has been partly overcome by deep-sea submersibles and remotely operated vehicles, which provide researchers with the opportunity to explore the sea floor close-up for hours to weeks at a time. But even the sea-floor coverage that can be achieved with these devices is greatly restricted relative to side-scan sonar, multibeam sonar, and satellite altimetry. Hence, with each technology there is a trade-off in sea-floor coverage with resolution, and which technology is used depends on what information is needed.

Sea-floor panoramas. Diverse sea-floor terrains, covering some 70% of the Earth's surface, are concealed from view by the oceans. Images have been produced that show the ocean stripped away and the sea floor laid bare. This has been accomplished with the use of computers by mathematically fitting a continuous surface to numerous, closely spaced, high-resolution depth measurements to render a detailed model of what the sea floor in various regions looks like (**Fig. 2**).

A wide range of submarine terrains border the continental United States. Continental slopes are where the sea floor drops at a relatively steep angle (on average, 2–6°) from the edge of the shallow shelf plateaus (on average, <400 ft or 120 m water depth) that rim the continents down to the gradually dipping (≤0.5°) rise and abyssal plains that underlie the deep sea (generally, >2 mi or 3000 m water depth). The continental slope off New Jersey (Fig. 2*a*) is scarred by numerous submarine canyons. A primary agent in cutting the canyons has been subsea avalanches of shelf-edge and slope sediments that became unstable and failed.

On the continental slope off Louisiana (Fig. 2*b*), the movement of buried salt has deformed the sea floor to create a lunarlike surface. The salt was deposited earlier in geologic time when the Gulf of Mexico was first forming. It became buried by the massive amounts of sediments eroded from North America and dumped into the Gulf of Mexico by the Mississippi River. Under the great weight of these sediments, pockets of salt are being squeezed upward and seaward, rearranging the seascape above them like groundhogs burrowing through soil.

On the continental slope off Oregon (Fig. 2*c*), the North American plate, which is one of the 12 lithospheric plates forming the Earth's outer shell and in which the United States is embedded, is overriding the Juan de Fuca plate, which lies beneath flat sea-floor sediments extending from the base of the slope. As the plates collide, the overthrusting North American plate is bulldozing the sediments off the downgoing Juan de Fuca plate and folding them into the ridges that run parallel to the continental slope.

Lincoln F. Pratson

Bibliography. L. F. Pratson and W. F. Haxby, Panoramas of the seafloor, *Sci. Amer.*, 276(6):66–71, 1997; L. F. Pratson and W. F. Haxby, What is the slope of the U.S. continental slope?, *Geology*, 24(1):3–6, 1996; D. T. Sandwell, Geophysical applications of satellite altimetry, *Rev. Geophys. Suppl.*, pp. 132–137, 1990; R. Vogt and B. E. Tucholke (eds.), *Imaging the Ocean Floor: History and State of the Art—Geology of North America: The Western North Atlantic Region*, 1986.

Sea ice

Ice formed by the freezing of seawater. Ice in the sea includes sea ice, river ice, and land ice. Land ice is principally icebergs which are prominent in some areas, such as the Ross Sea and Baffin Bay. River ice is carried into the sea during spring breakup and is important only near river mouths. The greatest part, probably 99% of ice in the sea, is sea ice. *See* ICEBERG.

Properties. The freezing point temperature and the temperature of maximum density of seawater

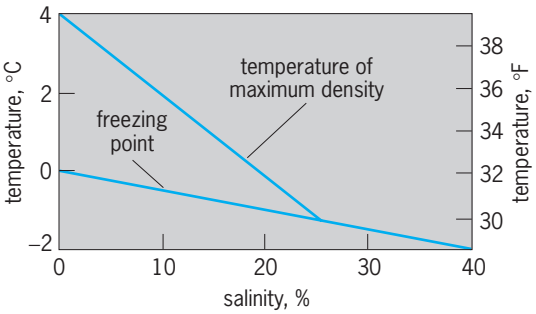


Fig. 1. Change of freezing point and temperature of maximum density with varying salinity of seawater.

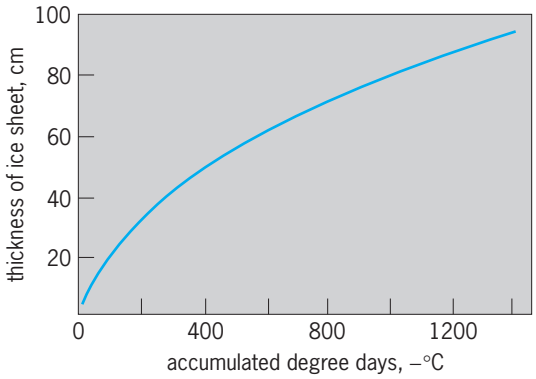


Fig. 2. Growth of undisturbed ice sheet. °F = (°C × 1.8) + 32. 1 cm = 0.4 in.

vary with salinity (**Fig. 1**). When freezing occurs, small flat plates of pure ice freeze out of solution to form a network which entraps brine in layers of cells. As the temperature decreases more water freezes out of the brine cells, further concentrating the remaining brine so that the freezing point of the brine equals the temperature of the surrounding pure ice structure. The brine is a complex solution of many ions. With lowering of temperature below −8°C (18°F), sodium sulfate decahydrate ($Na_2SO_4 \cdot 10H_2O$) and calcium sulfate dihydrate ($CaSO_4 \cdot 2H_2O$) are precipitated. Beginning at −24°C (−11°F), sodium chloride dihydrate ($NaCl \cdot 2H_2O$) is precipitated, followed by precipitation of potassium chloride (KCl) and magnesium chloride dodecahydrate ($MgCl \cdot 12H_2O$) at −34°C (−29°F), and the remaining ions with further lowering of temperature.

The brine cells migrate and change size with changes in temperature and pressure. The general downward migration of brine cells through the ice sheet leads to freshening of the top layers to near zero salinity by late summer. During winter the top surface temperature closely follows the air temperature, whereas the temperature of the underside remains at freezing point, corresponding to the salinity of water in contact. Heat flux up through the ice permits freezing at the underside. In summer freezing can also take place under sea ice in regions where complete melting does not occur. Surface melt water (temperature 0°C or 32°F) runs down through cracks in the ice to spread out underneath and contact the still cold ice masses and underlying colder seawater. Soft slush ice forms with large cells of entrapped seawater which then solidifies the following winter.

The salinity of recently formed sea ice depends on rate of freezing; thus sea ice formed at −10°C (14°F) has a salinity from 4 to 6 parts per thousand (‰), whereas that formed at −40°C (−40°F) may have a salinity from 10 to 15‰. Sea ice is a poor conductor of heat and the rate of ice formation drops appreciably after 4–6 in. (10–15 cm) are formed. An undisturbed sheet grows in relation to accumulated degree-days of frost. **Figure 2** shows an empirical relation between ice thickness and the sum of the mean diurnal negative air temperature (degrees Celsius). The thermal conductivity varies greatly with the air bubble content, ranging perhaps between 1.5 and 5.0×10^{-3} cal/(cm)(s)(°C) [0.6 and 2.0 J/(m)(s)(°C)].

The specific gravity of sea ice varies between 0.85 and 0.95, depending on the amount of entrapped air bubbles. The specific heat varies greatly because changing temperature involves freezing or melting of

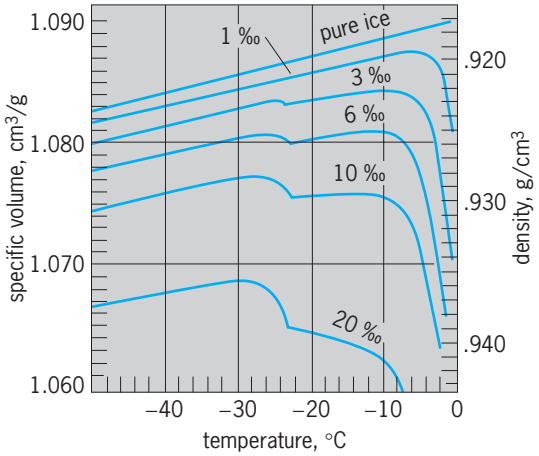


Fig. 3. Specific volume of sea ice for varying salinity and temperature, computed on basis of chemical model. °F = (°C × 1.8) + 32; 1 cm³/g = 1.730 in.³/oz; 1 g/cm³ = 0.578 oz/in.³ (*After D. L. Anderson, based on data in Arctic Sea Ice, NAS-NRC Publ. 598, 1958*)

Fig. 4. Open water in ice field. (*U.S. Navy*)

Fig. 5. Pressure ridge formation. (*U.S. Navy*)

ice. Near 0°C (32°F), amounts that freeze or melt at slight change of temperature are large and "specific heat" is anomalous. At low temperatures the value approaches that of pure ice; thus, specific heat for 4‰ saline ice is 4.6 cal/(g)(°C) [19 kJ/ (kg)(°C) at −2°C and 0.6 [2.5] at −14°C; for 8‰ saline ice, 8.8 [37] at −2°C and 0.6 [2.5] at −14°C.

High-saline sea ice may expand when cooled because further freezing out occurs with an increase of specific volume, for example, ice of salinity 8‰ at −2°C expands at a rate of about 93×10^{-4} cm³/g per degree Celsius decrease in temperature, at −14°C expands 0.1×10^{-4}, but at −20°C contracts 0.4×10^{-4} per degree Celsius decrease. Change of specific volume with temperature and salinity is shown in **Fig. 3**.

Sea ice is viscoelastic. Its brine content, which is very sensitive to temperature and to air bubble content, causes the elasticity to vary widely. Young's modulus measured by dynamic methods varies from 5.5×10^{10} dynes/cm² (Pa) during autumn freezing to 7.3×10^{10} dynes/cm² (Pa) at spring breakup. Static tests give much smaller values, as low as 0.2×10^{10} dynes/cm² (Pa). The flexural strength varies between 0.5 and 17.3 kg/cm² over salinity range of 7–16‰ and temperatures −2 to −19°C. Acoustic properties are highly variable, depending principally on the size and distribution of entrapped air bubbles.

Electrical properties vary greatly with frequency because of ionic migration within the brine cells. For example, for sea salinity of 10‰ at −22°C, the dielectric coefficient is very large, about 10^6 at 20 Hz, and decreases with increasing frequency to about 10^3 at 10 kHz and to 10, or less, at 50 MHz. The effective electrical conductivity decreases with

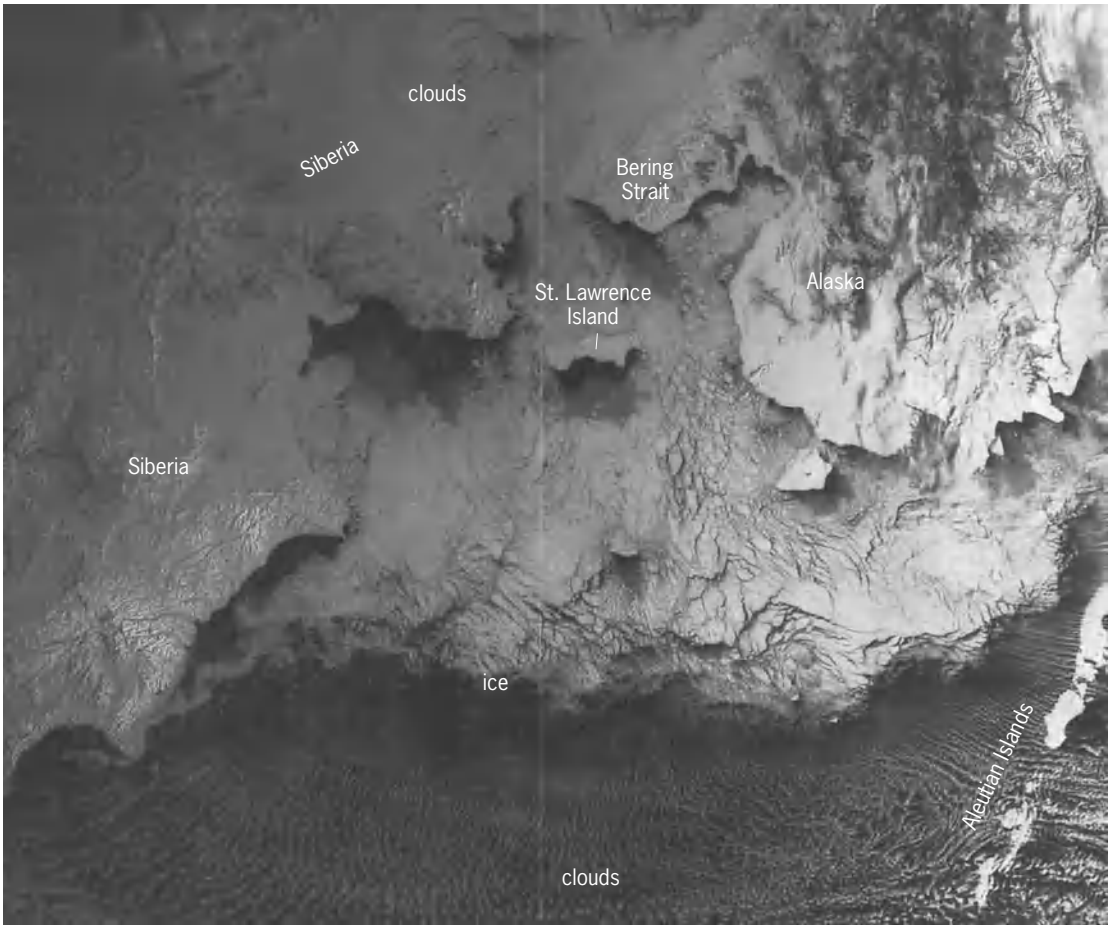


Fig. 6. Enhanced infrared image of Bering Sea by satellite, December 1976. (*Environmental Products Branch*, *NOAA*)

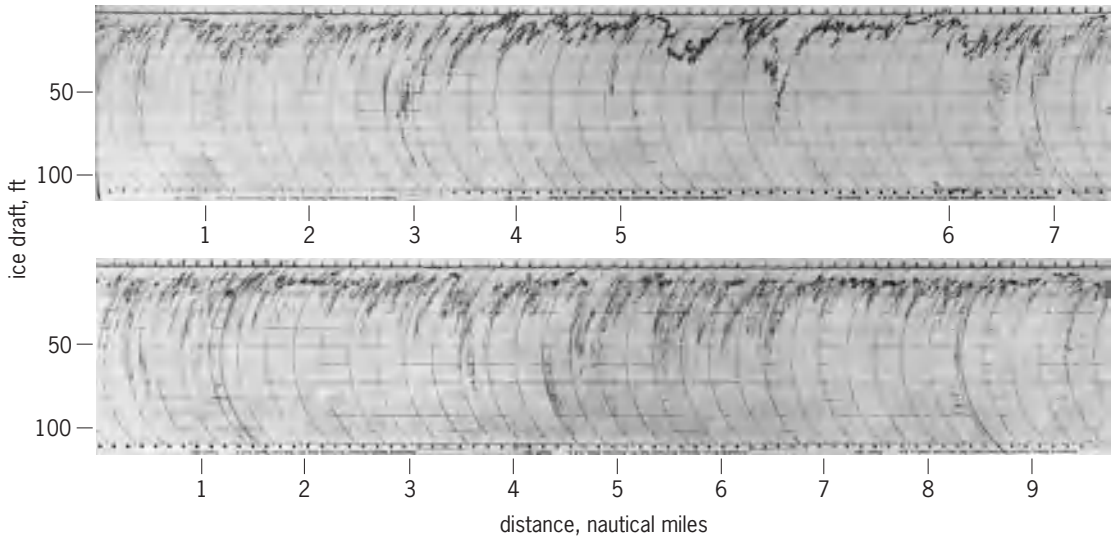


Fig. 7. Underside profile records of sea ice by submarine. Elongation of the distance scale between mile 5 and mile 6 on the upper record is due to the slower speed of the submarine during this portion of the observation. 1 nautical mile = 1.8 km; 1 ft = 0. 3 m. (*U.S. Navy*)

lowering of temperature, for example, from less than 10^{-3} mho/cm at −5°C to 10^{-6} mho/cm at −50°C (frequency 1 to 10 kHz).

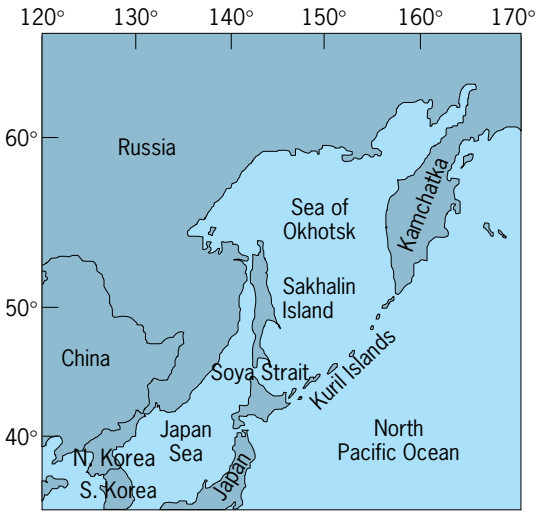
Extent and observation. The sea ice in any locality is commonly a mixture of recently formed ice and old ice which has survived one or more summers. Except in sheltered bays, sea ice is continually in motion because of wind and current. The weaker parts of the sea ice canopy break when overstressed in tension, compression, or shear, pulling apart to form a lead (open water) or piling block on block to form a pressure ridge (**Figs. 4** and **5**). Ridges may form in ice of any thickness, from thin sheets (10 cm or 4 in. in thickness) to heavy blocks (3 m or 10 ft or more in thickness). The ridges may pile 13 m (43 ft) high above and extend 50 m (164 ft) below the sea surface. Massive ridges become grounded in coastal zones. The extent, movement, and type of sea ice are routinely observed by satellite or aircraft-carried sensor systems. **Figure 6** is an infrared photograph of ice covering the Bering Sea. The underside can be observed from a submarine transiting under the sea ice and using a sonar to record the draft profile of the ice canopy (**Fig. 7**). Waldo Lyon

Bibliography. J. C. Comiso, Sea ice microwave emissivities from satellite passive microwave and infrared observations, *J. Geophys. Res.*, vol. 88, no. C12, September 1983; *Proceedings of the Conference on Sea Ice*, NRC, Iceland, 1972; Symposium on Applied Glaciology, *J. Glaciol.*, vol. 19, no. 81, 1977; Symposium on Remote Sensing in Glaciology, *J. Glaciol.*, vol. 15, no. 73, 1975; P. Wadhams, *Ice Characteristics in the Seasonal Ice Zone*, *Cold Regions Science and Technology*, vol. 2, 1980.

Sea of Okhotsk

A semienclosed basin adjacent to the North Pacific Ocean, bounded on the north, east, and west by continental Russia, the Kamchatka Peninsula, and northern Japan (see **illus.**). On its southeast side the Sea of Okhotsk is connected to the North Pacific via a number of straits and passages through the Kuril Islands. The sea covers a surface area of approximately 590,000 mi^2 (1.5 million km^2), or about 1% of the total area of the Pacific, and has a maximum depth of over 9000 ft (3000 m). Its mean depth is about 2500 ft (830 m). Owing to the cold, wintertime Arctic winds that blow to the southeast, from Russia toward the North Pacific Ocean, the Sea of Okhotsk is partially covered with ice during the winter months, from November through April. *See* BASIN; PACIFIC OCEAN.

The basic circulation of surface water in the Sea of Okhotsk is thought to be generally counterclockwise, with water entering the basin south of Kamchatka through passages in the Kuril Islands, flowing north, west, and south through the sea, and exiting through other passages in the islands, roughly halfway between northern Japan and Kamchatka. Since most of the sills in the passages of the Kuril



Sea of Okhotsk and adjacent landmasses.

Islands are shallower than about 900 ft (300 m), there is a limit to the amount of water that can be exchanged with the North Pacific. Below the sill depth, the water of the Sea of Okhotsk is trapped and can leave only via very slow vertical mixing with the surface waters.

The amount of water exchanged between the Sea of Okhotsk and the North Pacific Ocean is not well known, but it is thought that the waters of the Sea of Okhotsk that do enter the North Pacific may play an important role in the Pacific's large-scale circulation. The reason is the extreme winter conditions over the Sea of Okhotsk: its waters are generally colder and have a lower salinity than the waters at the same density in the North Pacific. It has been suggested that this flow from the Sea of Okhotsk into the North Pacific is the source of the large-scale subsurface tongue of low-salinity water at depths of 600–1800 ft (200–600 m) that can be seen over much of the subtropical North Pacific, from Asia to nearly the North American continent and south into the near-equatorial regions of the Pacific Ocean. *See* OCEAN CIRCULATION; SEAWATER.

The Sea of Okhotsk was generally off limits for scientific studies or commercial ventures from the 1940s through the mid-1990s. In recent years, the Russian government has relaxed restrictions on access to the sea. As a result, commercial fishing of salmon and pollack in the region by several nations is increasing rapidly, and offshore oil drilling has recently begun in the area just east of Sakhalin Island. A number of scientific studies of the Sea of Okhotsk have been undertaken that should help to ensure that these abundant natural resources are managed wisely. Since 1998, a joint Japanese, Russian, and United States observation program has examined some of the more important scientific questions pertaining to the interaction of ice, the atmosphere, and the ocean in the region. The new measurements suggest that tidal flows through the straits in the Kuril Islands are as strong as anywhere in the world, providing a source for the transport of the Sea of Okhotsk's properties into the North Pacific. It also appears that there is a great deal of year-to-year variability in ice production of in the Sea of Okhotsk that is closely related to regional changes in the state of the atmosphere. With limited exploration of the Sea of Okhotsk now permitted, it is expected that scientific knowledge of the region will continue to grow. *See* TIDE.

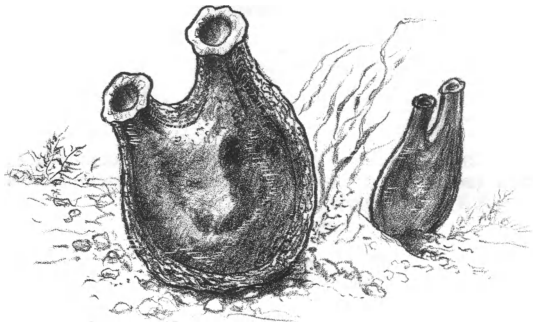
Stephen C. Riser

Bibliography. S. Martin, R. Drucker, and K. Yamashita, The production of ice and dense shelf water in the Okhotsk Sea polynyas, *J. Geophys. Res.*, 103:27,771–27,782, 1998; I. Polyakov and S. Martin, Interaction of the Okhotsk Sea diurnal tides with the Kashevarov Bank polynya, *J. Geophys. Res.*, 105:3281–3294, 2000; J. L. Reid, Jr., *Northwest Pacific Ocean Waters in Winter*, Johns Hopkins University Press, 1973; L. D. Talley, Distribution and formation of North Pacific Intermediate Water, *J. Phys. Oceanog.*, 23:517–537, 1993; Y. You et al., Roles of the Okhotsk Sea and Gulf of Alaska in forming the North Pacific Intermediate Water, *J. Geophys. Res.*, 105:3253–3280, 2000.

Sea squirt

A sessile tunicate of the class Ascidiacea (phylum Chordata), so named because water is squirted from two openings when the animal is touched. Sea squirts may be solitary, colonial, or compound, but always have a permanent enclosing structure, the test or tunic, composed of polysaccharide material structurally similar to cellulose. In colonies, individuals are loosely attached together; in compound ascidians the individual zooids are embedded in a common test.

The individual is roughly cylindrical in shape and attached at the base (see **illus.**). The test is lined by the mantle or body wall, a structure containing blood vessels and muscle fibers. The mantle encloses a space or atrial cavity. Suspended within the atrial cavity is a pharynx or branchial sac perforated by numerous oval slits (stigmata), each one bearing a ring of marginal cilia. Constant beating of the cilia draws in water through the inhalant or oral aperture into the branchial sac. Water passes through the stigmata into the atrial cavity and is discharged at the exhalant or atrial aperture. In the branchial sac, food particles from the water are filtered on a mucous sheet which is then rolled up and passed to the esophagus. The anus is located in the atrial cavity below the exhalant aperture so that feces are ejected with the exhalant water current (as are renal and reproductive products).



Sea squirts; these saclike tunicates serve as scavengers and as food for higher forms.

Sea squirts are dioecious, and fertilization may be internal or external. In some the egg develops into a tadpole which swims briefly before attaching by the anterior and metamorphosing into an adult; in others this free-swimming larval stage is suppressed.

Sea squirts occur in all seas and at all depths in the ocean. Most require a hard substratum on which to settle, but a few live in sand or mud. They are particularly common in shallow-water environments such as rocky shores, piers, and boat hulls, and in mangrove lagoons. They thrive best where there is a large amount of suspended particulate matter in the water to provide a source of food. *See* ASCIDIACEA; TUNICATA.

Ivan Goodbody

Bibliography. S. P. Parker (ed.), *Synopsis and Classification of Living Organisms*, 2 vols., 1988; H. H. Plough, *Sea Squirts of the Atlantic Continental Shelf from Maine to Texas*, 1978.

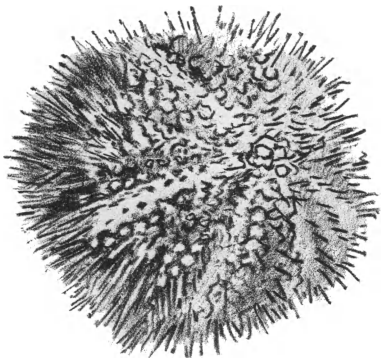
Sea urchin

A marine echinoderm of the class Echinoidea. These invertebrates are found commonly in shallow waters. The soft internal organs are enclosed in and protected by a test or shell (see **illus.**) consisting of a number of plates which fit closely together and are located under the skin. The oral surface is in contact with the substratum. Five teeth, located in the mouth, form part of Aristotle's lantern, a complex chewing structure. Average size for a sea urchin is about 2–3 in. (5–8 cm) in diameter; however, a specimen of *Sperosoma giganteum* from Japan measured 1 ft (30 cm) in diameter. Many species burrow into the sand, while others move into rock crevices to protect themselves from severe tidal action.

Like many other species of echinoderms, sea urchins use tube feet for locomotion and pincerlike structures called pedicellariae to keep the shell clean. In those species with long spines, tube feet serve a tactile function and the spines are used as locomotary structures. Sea urchins feed on most available animal and vegetable materials. The edible sea urchin from the Mediterranean (*Echinus esculentus*) is able to climb by using the tube feet as suction cups. The turban urchin (*Diadema antillarum*), a tropical species found in West Indian waters, has long, slender, and needle-sharp poisonous spines. These serve as defensive structures and can be moved in the direction of any threatening object "seen" by the numerous eye spots on the body surface. *See* EYE (INVERTEBRATE).

Fertilization is external. The larvae that develop are known as echinoplutei, bilaterally symmetrical organisms. The echinopluteus undergoes a transformation to radial symmetry during metamorphosis. When sea urchins are injured, regeneration occurs; this is especially true of the test. Their regenerative ability does not appear to be as great as that of starfishes. Severe injury to a mature animal containing reproductive cells has been advanced as a cause for what is termed a breeding epidemic. It is postulated that under such conditions a hormone is liberated which induces or stimulates adjacent animals to reproduce. Since sea urchins have the ability to produce large numbers of reproductive cells, they are used experimentally in developmental studies; the eggs have been sent into outer space to study the effects of that environment on development. *See* DEVELOPMENTAL BIOLOGY; ECHINODERMATA; ECHINOIDEA; INVERTEBRATE EMBRYOLOGY; REGENERATION (BIOLOGY).

Charles B. Curtin



The sea urchin shell is covered with spines arranged in five broad areas that are separated by narrow unprotected areas.

Seaborgium

A chemical element, symbol Sg, atomic number 106. Seaborgium has chemical properties similar to tungsten. It was synthesized and identified in 1974. This discovery of seaborgium took place nearly simultaneously in two nuclear laboratories, the Lawrence Berkeley Laboratory at the University of California and the Joint Institute for Nuclear Research at Dubna in Russia.

1												13	14	15	16	17	18
1 H	2																2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

The Berkeley group, under the leadership of A. Ghiorso, used as its source of heavy ions the SuperHeavy Ion Linear Accelerator (SuperHILAC). The production and positive identification of the isotope of seaborgium with the mass number 263 decays with a half-life of 0.9 + 0.2 s by the emission of alpha particles of principal energy 9.06 + 0.04 MeV. This isotope is produced in the reaction in which four neutrons are emitted: $^{249}Cf(^{18}O,4n)$.

The Dubna group, under the leadership of G. N. Flerov and Y. T. Oganessian, produced its heavy ions with a heavy-ion cyclotron. They found a product that decays by the spontaneous fission mechanism with the very short half-life of 7 ms. They assigned it to the isotope ^{259}Sg, suggesting reactions in which two or three neutrons are emitted: $^{207}Pb(^{54}Cr,2n)$ and $^{208}Pb(Cr^{54},3n)$. *See* NOBELIUM; NUCLEAR CHEMISTRY; PERIODIC TABLE; TRANSURANIUM ELEMENTS.

Glenn T. Seaborg

Bibliography. S. Hofmann, *On Beyond Uranium: Journey to the End of the Periodic Table*, 2002; G. T. Seaborg (ed.), *Transuranium Elements: Products of Modern Alchemy*, 1978; G. T. Seaborg and W. D. Loveland, *The Elements Beyond Uranium*, 1990.

Seamount and guyot

A seamount is a mountain that rises from the ocean floor; a submerged flat-topped seamount is termed a guyot. By arbitrary definition, seamounts must be at least 3000 ft (about 900 m) high, but in fact there is a continuum of smaller undersea mounts, down to heights of only about 300 ft (100 m). Some seamounts are high enough temporarily to form oceanic islands, which ultimately subside beneath sea level. There are on the order of 10,000 seamounts in the world ocean, arranged in chains (for example, the Hawaiian chain in the North Pacific) or as isolated features. In some chains, seamounts are packed closely to form ridges (for example, the Walvis Ridge in the South Atlantic). Very large oceanic volcanic constructions, hundreds of kilometers across, are called oceanic plateaus (for example, the Manihiki Plateau in the South Pacific). *See* OCEANIC ISLANDS.

Leaving aside small tectonic fragments of continents isolated in oceanic surroundings (for example, the Seychelles in the Indian Ocean), almost all seamounts are the result of submarine volcanism, and most are built within less than about 1 million years. Seamounts are made by extrusion of lavas piped upward in stages from sources within the Earth's mantle to vents on the seafloor. The lavas in most seamounts within the oceanic basins are composed mainly of the tholeiitic variety of basalt, but lavas are commonly of more alkalic varieties in later stages of eruption. The basaltic lavas generally erupt quietly, but the rapid escape of gases dissolved in the magmas may fragment the cooling lavas, forming pyroclastic (ashy and blocky) deposits. Pyroclastics become increasingly important as the seamount nears sea level, where confining hydrostatic pressures diminish. Seamounts closely associated with island arcs are commonly andesitic rather than basaltic, and these commonly are more explosive than basaltic volcanoes. *See* BASALT; MARINE GEOLOGY; VOLCANO.

Seamounts provide data on movements of tectonic plates on which they ride, and on the rheology of the underlying lithosphere. The trend of a seamount chain traces the direction of motion of the lithospheric plate over a more or less fixed heat source in the underlying asthenosphere part of the Earth's mantle. During cooling of the lavas, the magnetic minerals are magnetized in the direction of the Earth's magnetic field, and this direction is preserved in the rocks. From magnetic surveys or from oriented samples of seamount basalt, the paleolatitude of a seamount can be determined. The weight of a seamount flexes the underlying lithosphere, with the result that seamounts and seamount chains are generally flanked by a moatlike depression that may be in turn bordered by a broad arch. The form of the moat and arch depends on the rigidity of the lithosphere at the time of construction of the volcano: the older the lithosphere, the more rigid. *See* LITHOSPHERE; PLATE TECTONICS.

Many smaller seamounts are flat topped, possibly as a result of summit collapse, but larger seamounts tend to be more or less conical. As a seamount grows upward, some lavas, rather than being extruded from the summit of the volcano, are emplaced laterally along radial rift zones. Volcano slopes are built by lava flows, modified by undersea avalanching and sliding, which may alter the original conical form of a seamount to a starfishlike ground plan, with arms along the rift zones. Typical side slopes range about 15–30°.

Erosion of the subaerial summit region of an island seamount, or the accumulation of shallow-water reefal deposits of calcium carbonate atop a tropical seamount during subsidence, result in a nearly flat top. Termed guyots, some that were at sea level 100 million years ago have summits that are at present at water depths of about 1 mi (1500 m). *See* MARINE GEOLOGY; VOLCANO.

Edward L. Winterer

Bibliography. H. Aoki and S. Iizuka (eds.), *Volcanoes and Tectonosphere*, 1976; E. L. Winterer, D. M. Hussong, and R. W. Decker (eds.), *The Eastern Pacific Ocean and Hawaii*, 1989.

Seaplane

An airplane capable of navigating on, taking off from, and alighting upon the surface of water. Seaplanes are grouped into two main types: flying boats and float planes. In the flying boat, the hull, which provides buoyancy and planing area, is an integral part of the airframe, a specially designed fuselage which supports the wings and tail surfaces and houses the crew, equipment, and cargo. Although multihull flying boats have been built, modern use is confined to single-hull boats with lateral stability on the water provided by small floats or pontoons attached to the wings (**Fig. 1**). The float plane is a standard landplane made capable of water operation by the addition of floats which are attached to the airframe by struts. In practice the twin float is used exclusively, lateral stability on the water being provided by the separation of the two identical floats (**Fig. 2**). A seaplane with retracting wheels which permit either land or water operation is known as an amphibian. *See* AIRFRAME; FUSELAGE.

Takeoff dynamics. At rest or at low water speeds, the seaplane is supported by buoyancy and acts as a displacement vessel. Directional control in this

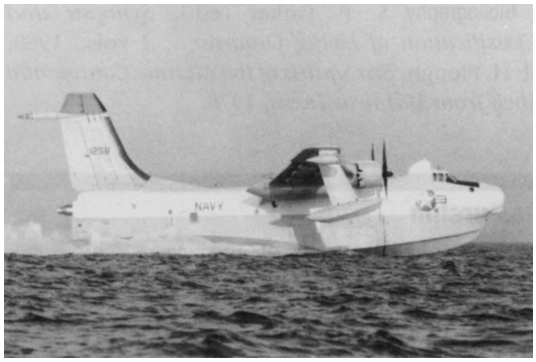


Fig. 1. Martin P5 Marlin twin-engine flying boat. (*Martin Marietta Corp.*)

Fig. 2. Single-engine float plane.

regime is achieved by water rudders or by asymmetric engine thrust in multiengine craft. As speed increases, hydrodynamic forces act on the bottom of the hull or floats, and the seaplane is supported by a combination of aerodynamic and hydrodynamic lift forces. The speed regime where the seaplane is supported by dynamic force rather than by buoyancy is known as the planing regime. While planing, hydrodynamic drag is a function of the apparent weight of the seaplane in relation to the water. Since aerodynamic lift supports more of the weight with increasing speed, the hydrodynamic drag decreases, allowing the seaplane to accelerate to flying speed. The water speed of maximum drag is known as the hump speed, and is the speed where the seaplane goes "on the step" or changes from a displacement to a planing body. Directional control in the planing regime is through the seaplane's aerodynamic control surfaces. *See* AERODYNAMICS; BUOYANCY; HYDRODYNAMICS.

Hull or float design. While the hull or float is designed to be as perfect a streamlined body as possible, certain deviations are necessary to ensure satisfactory planing action. These deviations are required because water exerts suction on convex surfaces. Although it is impossible to avoid the use of convex surfaces near the bow, the bow is in the water only at very low speeds, its curved surfaces rising above the surface even before full planing is developed. The lines formed by the intersection of the float bottom with vertical planes parallel to the float centerline are known as buttock lines. These buttock lines must be straight for most of the float forebody length in order to avoid suction and to allow dynamic planing forces to be generated efficiently. These lines must terminate at their aft ends in discontinuities called steps, since any upward change of direction of a continuous line would cause suction. The step is usually located near the longitudinal center of gravity of the seaplane, for it limits the aft travel of the hydrodynamic center of pressure. The unbalanced moment caused by hydrodynamic forces acting through a center of pressure not coincident with the center of gravity must be overcome by the seaplane's aerodynamic controls.

Although a flat surface is the most efficient planing bottom, most seaplanes have a vee bottom, with the sharpness of the vee expressed in dead-rise angle, measured from the horizontal to the nearest leg of the vee. A vee bottom lessens the impact forces experienced during landing and takeoff. The deadrise angle selected for a particular seaplane is a compromise between planing efficiency and impact reduction and is usually in the 20–30° range.

Spray control. Control of the spray thrown up by the floats or hull is an important consideration in seaplane design, for that spray can damage the propellers or structure and impair the pilot's vision. The most common spray-control technique is to install a vertical spray strip at the chine (where the bottom meets the side of the float). The spray strip alters the direction of the spray so that it does not touch vital areas.

Takeoff and landing conditions. Seaplanes are designed primarily for operation from fairly calm or sheltered waters. Seaplane landings or takeoffs from rough water are analogous to a landplane operating from very rough ground. Since impact forces on water contact are proportional to the square of the impact velocity, the only truly successful rough-water seaplane has been the Japanese Shin Meiwa PS-1, launched in 1967, which used high-lift devices to achieve touchdown and lift-off speeds of about one-half those of a conventional aircraft.

Development. The first seaplane was flown in France in 1910 by Henri Fabre. Rapid design progress led to large multiengined flying boats by World War I. Subsequently, long-distance and international flight was dominated by the seaplane, and culminated in regular passenger service across the Pacific and Atlantic by Martin and Boeing Clippers. During World War II, thousands of seaplanes were used by all combatants for reconnaissance, rescue, and logistic support missions. Wartime research led to the ultimate large seaplane, the four-jet, 600-mi/h (270-m/s) Martin P6M Seamaster (**Fig. 3**) launched in 1955.

Use. Seaplanes suffer a number of disadvantages in relation to landplanes: greater aerodynamic drag because of the float and hull shape; greater weight because of the additional structure required; susceptibility to corrosion when in salt water; inability to operate when the water is frozen or rough; and a requirement for a high level of pilot skill. The one main advantage of the seaplane, the ability to

Fig. 3. Martin P6M Seamaster four-jet flying boat. (*Martin Marietta Corp.*)

operate from an unprepared surface, has been negated by the worldwide availability of airports. Consequently, the large seaplane has all but disappeared from commercial and military service, and its rescue function has been assumed by the long-range helicopter. *See* HELICOPTER.

Most modern seaplanes are of the small single- or twin-engine type and are used for sport or recreational flying. They still serve a valuable logistic support rule in remote areas without airports but with many lakes and rivers. *See* AIRPLANE.

Richard A. Hoffman

Bibliography. M. Allward, *An Illustrated History of Seaplanes and Flying Boats*, 1982, reprint 1990; M. Faure, *Flying a Floatplane*, 3d ed., 1996; H. R. Palmer, *The Seaplanes*, 1980; J. Stroud, *The World's Civil Marine Aircraft*, 1979.

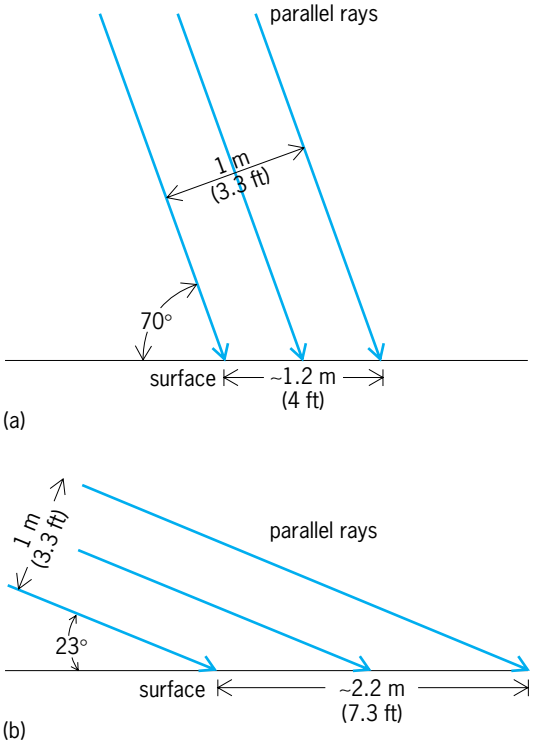
Seasons

The four divisions of the year based upon variations of sunlight intensity (solar energy per unit area at the Earth's surface) at local solar noon (noontime) and daylight period. The variations in noontime intensity and daylight period are the result of the Earth's rotational axis being tilted 23°.5 from the perpendicular to the plane of the Earth's orbit around the Sun. The direction of the Earth's axis with respect to the stars remains fixed as the Earth orbits the Sun. If the Earth's axis were not tilted from the perpendicular, there would be no variation in noontime sunlight intensity or daylight period and no seasons.

A common misconception is that the seasons are caused by variation of the Earth-Sun distance: many people think that the Earth is nearest to the Sun in summer. However, the Earth is farthest from the Sun during the first week of July (early summer in the Northern Hemisphere) and nearest to the Sun during the first week of January; also, winter and summer occur simultaneously on opposite sides of the Equator. The Earth-Sun distance does not influence the seasons because it varies only slightly, and it is overwhelmed by the effects of variations in sunlight intensity and daylight period due to the alignment of the Earth's axis.

The more hours the Sun is above the horizon, the greater the heating effect produced by the Sun. Daylight period varies for all locations except the Equator, which has 12 h of daylight every day. The poles have the widest variation in daylight period, experiencing 6 months of continuous daylight and 6 months of continuous night.

Sunlight intensity at a location depends upon the angle from the horizon to the Sun at local solar noon; this angle in turn depends upon the location's latitude and the position of the Earth in its orbit. At increased angles, a given amount of sunlight is spread over smaller surface areas, resulting in a greater concentration of solar energy, which produces increased surface heating (see **illus.**). Intensity is a more important factor than the number of daylight hours in determining the heating effect at the Earth's surface.



Relation of the angle of the Sun above the horizon to the area covered by a beam of sunlight 1 m (3.3 ft) on a side at 43° latitude. (a) The beam at solar noon on the summer solstice; the beam covers an area 1.2 m (4 ft) on a side, or about 1.4 m^2 (16 ft^2). (b) The beam at solar noon on the winter solstice; the beam covers an area 2.2 m (7.3 ft) on a side, or about 4.8 m^2 (53 ft^2). Since sunlight at noon on the winter solstice is spread over an area more than three times larger than that covered on the summer solstice, less solar heating occurs in winter than in summer.

For example, the poles receive 6 months of continuous daylight, a situation that would seem likely to produce significant heating. However, the maximum sunlight intensity at the poles is equal only to the noontime intensity on the first day of winter at 43° latitude (for example, the northern United States). The many hours of daylight at the poles provide relatively little heating effect because of the low sunlight intensity. By contrast, the maximum noontime intensity at 43° latitude equals the solar energy per unit area received at the Equator's surface in late November.

For locations on or north of the Tropic of Cancer (23°.5 north) and on or south of the Tropic of Capricorn (23°.5 south), maximum sunlight intensity and daylight period occur on the summer solstice; minimum intensity and period occur on the winter solstice. The solstices occur on about June 21 and December 21, the June solstice marking the start of summer in the Northern Hemisphere and the beginning of winter in the Southern Hemisphere. *See* EARTH ROTATION AND ORBITAL MOTION.

Harold P. Coyle

Bibliography. J. Bennett et al., *Cosmic Perspective*, 3d ed., Addison Wesley, 2004; E. Chaisson and S. McMillan, *Astronomy Today*, 5th ed., Prentice Hall, 2005; J. M. Pasachoff, *Science Explorer: Astronomy*, Prentice Hall, 2005; J. M. Pasachoff and

A. Filippenko, *The Cosmos: Astronomy in the New Millenium*, 3d ed., Brooks Cole, 2007.

Seawater

An aqueous solution of salts of a rather constant composition of elements whose presence determines the climate and makes life possible on the Earth and which constitutes the oceans, the mediterranean seas, and their embayments. The physical, chemical, biological, and geological events therein are the studies that are grouped as oceanography. Water is most often found in nature as seawater (about 98%). The rest is ice, water vapor, and fresh water. The basic properties of seawater, their distribution, the interchange of properties between sea and atmosphere or land, the transmission of energy within the sea, and the geochemical laws governing the composition of seawater and sediments are the fundamentals of oceanography. *See* HYDROSPHERE; OCEANOGRAPHY.

Physical Properties

Water is arguably the most important chemical substance in the terrestrial system, its large specific heat (the highest of almost any substance) is of extreme import to weather, climate, and the ocean. Its physical properties include a large dielectric constant, elevated boiling and freezing points, high osmotic pressure, and strong dissolving power—properties that are modified by the dissolved salts and the other materials found in the ocean.

The major chemical constituents of seawater are cations (positive ions) and anions (negative ions) [**Table 1**]. In addition, seawater contains the suspended solids, organic substances, and dissolved gases found in all natural waters. A standard salinity of 35 practical salinity units (psu; formerly parts per thousand, or ‰) has been assumed. While salinity does vary appreciably in oceanic waters, the fractional composition of salts is remarkably constant throughout the world's oceans. In addition to the dissolved salts, natural seawater contains particulates in the form of plankton and their detritus, sediments, and dissolved organic matter, all of which lend additional coloration beyond the blue coming from Rayleigh scattering by the water molecules. Almost every known natural substance is found in the ocean, mostly in minute concentrations. *See* SCATTERING OF ELECTROMAGNETIC RADIATION.

The density of seawater is an important quantity that enters sensitively into many dynamical and thermodynamical processes. An internationally accepted equation of state has been derived that allows the accurate computation of density, given measurements of salinity, temperature, and pressure. A high accuracy is required to compute the small density variations that set up horizontal pressure gradients which, together with the Earth's Coriolis force, balance large-scale geostrophic flows such as the Gulf Stream. Salinity is usually measured via electrical conductivity. *See* CORIOLIS ACCELERATION; GULF STREAM.

Seawater is slightly compressible, implying that sea level is approximately 30 m (100 ft) lower than it would be if it water were incompressible. A parameter known as the thermodynamic space is quite convenient for displaying measurements of temperature and salinity taken during vertical probes of the water column, and allows the identification of the origins of various water masses via a technique called isentropic analysis.

The density of sea ice drops to the range of 857–924 kg/m^3, depending on the amount of brine. These effects of salinity are related to the very large osmotic pressure of seawater.

Various thermodynamic quantities define other properties of seawater (**Table 2**). All of these coefficients are functions of salinity, temperature, and density, and some of their variations with pressure, for example, lead to subtle but profound effects in the sea.

Acoustics. Because of the complex temperature and salinity structure of the upper ocean, highly variable sound propagation paths occur there. The oceanic mixed layer often serves as a leaky waveguide for a sound source located in that layer; in the absence of a vertical temperature gradient, sound propagates along the mixed layer by a combination of upward refraction and surface reflection–forward scattering. However, for a source located beneath the thermocline, the colder water initially refracts propagation paths downward as the sound speed decreases. This is partially offset by the slow increase in speed due to pressure at depth, with the result that there develops a minimum in overall sound speed, the so-called deep sound channel, a secondary waveguide. In tropical and temperate seas, this minimum lies near 1200 m (4000 ft) depth with the result that detectable levels of sound may be heard across entire ocean basins. In a typical sound

TABLE 1. Major constituents of seawater (salinity 35 psu)*

Positive ions	Amount, g/kg	Negative ions	Amount, g/kg
Sodium (Na^+)	10.752	Chloride (Cl^-)	19.345
Magnesium (Mg^{2+})	1.295	Bromide (Br^-)	0.066
Potassium (K^+)	0.390	Fluoride (F^-)	0.0013
Calcium (Ca^{2+})	0.416	Sulfate (SO_4^{-})	2.701
Strontium (Sr^{2+})	0.013	Bicarbonate (HCO_3^-)	0.145
		Boron hydroxide [$B(OH)_3^-$]	0.027

* Water, 965 psu: psu dissolved materials, 35 psu.

TABLE 2. Thermodynamic properties of seawater*

Quantity	Symbol	Value
Isothermal-isohaline compressibility	a_p	4.27×10^{-10}/Pa
Specific heat at constant volume and salinity	$C_{\alpha s}$	3939 J/kg-°C
Specific heat at constant pressure and salinity	C_{ps}	3994 J/kg-°C
Isobaric-isohaline thermal expansion coefficient	a_T	2.41×10^{-4}/°C
Isobaric-isothermal saline contraction coefficient	a_s	7.45×10^{-4}/psu
Speed of sound	c	1520 m/s (3280 mi/h)
Thermal conductivity	kq	0.596 W/m-°C
Latent heat of vaporization	L_v	2.453×10^6 J/kg
Latent heat of fusion	L_f	0.335×10^6 J/kg

*All quantities are evaluated at $s = 35$ psu and $T = 20$°C (68°F), except L_f, for which $T = 0$°C (32°F).

speed profile for a midlatitude ocean, the major variability occurs in the seasonal mixed layer.

At even greater depths, the pressure continues to increase the sound speed and to bend rays upward. Near-surface rays that are launched beneath the mixed layer at shallow angles will initially penetrate into the deep ocean while undergoing refraction, until near depths of 5000 m (16,500 ft) the rays will have been refracted through the horizontal; subsequently they will bend upward toward the surface again to intersect it at distances near 60–65 km (35–40 mi) from the source. Thus, there is a focusing of sound intensity into an annular region at the surface centered on the source, the so-called first convergence zone; sound levels there may be as much as 20 decibels above those expected from nominal spreading and attenuation. Within the annular region at the surface, reflection and scattering again launch acoustic energy downward, which then goes through the same processes once again; as a result, a series of high-intensity convergence zones are developed in the form of circular annuli surrounding the source, which have radii that are integral multiples of the basic 60–65 km (35–40 mi) and widths of perhaps 5 km (3 mi).

In addition to these intensity variations due to spreading and refraction losses, there exists attenuation from absorption and volume scattering in the seawater. While the attenuation coefficient nominally varies as the square of the frequency, there are significant modifications from dissolved chemicals and acoustic scattering.

Underwater noise levels limit the detection of sound in the sea; they vary enormously with frequency, wind, rain, and shipping traffic, ranging from approximately 40 dB at 10 Hz to −30 dB at 10 kHz. At higher frequencies, receiver thermal noise is often limiting. *See* ACOUSTICS; SOUND; UNDERWATER SOUND.

Dynamical properties. The most important dynamical characteristic of the sea are waves, ocean currents, and tides; however, this discussion is limited to air-sea interaction and diffusion. *See* OCEAN WAVES; TIDE.

Air-sea interaction. The sea absorbs roughly 98% of the visible energy incident on it at high sun angles, and converts it to heat, thereby warming the upper layers of the ocean. The solar spectral irradiance is attenuated as sunlight penetrates beneath the sea surface. Under the influences of wind wave stress, turbulence, and convective overturning of the upper layers, the thermalized solar energy that is deposited into the upper layers is mixed down to depths ranging perhaps 30–150 m (100–500 ft), depending on location and season. Because of the high specific heat of seawater, this mixed-layer storage represents by far the largest heat capacity on the face of the Earth. Subsequent exchanges across the interface are governed by wind stress and air-sea temperature differences. When the air is cooler than the ocean, as occurs during cold-air outbreaks or at night, significant evaporation sets in. This results in turbulence in the atmospheric boundary layer, which causes rougher seas than when the air-sea temperature difference is positive. Values of air-sea temperature differences in excess of roughly −1°C (−1.8°F) lead to strong interchanges during such unstable conditions. When the air and sea temperatures are nearly the same, the exchanges are greatly reduced. *See* WIND STRESS.

Evaporation of water from the sea results in rising moist air, cloud formation and, upon condensation into clouds at altitude, the release of the latent heat of evaporation to the atmosphere, warming the air. Because the elevated warm air is farther from the Earth's axis of rotation than the air was at sea level, its changed angular momentum is transformed into winds via the Coriolis force. Thus the combination of latent heat release and Coriolis effect initiates the global atmospheric circulation, the wind stress on the ocean, and ultimately both the wind-driven and density-driven currents in the sea. Starting in the tropics then, the general circulation of the atmosphere is initiated via air-sea interaction. *See* ATMOSPHERIC GENERAL CIRCULATION.

The great poleward-directed systems (such as the Gulf Stream) that are found in every ocean basin transport heat out of the tropics and into temperate and polar regions. During their journey, the warm currents surrender to the atmosphere large amounts of latent and sensible heat, moisture, and momentum via the processes of evaporation, turbulent fluctuations, radiation, and conduction, in decreasing order of importance. In midlatitudes, for example, this warming of the atmosphere is responsible for maintaining the western parts of Europe at significantly

higher average temperatures than in the same latitudes in North America. Over longer times, these large-scale air-sea exchanges constitute one of the most important processes governing weather and climate dynamics.

Turbulence in the planetary boundary layer and the oceanic mixed layer is largely responsible for many of the secondary processes conditioning the air-sea interface. Surface wave generation, mixing, and exchanges of heat, momentum, and moisture all rely on turbulent diffusion for their efficacy. *See* HEAT BALANCE, TERRESTRIAL ATMOSPHERIC; OCEAN CIRCULATION.

Diffusive properties. Beyond waves and advection by currents, properties such as heat, salt, and momentum are transported by diffusive effects in the sea. Molecular diffusion coefficients are quite small; but eddylike fluctuations lead to turbulence and eddy diffusion coefficients which are orders of magnitude greater than their molecular counterparts and which depend on the scale of the fluctuations. The coefficients also differ for horizontal diffusion as compared with vertical diffusion, because buoyancy greatly inhibits transport in the vertical. The energy going into waves, currents, and diffusion of properties is ultimately removed from the system at scales near the dissipation length, that is, at about 1 mm. This fluid friction, along with bottom friction, acts to slow down both the atmosphere and ocean, but their large-scale dynamics are maintained by a continuing but unsteady input of energy from the Sun and the atmosphere. *See* DIFFUSION.

Electromagnetics. Electromagnetic and optical properties of seawater include both surface characteristics and bulk attributes.

Conductivity. The ocean is a moderately good conductor of electricity. Electrical conductivity, like the other material properties, depends not only on salinity but on temperature and pressure as well, although the dependence on pressure is quite weak. In seawater, conductivity is conventionally measured, and an algorithm is then used to convert it to salinity; a method for computing the inverse relationship is available as well. Not only is conductivity important for determination of the oceanic density field, but along with the dielectric constant, it is relevant to problems involving remote sensing of the sea surface by microwave methods.

Scattering and emission. Microwave power reflection coefficients (or reflectivities) are dependent on polarization and incidence angle. Waves on the sea modify the reflection coefficients significantly. The customary measure of reflection-scattering from a rough surface is the normalized radar cross section, relative to 1 m^2 of ocean surface. This is a function of frequency, incidence, azimuth, and reflection angles; polarization; wind speed–sea state; and surfactant coverage. Emission of thermal energy at microwave frequencies is used to determine thermodynamic temperatures, wind speed and, at lower frequencies, salinity of the sea surface. *See* ABSORPTION OF ELECTROMAGNETIC RADIATION; MICROWAVE; POLARIZED LIGHT.

Optics. The optical characteristics of seawater derive from a combination of physical, chemical, biological, and geological properties. Central to the subject are the questions of light transmission in the sea (for studies of biological productivity and underwater signaling), the color of seawater, and the use of variations in water color as a means for flow visualization in optical images of the sea.

Surface reflectivities. Surface reflectivities are due to a combination of intrinsic Fresnel reflection and rough surface scattering. The smooth-surface reflectivity is polarization sensitive and is similar to that for microwave frequencies. Since direct sunlight is unpolarized, the variation of reflectivity with incidence angle is given by the average of vertical and horizontal values. As the wind speed increases, reflection falls off.

Index of refraction. Over the nominal range of variation of wavelength, temperature, salinity, and pressure, the index of refraction does not depart from 1.34 by more than 0.015.

Volume absorption and scattering. While seawater is a reasonable conductor at microwave frequencies and below, at optical frequencies clearest seawater is a remarkably low-loss dielectric with a broad transparency band in the visible. The spectral beam absorption coefficient is a measure of the fractional absorption of a collimated beam of light in traversing 1 m of water. The addition of photosynthetic biological material profoundly alters the adsorption coefficient; chlorophyll *a* and phaeophytin *a* are the major absorbers, along with so-called yellow substance, or dissolved tanninlike compounds. Beyond absorption, the process of scattering is important, and this is described by a beam-scattering coefficient which is large in the forward direction but much smaller in the backscatter direction. Thus, the total beam attenuation coefficient is due to absorption and scattering.

Remote sensing with imaging optical spectrometers on satellites has yielded the first global estimates of near-surface chlorophyll and sediment concentrations by utilizing the concepts of ocean optics in the interpretation of multispectral images from satellites and aircraft. *See* OPTICAL DETECTORS; REMOTE SENSING; SCIENTIFIC AND APPLICATIONS SATELLITES.

John R. Apel

Chemical Properties

The complex mixture of salts and other substances found dissolved in seawater is quantitatively dominated by chloride and sodium ions, with somewhat lesser amounts of sulfate, magnesium, calcium, and other ions. The total salt content of seawater is expressed by the term salinity. This measure of the salt concentration has been rigorously defined in several ways; since 1978 it has been defined and measured according to the ratio of the electrical conductivity of a seawater sample to that of a standard solution of potassium chloride. For both technical and historical reasons, the salinity is not exactly the total mass of salts in solution, but the two values differ by only about 0.5%.

TABLE 3. Major inorganic nongaseous constituents in seawater*

Substance	g/kg	mmol/kg
Sodium ion (Na^+)	10.781	468.96
Potassium ion (K^+)	0.399	10.21
Magnesium ion (Mg^{2+})	1.284	52.83
Calcium ion (Ca^{2+})	0.4119	10.28
Strontium ion (Sr^{2+})	0.00794	0.0906
Chloride ion (Cl^-)	19.353	545.88
Sulfate ion (SO_4^{2-})	2.712	28.23
Bicarbonate ion (HCO_3^-)	0.126	2.06
Bromide ion (Br^-)	0.0673	0.844
Boron hydroxide [$B(OH)_3^-$]	0.0263	0.425
Fluoride ion (F^-)	0.00130	0.086

* Total mass concentration of above constituents, 35.170; total concentration of all other dissolved substances, ~0.03; water, ~964.80.

The average salinity of the world ocean is 34.73‰ (parts per thousand by weight, or grams per kilogram of seawater) and is remarkably constant. Only about one-tenth of all the water in the ocean departs by more than 1% from this mean value, so measurement techniques of the most exquisite sensitivity must be used to discern the real, and important, differences in salinity from place to place within the oceans. The salinity departs from this relative constancy in estuaries and other near-coastal regions where seawater is diluted with river water, in polar regions where ice melts in the spring, and in coastal lagoons or small seas where evaporation may sometimes exceed the input of fresh water by rivers and rain. Since evaporation and rainfall generally change only the total salt content and not the relative proportions of the salts, it is conventional to express the concentrations of substances in the ocean on the basis of a standard ocean water with a salinity of 35‰ (**Tables 3** and **4**).

Major and minor elements. It is also conventional to consider that the major dissolved constituents of seawater consist of 11 substances (Table 3), all present in concentrations of greater than 1 mg/kg or 1 part per million (ppm). In addition, however, the dissolved gases nitrogen and oxygen, while variable in concentration, generally amount to a total of about 10–30 mg/kg, and dissolved organic matter is present in variable concentrations of about 1 to several milligrams per kilogram.

The total concentration of salts in seawater, about 35 g in a kilogram, is great enough that the chemical activity of each ion is strongly affected by the presence of the others, and in addition numerous ion pairs and other complexes are present. These effects must be taken into account in any quantitative evaluation of the behavior of substances in seawater.

TABLE 4. Concentrations of dissolved gases in seawater*

Gas	mg/kg	μmol/kg
Nitrogen (N_2)	12.04	428.5
Oxygen (O_2)	7.48	233.7
Helium (He)	0.0000067	0.0017
Neon (Ne)	0.00014	0.0069
Argon (Ar)	0.457	11.44
Krypton (Kr)	0.00021	0.0025
Xenon (Xe)	0.000043	0.00035

*At 18°C (64°F) and salinity of 35‰, and at pressure of one standard atmosphere.

With the exceptions of oxygen and helium, the dissolved gases in seawater (Table 4) are not subject to any processes that can significantly change their concentrations once these concentrations have been established by coming to equilibrium with the atmosphere. They will not change during the circulation of that water into deeper parts of the ocean, except by mixing with other parcels. The concept of attaining equilibrium with the atmosphere is, however, a little fuzzy. It is common to assume standard atmospheric pressure, but atmospheric pressure varies. During windy conditions, breaking waves cause the entrainment of bubbles to some depth in the water; as these dissolve, they increase the concentrations, sometimes up to several percent above saturation. Observation shows that all ocean water has been affected to a measurable extent by these effects, so that they have to be taken into account when making careful calculations. Once below the surface, however, the concentrations of nitrogen and the noble gases (with the exception of helium) are not subject to further change. Helium, especially the rare isotope helium-3, is increased in deep water by its release from the mantle in places where fresh hot basalt comes in contact with seawater. Nitrogen is used and produced by biological processes, but not in amounts that can easily be measured in most of the ocean. Oxygen is both produced and consumed in biological reactions, and measurements of the concentration of oxygen are thus an important tracer of oceanic processes.

The minor elements exhibit a great range of concentrations. Most are present in only tiny fractions of a part per billion, providing a great challenge to the analyst. In many cases the limitation has been the great difficulty of collecting seawater from ships without contamination being far greater than the concentrations in the water, as well as the difficulty of obtaining reagents and equipment that do not contain more of the trace elements than the water samples themselves. It is possible to measure such low concentrations by combining great attention to ultraclean sampling and processing with modern highly sensitive instrumentation.

Nutrients. Unlike the major elements and the conservative minor elements, the important plant nutrients (phosphate, nitrate, and silicate) are not uniformly distributed in the ocean according to the salinity of the water. Instead, they exhibit marked changes from place to place, both vertically and horizontally. Characteristically, the nutrients are depleted in surface waters and are enriched at depth (**Fig. 1**). Similar and related profiles from much of the ocean have provided the general understanding of the interaction of ocean circulation and biogeochemical processes.

Small plants (the mostly unicellular photosynthetic plankton) in the sunlit surface waters use these

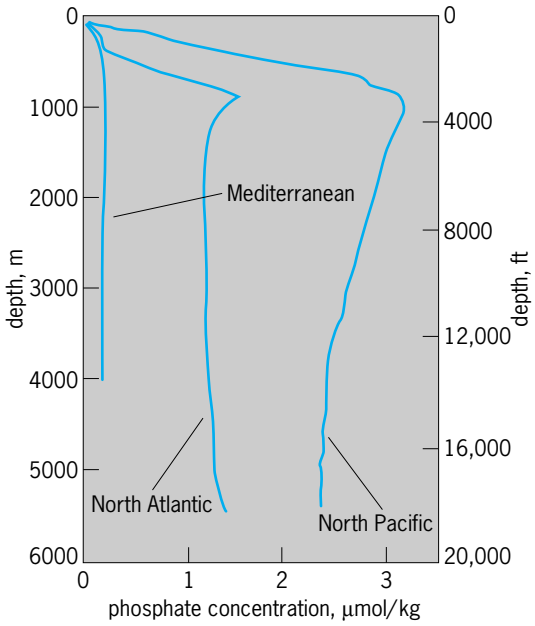


Fig. 1. Vertical distribution of phosphate (PO_4^{3-}) at three stations to show a range of characteristics found in major oceanic regions. North Atlantic is from a region in the central North Atlantic northeast of Bermuda, and North Pacific is from the central North Pacific northeast of Hawaii.

nutrients (along with carbon and smaller amounts of other nutrients) to form their body structures; these in turn provide the base of the food chain from zooplankton all the way up to the largest fish and whales. Particulate matter, including plant cells, feces, and other debris, sinks below the surface and is eaten by deeper-living plankton, fish, bacteria, and other organisms throughout the water column and on the bottom. The metabolism of these deeper living organisms results in the release and enrichment of the deeper water with carbon dioxide, phosphate, and the other nutrients, and the consumption of oxygen. *See* FOOD WEB; MARINE ECOLOGY; PHYTOPLANKTON.

The difference in the concentrations of the nutrients between the Mediterranean and the Atlantic is explained by the circulation of seawater. Surface water from the Atlantic flows into the Mediterranean, and is already depleted in nutrients. The return flow through the Strait of Gibraltar is somewhat deeper, underneath the surface inflow, so that the small accumulation of nutrients in the deeper water tends to be removed. The Pacific Ocean provides the reverse example. Deep water from the Atlantic flows into the Pacific, carrying its load of nutrients; in the Pacific it upwells, and the return flow to the Atlantic includes a considerable amount of surface water, depleted in nutrients through the activities of the biota that caused the transport of the nutrients back to the deep water in the Pacific. The nutrients therefore tend to be retained and to build up in the Pacific; the same is also true of the Indian Ocean. In many cases, more detailed analyses also provide analogous explanations of the structure seen in the vertical profiles. *See* ATLANTIC OCEAN; INDIAN OCEAN; MEDITERRANEAN SEA; PACIFIC OCEAN.

Surface water concentrations of nutrients vary considerably throughout the ocean, depending on the detailed physical processes that bring water from some depth to the surface. These upwelling regions are especially notable, for example, off the western shores of North and South America, off the southwestern shore of Africa, and in the complex current systems flowing along the Equator. Accordingly, these are regions of high biological productivity and also regions where there is a very strong transport of nutrients from the surface downward as the debris sinks. Shallow water regions tend also to be very productive, for two reasons. First, the interaction of ocean currents with these shallow regions causes turbulence that enhances the mixing of surface and deeper nutrient-rich water. Second, the sinking debris cannot go very deep and much is metabolized on the shallow bottom, so that the resulting nutrients are easily mixed into the surface water again. *See* SEAWATER FERTILITY; UPWELLING.

In general, an inverse relationship prevails between the concentration of oxygen (O_2) in the deep water of the ocean and the concentrations of phosphorus (P) as phosphate [and nitrogen (N) as nitrate] and carbon dioxide (CO_2). The reason is that, on average, the debris that sinks to depth has a generally uniform composition, and the nutrients are released in proportion to the oxygen used in metabolism of the debris. The quantitative relationship is expressed in the Redfield ratio: $175O_2$:122C:16N:1P; that is, for every 175 molecules of oxygen that are utilized, 122 atoms of carbon (as carbon dioxide), 16 atoms of nitrogen (as nitrate), and 1 atom of phosphorus (as phosphate) are released. Variations in this ratio from place to place appear to be only a few percent. Additional carbon dioxide, in the form of carbonate ion (CO_3^{2-}), is released by particles of calcium carbonate in places where these dissolve.

In addition to phosphorus, nitrogen, and carbon, organisms need a number of other substances in small amounts. For example, some metals are needed, such as iron, copper, molybdenum, zinc, and cobalt, with iron in the greatest amount. Iron is quite insoluble in seawater, however, and is much depleted in the surface water of the open ocean. There are a number of places in the ocean, generally far from land, where the other nutrients are available, but the growth of plankton is limited by the availability of iron. In these cases the surface depletions of phosphate and nitrate are less evident.

Many trace elements show vertical distributions that, to varying degrees, resemble the vertical distributions of phosphate. For those that are known to be plant nutrients, these distributions are easily explicable, as they must be carried down with the sinking debris from life at the surface. In other cases it appears that some elements may be simply adsorbed to the surfaces of sinking particles, carried down, and released at depth. The vertical profiles of concentration of some elements exhibit quite different shapes; in these cases other processes must be called upon to explain them.

Carbon dioxide. The chemistry of carbon dioxide in seawater is complicated and important, not least

because the oceans are taking up a considerable part of the carbon dioxide produced by the burning of fossil fuels and by other human activities. When carbon dioxide dissolves in seawater, it enters into a number of chemical equilibrium reactions. The equilibrium constants of all such reactions are affected by temperature and other parameters. Of particular importance are those reactions that are strongly affected by the temperature and salinity of the water and by hydrostatic pressure, and accurate knowledge of all these parameters is important in assessing the chemical state of the carbonate system. Because of the participation of the hydrogen ion (H^+) in some reactions, the concentration of this ion (expressed usually as the pH) is an important indicator of the state of the system. When the concentration of carbon dioxide increases, the pH drops; and when carbon dioxide decreases, the pH rises. Accordingly, as anthropogenic carbon enters the ocean it causes a small decrease in the pH. The effect is, so far, much smaller than the changes commonly caused by the growth of plants (increase in pH) and by the respiration of all living things (decrease in pH).

An additional complication is the precipitation of calcium carbonate by planktonic plants and animals (as well as by many types of organisms on coral reefs). The solubility of calcium carbonate is also strongly affected by temperature, salinity, and pressure. The shells and other bits of calcium carbonate formed in the plankton also sink, directly or in fecal pellets, out of surface layers and tend to dissolve at depth. In many places they dissolve completely because of the combined effects of pressure and the acidification of the deep water by respiratory carbon dioxide. The concentration of carbon dioxide in deep water is therefore increased, relative to the surface (**Fig. 2**). As with the nutrients, this concentration increases from the North Atlantic to the South Atlantic, through the South Pacific to the North Pacific. *See* CHEMICAL EQUILIBRIUM; HYDROGEN ION; PH.

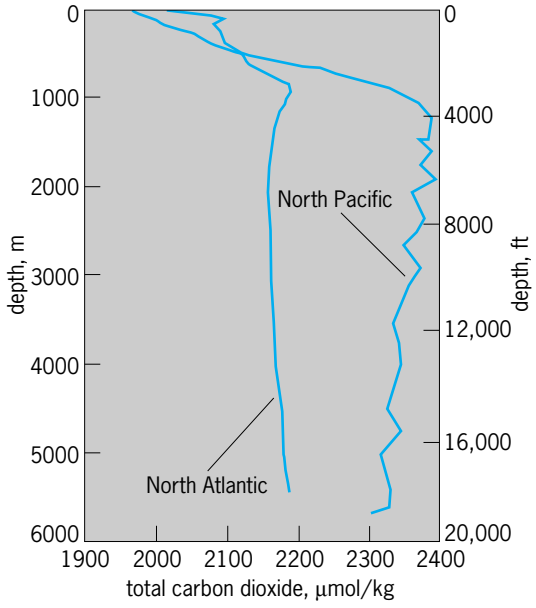


Fig. 2. Vertical distribution of total carbon dioxide concentrations at locations in the North Atlantic and North Pacific. Below depths of a few hundred meters, the distributions are generally characteristic of fairly large regions in the two oceans.

TABLE 5. Partial list of radioactive nuclides in seawater*

Isotope	Principal source
Potassium-40 (^{40}K)	Primordial
Rubidium-87 (^{87}Rb)	Primordial
Uranium-238 (^{238}U)†	Primordial
Uranium-235 (^{235}U)†	Primordial
Uranium-234 (^{234}U)†	Decay of ^{238}U
Strontium-90 (^{90}Sr)	Bomb explosions
Cesium-137 (^{137}Cs)	Bomb explosions
Carbon-14 (^{14}C)	Cosmic rays in atmosphere and bomb explosions
Hydrogen-3 [tritium] (^{3}H)	Cosmic rays in atmosphere and bomb explosions

*The total number detected is well over 40, but those not listed are present in very small concentrations.
†Nuclides of uranium that are the parents of long series of radioactive decay products.

Radioactive materials. Most radioactivity in seawater is due to nuclides present since the world began (**Table 5**). The distributions of many of the radioactive nuclides resulting from the decay of uranium isotopes and those produced by cosmic-ray-induced processes in the atmosphere, as well as those produced artificially and deposited into the ocean after nuclear explosions in the atmosphere, have been utilized to help trace mixing and transport processes in the ocean, and have helped provide quantitative understanding of the rates of these processes. *See* RADIOACTIVITY; RADIOISOTOPE; URANIUM.

Changes in composition. The overall chemical composition of the ocean results from the balance of numerous processes. These processes include (1) the weathering of exposed rocks on continents and islands, with the transport of the resulting materials to the oceans; (2) the uptake and release of various substances when seawater comes in contact with hot spots on the sea floor; (3) exchange with the atmosphere, mostly gases; (4) uptake and release of many substances by organisms; (5) sinking of particles from surface water into deep water; (6) chemical exchanges between seawater and sediments, and chemical reactions occurring within the sediments; and (7) precipitation of complex mixtures of salts in places where seawater can evaporate to dryness, a process that has varied greatly in magnitude over the course of geological time. *See* HOT SPOTS (GEOLOGY).

It should not be assumed that all these processes are constant over the span of geologic time. Known changes in climate, rates of weathering of the continents, rates of sea-floor spreading, and rearrangements of geological structures must have caused some changes in the chemistry of the ocean. Most of these changes would not likely be perceptible on the time scale of human history. However, the activities of human civilization since the beginning of the industrial revolution have resulted in detectable chemical signals throughout most of the ocean, especially in surface water.

Michael E. Q. Pilson

Distribution of Properties

The distribution of physical characteristics in the ocean is principally the result of radiation (of heat), exchange with the land (of water and salt), exchange with the atmosphere (water and heat), and mixing and stirring processes within the ocean.

Temperature. The ocean is heated by the sun, but very little of its radiation penetrates more than a few meters beneath the surface. Essentially all the heat lost by the ocean is across the air-sea interface by long-wave radiation, evaporation, and conduction. Although the heat losses balance the heat gains over an annual cycle, they seldom balance at any one location. The surface ocean gains more heat than it loses equatorward of about 35°, and it loses more heat than it gains poleward of 35°. As a consequence, the sea surface temperature is high (more than 28°C or 82°F) in equatorial regions and low (less than 1°C or 34°F) in polar regions. Below a shallow surface layer of a few hundred meters, the temperature structure of the ocean is a result of this continuing poleward transfer of the order of 10^{15} W of heat energy.

Salinity. Various dissolved solids have entered the sea from the land and have been so mixed that their relative amounts are everywhere nearly constant. For well over 95% of the ocean, the salt content is 3.43–3.51% by weight. Oceanographers routinely measure salinity to a few thousands of a percent, and salinity is reported in parts per thousand (‰). Considerable insight in oceanic processes can be achieved by examining changes in seawater salinity distributions of the order of ±0.02‰. In the middle latitudes the evaporation of water exceeds precipitation, and the surface salinity is high; in low and high latitudes precipitation exceeds evaporation, and dilution reduces the surface salinity. The Atlantic Ocean is more saline than the Pacific and Indian oceans.

Density. The density of seawater varies with temperature, salinity, and pressure. It can vary horizontally only in the presence of currents, and hence its distribution depends closely upon the structure of the flow.

Pycnocline, thermocline, and halocline. The ocean is stratified. The upper layer, in contact with the atmosphere, is warmed and cooled by exchange of heat through the surface and freshened and made more saline by rainfall and evaporation. Over the ocean this layer varies in depth from less than 100 m (330 ft) to more than 300 m (1000 ft). During summer heating, a thin cap of warmer water may be formed at the top, disappearing in fall and winter by cooling and mixing with the underlying water.

Beneath the upper layer, there is a strong increase in density, called the pycnocline. This increase may include a sharp drop in temperature (the thermocline) or a rise in salinity (the halocline). In most middle latitudes, both temperature and salinity decrease. In high latitudes, both temperature and salinity may increase downward if the vertical gradient of salinity is strong enough.

Characteristics beneath the surface. The oceans of the world are comprised of layers (**Figs. 3–5**). Because seawater is slightly compressible, it is made warmer with pressure. The potential temperature which is defined as the temperature that a parcel of water would have if it were acted upon by atmospheric pressure alone. Near 4000 m (14,000 ft), where pressures exceed 400 atm (41,000 kilopascals), the actual temperatures are a few tenths of a degree higher than the potential temperatures.

Abyssal waters. The densest waters of the open ocean are formed in high latitudes by cooling of some of the waters that have flowed poleward from the more saline regions. The principal source of the abyssal waters is the Weddell Sea, with some contribution from the Ross Sea, and some from the Norwegian-Greenland Sea passing over the shallow sill between Greenland and Iceland into the North Atlantic Ocean (Fig. 3). The dense waters formed in this fashion are cold and moderate in salinity. They flow equatorward along the western boundaries of the oceans and become progressively warmer, more saline, and less dense by mixing with the overlying water.

The abyssal waters formed in the Weddell and Ross seas are not formed by convection to the bottom in the open ocean. Though these seas are at freezing temperature most of the year, the salinity in the upper layer is so low (about 34.0 parts per thousand) that even at the freezing point the surface waters are not as dense as the underlying waters. The densest waters are formed on the continental shelves, where two processes raise the salinity. Higher-salinity waters from lower latitudes extend southward beneath the surface and raise the salinity on the continental shelves. Leaching out of brine from the ice adds significant amounts of salt to the shallower layer on the shelves. At these raised salinities and the freezing point, the shelf waters become dense enough to spill down the continental slopes to abyssal depths. These waters mix with the open-ocean waters as they pour down. As they reach the bottom, their temperature is near −0.25°C (31.6°F). Their salinity is 34.66–34.7 parts per thousand, far higher than any of the open-ocean salinity values of the Antarctic Ocean.

Overturn to the bottom. Overturn to the bottom does not occur in the open ocean, but takes place only in adjacent seas. The warm and saline waters of the open North Atlantic Ocean flow into the Norwegian-Greenland Sea, where the intense cooling increases their density and convection occurs. The passages from the Norwegian-Greenland Sea to the open ocean are narrow and shallow; and where the dense Norwegian-Greenland Sea waters spill back into the Atlantic, they mix with the less dense waters and their density is reduced. The waters that pour down from the Denmark Strait, between Greenland and Iceland, and still dense enough to reach the bottom in the North Atlantic Ocean, but their density is not as great as those at the bottom of the Weddell Sea. The mixture from the overflow between Scotland and Iceland is less dense and reaches only about 3000 m (10,000 ft). Their major contribution is to the middepth water.

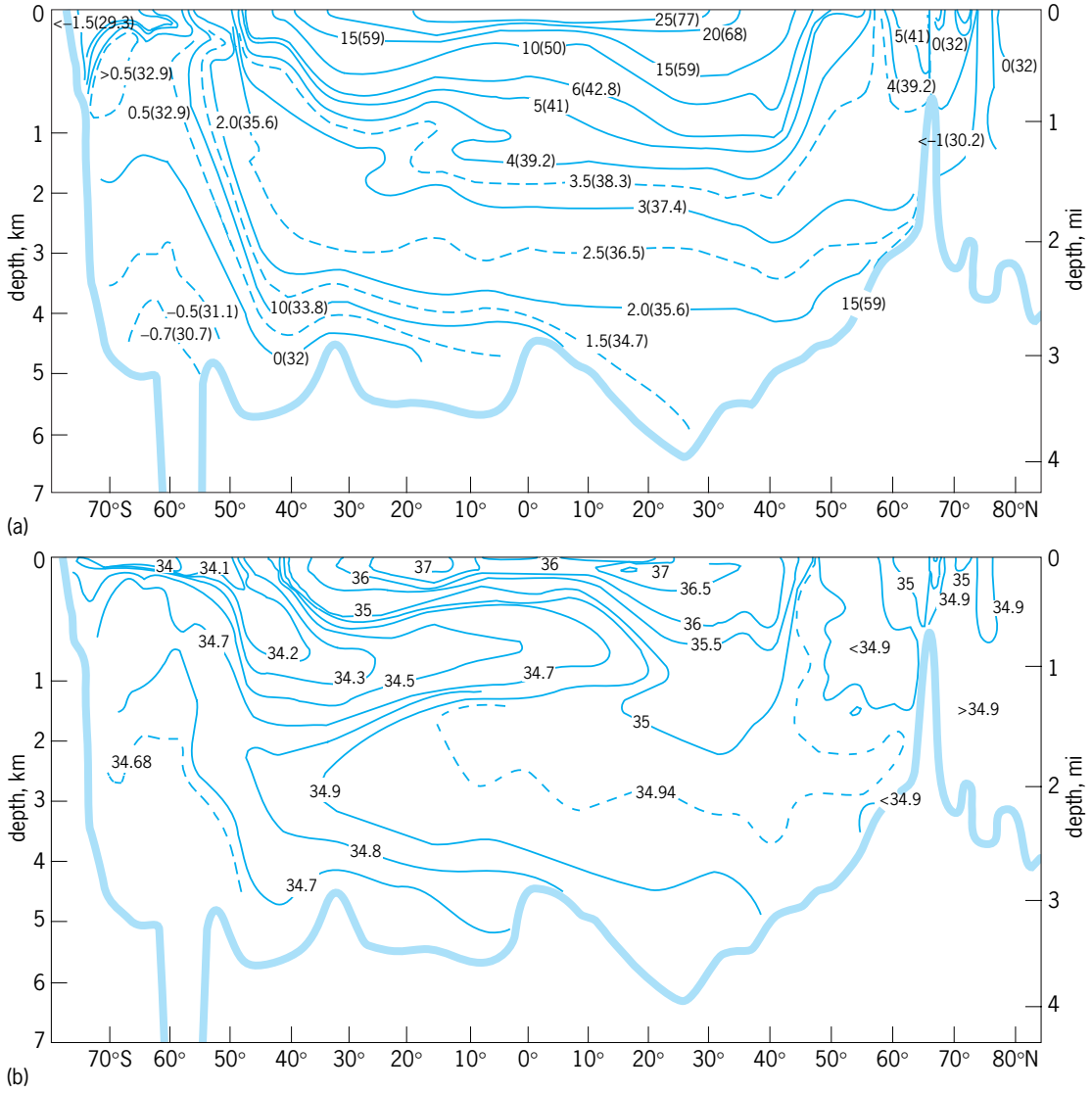


Fig. 3. Layers in seawater in a section through the Atlantic Ocean from the Greenland Sea (right) to the Weddell Sea (left). The thick line represents depth. (*a*) Potential temperature; contour values are in °C (°F). (*b*) Salinity; contours are in parts per thousand.

The Mediterranean Sea is an evaporation basin. Water from the Atlantic flows in at the surface to make up the loss to the atmosphere. Evaporation raises the salinity so high (more than 38 parts per thousand) that even at a temperature of 12°C (36°F) this water is denser than any of the Atlantic waters, and the Mediterranean Sea overturns to the bottom. Then the water pours out through the Strait of Gibraltar, beneath the inflowing Atlantic water, and spills down the slope. Rapid mixing with the Atlantic waters and the water's low compressibility (from its high temperature) keep it from penetrating to the bottom. It provides the very high salinity and temperature that characterize the middepth Atlantic Ocean.

Middepth water. Between the upper layer and the dense abyssal layer, the middepth waters are also supplied from the higher latitudes. Warm and extremely saline water from the Mediterranean Sea makes the North Atlantic the warmest and saltiest of the oceans at depths of 1000–2500 m (3300–8300 ft; Figs. 3–5). Between Iceland and Scotland, warm and saline water from the near-surface waters of the eastern North Atlantic meets the cold and extremely dense water from the Norwegian Sea. Their mixture produces a denser layer of warm and saline water. This water flows westward south of Iceland and Greenland and into the Labrador Sea, where it turns equatorward along the western boundary.

These three North Atlantic sources (Greenland Sea, Mediterranean Sea, and Norwegian Sea–Northeastern Atlantic) account for the characteristics of much of the middepth water in the world ocean. The high salinity in this layer extends southward through the Atlantic (Fig. 3) at depths of 1500–2500 m (5000–8300 ft). It rises with the Antarctic Circumpolar Current near 45 to 60°S latitude and turns westward into the South Indian Ocean (Fig. 4). It continues through the Pacific Ocean (Fig. 5), where part of it turns northward and part continues

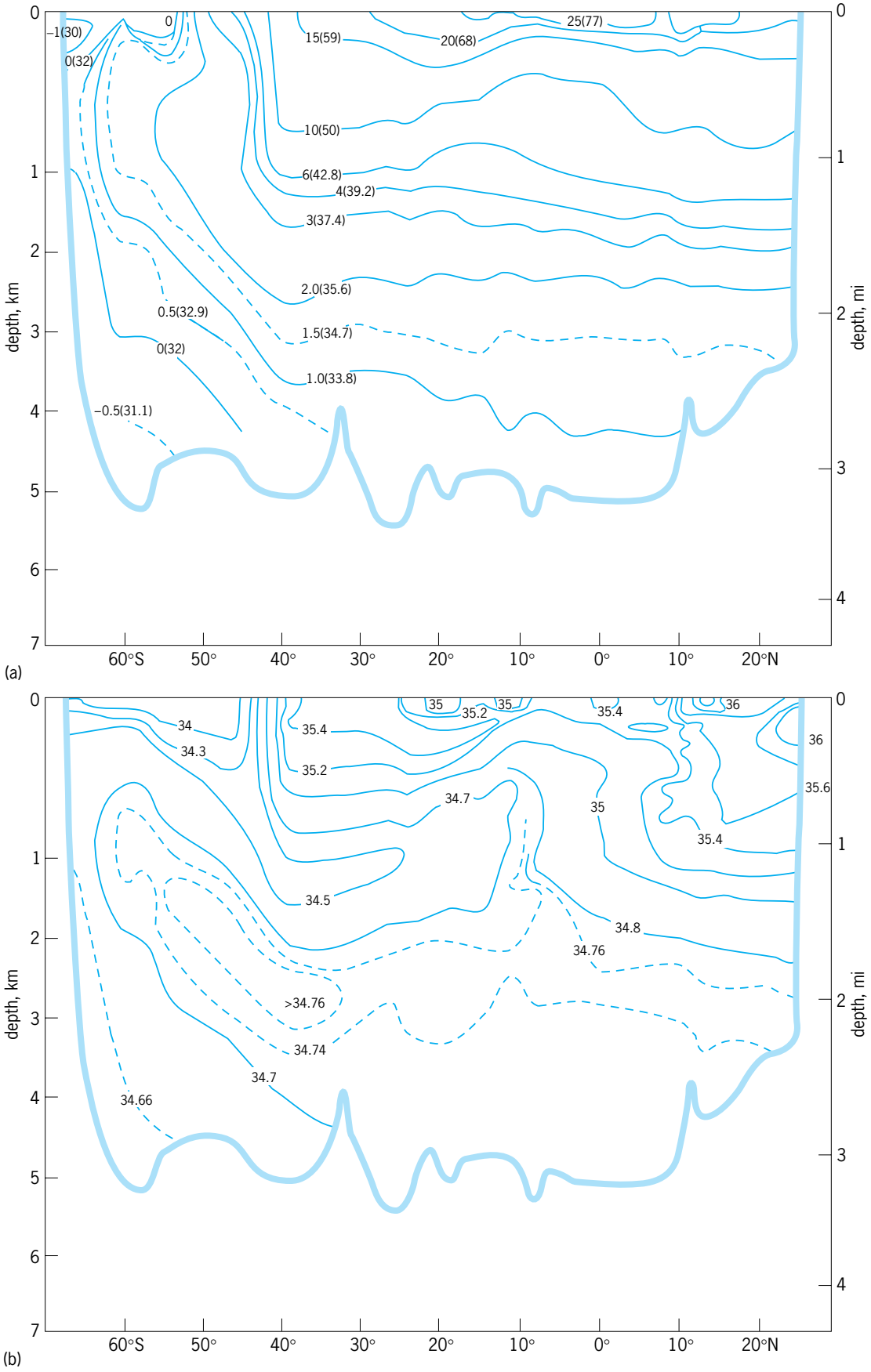


Fig. 4. Layers in seawater in a section through the western Indian Ocean from India (right) to Antarctica (left). The thick line represents depth. (*a*) Potential temperature; contour values are in °C (°F). (*b*) Salinity; contours are in parts per thousand.

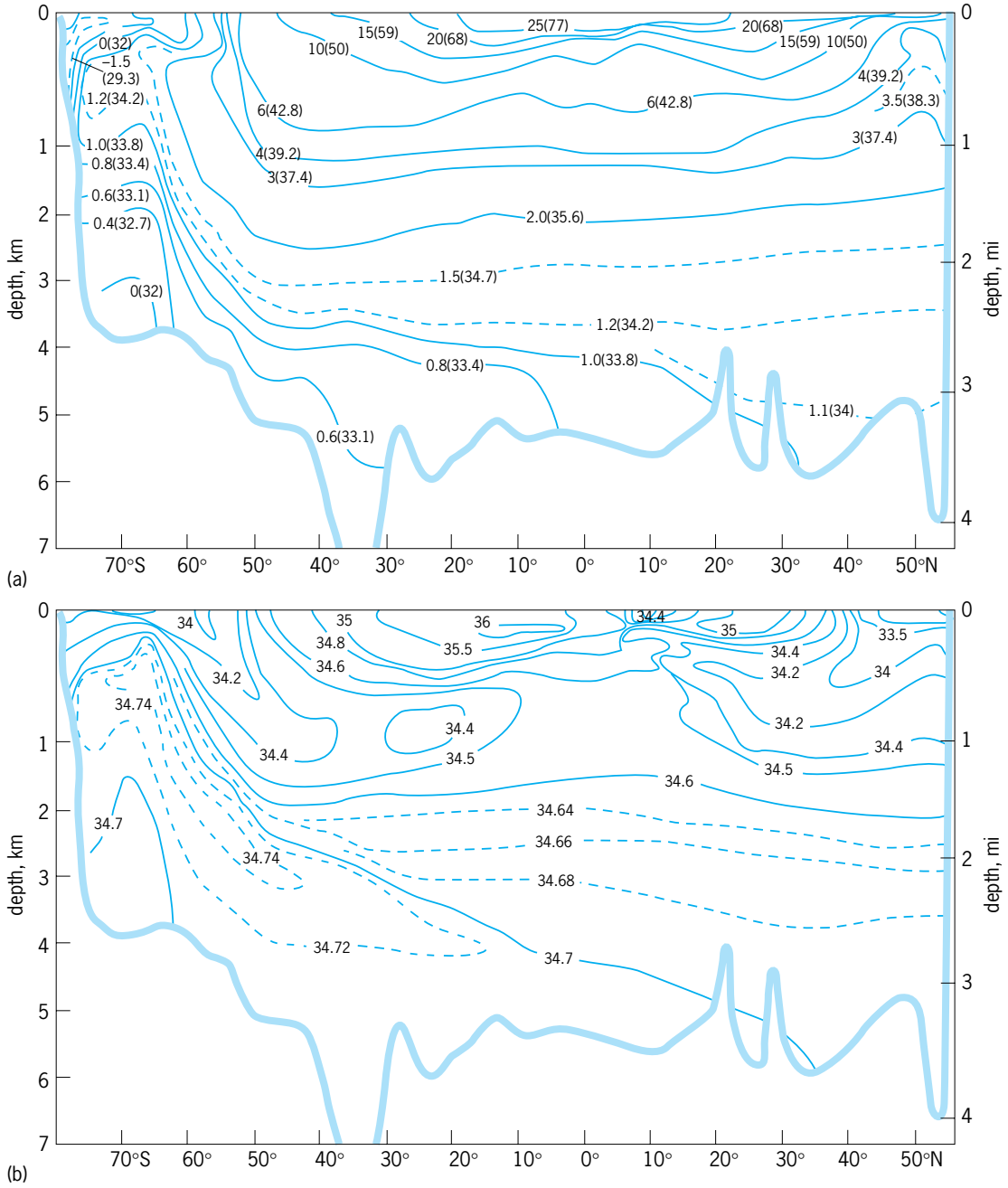


Fig. 5. Layers in seawater in a section through the Pacific Ocean from the Aleutian Islands (right) to the Ross Sea (left). The thick line represents depth. (*a*) Potential temperature; contour values are in °C (°F). (*b*) Salinity; contours are in parts per thousand.

eastward to return, in a diminished form, to the Atlantic Ocean.

Layer of low salinity. The surface waters in high latitudes are made colder and less saline and dense than the adjacent midlatitude waters by exchange of heat and water with the atmosphere. As they extend equatorward, they descend below the upper layer because of their higher density. They are recognized near 800–1000 m (2700–3300 ft) as layers of low salinity from the high southern latitudes of all three oceans, and also from the north in the Pacific.

Layers of high salinity. The highest salinity in the open ocean is found in the upper layer near the tropic circles, where evaporation exceeds precipitation. Vertical mixing through the pycnocline takes place only slowly, and the high surface salinity does not penetrate very well. But these high salinities do spread laterally beneath the less dense surface waters of lower latitudes. In all three oceans (Figs. 3–5), shallow layers of high salinity can be seen extending equatorward from the great high-salinity cells around the tropic circles.

Joseph L. Reid

Thermal Structure

There is a rich spectrum of temperature variations in the ocean. Those with lateral scales of thousands

of kilometers affect the horizontal density distribution and, hence, have a crucial role in determining the geostrophic currents. At the smallest scales (on the order of a centimeter), the temperature variations are usually so small that they have no effect on the density. In these cases the temperature behaves as a passive tracer that shows the stirring of water parcels. Between these extreme spatial scales, temperature can be a tracer of the motion or an active participant in the motion via its effect on the density.

Since temperature is relatively easy to measure, it has been a popular way to study the processes which stir and mix the oceans. Two general types of measurements predominate: (1) vertical profiles of temperature or its gradient, either single profiles or repeated profiles from a steaming or drifting vessel; and (2) horizontal sections of isotherm depth, derived from vertical arrays of many thermistors towed through the water. Operational problems are significant, and the data are always limited by the fact that the vertical profiles are spaced horizontally and the towed arrays have fixed and finite vertical spacing between the sensors.

Fine structure. Temperature generally decreases with depth in the ocean. This decrease is neither steady nor monotonic. The detailed fluctuations with vertical scales on the order of meters are called fine structure. Because of the stratification of the ocean, these features are relatively thin and wide; that is, they have a large aspect ratio. Fine-structure features extend horizontally 100–1000 times (or more) further than they extend vertically.

A parcel of water that is warmer or colder than the water found both above and below must have come into the water column laterally. Known as intrusions, they are quite common (and especially dramatic) in frontal regions where waters with substantially different salinity and temperature characteristics meet. Intrusions 20 m (66 ft) thick with horizontal scales of several kilometers are observed in coastal regions. The rate of change of temperature with depth varies greatly in magnitude and even changes sign. Intrusions are driven by the horizontal pressure gradients that arise from the horizontal density differences. It is possible to have intrusions without a temperature maximum or minimum. These cannot be recognized from the temperature alone, but require examination of the salinity, the density, or some other property such as oxygen.

Another form of temperature fine structure is created by vertical mixing. As the mixing events in the ocean are rarely thorough enough to produce a truly homogeneous patch of water, the result is a relatively low gradient in the patch and relatively high gradients at the edges. A special case of fine structure due to mixing is associated with double diffusive convection, and leads to very strong layering of almost isothermal water with a horizontal extent of many kilometers.

Fine structure is also created by internal waves. Waves produce differential motion with depth, producing alternating regions of horizontal convergence and divergence. Even in an ocean with a smooth temperature profile that decreases monotonically, internal waves can produce local regions of higher and lower gradient. As a result, a smooth temperature profile develops kinks but not inversions. The irregular fine-structure features in the profile can be observed to propagate vertically.

Fine structure due to internal wave motion appears and disappears with the arrival and departure of the wave energy. Hence, it is called reversible fine structure. Lateral intrusions and vertical mixing are not reversible processes; they generate irreversible fine structure. Irreversible fine structure is left behind after the processes that produced it have finished their work. Measurements indicate that in most regions free from intrusions the reversible fine structure dominates over the irreversible fine structure, which would arise from vertical mixing. In regions with intrusions, no general statement can be made.

Microstructure. At scales smaller than the fine structure, from less than 1 m (3.3 ft) down to a few millimeters, the temperature fluctuations are called microstructure. This microstructure is usually due to turbulent mixing processes, with temperature acting as a tracer of the motion. The aspect ratio of these features (ratio of horizontal to vertical scale) decrease with the vertical scale. It might be as large as 100 for 1-m vertical scales and shrink toward 1 at the centimeter scale. The magnitude of the temperature variations decreases with decreasing size.

The magnitude of the fluctuations is determined by the local temperature gradient as well as the size and intensity of the mixing event. Thus, there is no simple relation between the temperature variations and the motion in the water. Molecular thermal diffusivity smooths out the inhomogeneities at all scales, but this effect is most noticeable at scales of a few centimeters and smaller. At these smallest scales, the time for the temperature gradient to diffuse away is from a few seconds (1 mm) to 10 min (2 cm). Therefore, each patch of microstructure is transient, lasting only slightly longer than the processes which produce it.

The detailed features of the thermal structure portend similar features in other parameters such as salinity, density, oxygen, and nutrients. The physical processes that produce fine structure and microstructure also affect the distribution of phytoplankton, small herbivores, and carnivores. However, there are many other processes that affect the chemical and biological distributions in addition to the physical processes that dominate the thermal structure.

Thomas Osborn

Bibliography. J. R. Apel, *Principles of Ocean Physics*, 2d ed., 2001; A. E. Gill, *Atmosphere-Ocean Dynamics*, 1982; M. N. Hill (ed.), *The Sea*, vol. 1: *Physical Oceanography*, 1982, vol. 2: *Composition of Seawater*, 1982; J. A. Knauss, *Introduction to Physical Oceanography*, 2d ed., 1996; E. B. Kraus and J. Businger, *Atmosphere-Ocean Interaction*, 1994; S. K. Majumdar et al. (eds.), *The Oceans: Physical-Chemical Dynamics and Human Impact*, 1994; H. Medwin, *Sounds in the Sea: From Ocean*

Acoustics to Acoustical Oceanography, 2005; F. J. Millero and M. L. Sohn, *Chemical Oceanography*, 2d ed., 1996; M. E. Pilson, *Introduction to the Chemistry of the Sea*, 1998; R. W. Spinrad, K. L. Carder, and M. J. Perry, *Ocean Optics*, 1997; T. Teramoto (ed.), *Deep Ocean Circulation: Physical and Chemical Aspects*, 1993.

Seawater fertility

A measure of the potential ability of seawater to support life. Fertility is distinguished from productivity, which is the actual production of living material by various trophic levels of the food web. Fertility is a broader and more general description of the biological activity of a region of the sea, while primary production, secondary production, and so on, is a quantitative description of the biological growth at a specified time and place by a certain trophic level. Primary production that uses recently recycled nutrients such as ammonium, urea, or amino acids is called regenerated production to distinguish it from the new production that is dependent on nitrate being transported by mixing or circulation into the upper layer where primary production occurs. New production is organic matter, in the form of fish or sinking organic matter, that can be exported from the ecosystem without damaging the productive capacity of the system.

An example illustrates the difference between fertility and productivity. Waters around Antarctica during the austral summer have dense blooms of diatoms and large swarms of krill, the shrimplike zooplankton that feed on the diatoms and in turn are eaten by whales, seals, fish, and penguins. Krill are harvested in Antarctic waters, which contain the only unexploited fishery resource that can significantly increase the supply of protein from the sea (**Fig. 1**). In the austral summer, both primary production, the synthesis of organic matter by photosynthesis, and secondary production, the growth of krill, proceed at high rates. In contrast, during the dark and stormy Antarctic winter there is little sunlight for photosynthesis, and winter storms deeply mix the surface waters. Primary and secondary production proceed at undetectably low rates. While productivity of Antarctic waters varies throughout the year from very low to very high, it is correct to refer to the Antarctic as a fertile region because as a region it has the potential to support high levels of biological activity.

Another analogy that clearly illustrates the distinction between fertility and productivity deals with soil. The rich black soil of the Corn Belt in Illinois and Iowa is always fertile, but in winter the actual productivity of the soil is very low because of low temperatures, short day lengths, and snow cover. *See* BIOLOGICAL PRODUCTIVITY.

The potential of the sea to support growth of living organisms is determined by the fertilizer elements that marine plants need for growth. Fertilizers, or inorganic nutrients as they are called in oceanography, are required only by the first trophic level in the food web, the primary producers; but the supply of inorganic nutrients is a fertility-regulating process whose effect reaches throughout the food web. When there is an abundant supply to the surface layer of the ocean that is taken up by marine plants and converted into organic matter through photosynthesis, the entire food web is enriched, including zooplankton, fish, birds, whales, benthic invertebrates, protozoa, and bacteria. Since the level of seawater fertility is initially set by the primary producers' response to nutrient conditions, this article is limited to the geographic and temporal variations in nutrients and primary productivity. *See* DEEP-SEA FAUNA; FOOD WEB; MARINE FISHERIES.

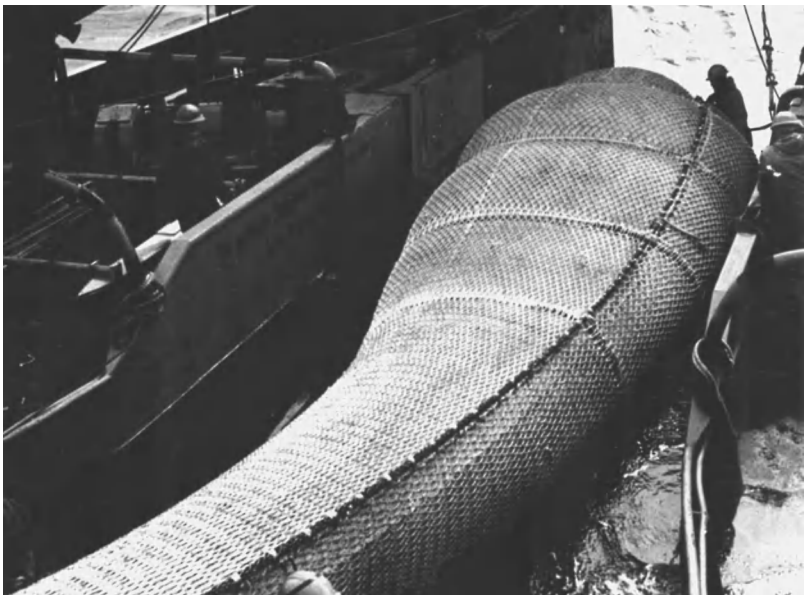


Fig. 1. A 17-ton (15-metric-ton) catch of krill harvested in Antarctic waters. (*Photograph by Dr. Gotthilf Hempel*)

Nutrients. The elements needed by marine plants for growth are divided into two categories depending on the quantities required: The major nutrient elements that appear to determine variations in ocean fertility are nitrogen, phosphorus, and silicon. The micronutrients are elements required in extremely small, or trace, quantities including essential metals such as iron, manganese, zinc, cobalt, magnesium, and copper, as well as vitamins and specific organic growth factors such as chelators. Knowledge of the fertility consequences of variations in the distribution of micronutrients is incomplete, but consensus among oceanographers is that the overall pattern of ocean fertility is set by the major fertilizer elements—nitrogen, phosphorus, and silicon—and not by micronutrients.

Nitrogen, phosphorus, and silicon are required for plants in relatively large quantities because these elements are building blocks of living material. While the other elements in proteins, carbohydrates, and fats such as carbon, oxygen, or sulfur are necessary to living organisms, seawater contains a large excess of them in proportion to the demands of marine plants in the synthesis of new organic matter. Hence the supply of carbon, sulfur, or oxygen does not

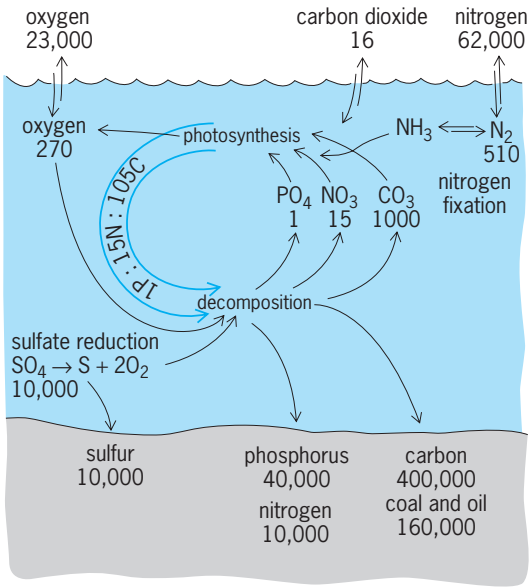


Fig. 2. Biochemical cycle of nutrient elements in the sea. Numbers represent the quantities of elements relative to phosphorus in the ocean and show the excess abundance of carbon and sulfur in comparison with phosphorus and nitrogen. (*After A. C. Redfield, The biological control of chemical factors in the environment, Amer. Sci., 46:205–221, 1958*)

determine variations in ocean fertility (**Fig. 2**). Understanding the pattern of fertility requires knowledge of the pattern of nitrogen, phosphorus, and silicon distribution in the sea.

The nutrient requirement for nitrogen is met in the ocean by forms of fixed inorganic nitrogen such as nitrate, nitrite, or ammonium and by dissolved organic nitrogen species such as urea or amino acids. The nitrate transported across the thermocline into the upper layer supports most of the new primary production in the ocean, while organic nitrogen recycled within the upper layer supports regenerated production. The requirement for phosphorus is met by the phosphate ion and that for silicon by ions of silicate and silicic acid.

Deep waters of the ocean beneath the permanent thermocline contain high concentrations of nitrate, phosphate, and silicate—dissolved inorganic nutrients maintained by a balance of processes. A steady but very small flux of sinking organic particles, especially zooplankton fecal pellets, into the subthermocline water exports material from the upper layer to deep water. In deep water, nitrogen, phosphorus, and silicon are released as the organic matter is decomposed by bacteria and protozoa; the released elements pass from particulate phase back to dissolved phase and are eventually remineralized to the stable terminal oxidation products of nitrate, phosphate, and silicate. In the deep ocean there is no light for photosynthesis, so the abundant inorganic nutrients cannot be taken up by plants; the fertilizer elements accumulate in solution until the nutrient-rich deep water returns to the surface by circulation or mixing. When and where subsurface water is transported to the surface layer, nutrients are brought together with light, and the stage is set for enhanced primary production and greatly increased fertility.

Oceanographers express the atom ratios of inorganic nutrients in seawater in terms of concentrations in microgram-atoms (of N or P) per liter of seawater or the more typical chemical units of micromoles (of nutrient ions) per kilogram of seawater. Since nitrate, phosphate, and silicate are monoatomic ions, the two conventions are quantitatively equal. Deep waters of the Pacific and Indian oceans have a nitrate concentration of about 40 μmol/kg, while Atlantic deep water has half that. Surface waters having a nitrate concentration of more than 10 μmol/kg are highly fertile; those having less than 0.1 μmol/kg are biological deserts. The relative abundances of phosphate and silicate vary in proportion to nitrate.

Primary producers. Two types of marine plants carry out primary production in the ocean: microscopic planktonic algae collectively called phytoplankton, and benthic algae and sea grasses attached to hard and soft substrates in shallow coastal waters.

The benthic and planktonic primary producers are a diverse assemblage of plants adapted to exploit a wide variety of marine niches; however, they have in common two basic requirements for the photosynthetic production of new organic matter: light energy and the essential elements of carbon, hydrogen, nitrogen, oxygen, phosphorus, sulfur, and silicon for the synthesis of new organic molecules. These two requirements are the first-order determinants of photosynthetic growth for all marine plants and, hence, for primary productivity everywhere in the ocean.

Phytoplankton. The term phytoplankton refers to a wide variety of planktonic plants, including diatoms, dinoflagellates, coccolithophorids, microflagellates, blue-green algae, and even ciliates that have endosymbiotic chloroplasts. Diatoms are abundant in the more fertile upwelling and coastal regions of the ocean, and are probably the major group responsible for the new production that provides material to the pelagic food web and to the bottom for consumption by the benthos (see **table**). Coccolithophorids and microflagellates maintain relatively constant but

Primary production and fish production in various provinces of the ocean

Province	Percent of ocean area	Area, mi^2 (km^2)	Average primary production, g C/(m^2)(yr)	Total primary production, tons/yr (metric tons/yr)	Number of trophic levels	Trophic efficiency, percent	Fish production, tons/yr (metric tons/yr)
Offshore ocean	90.0	125×10^6 (326×10^6)	50	17.9×10^9 (16.3×10^9)	5	10	18×10^5 (16×10^5)
Coastal zone	9.9	14×10^6 (36×10^6)	100	4.0×10^9 (3.6×10^9)	3	15	13×10^7 (12×10^7)
Upwelling areas	0.1	1.4×10^5 (3.6×10^5)	300	0.1×10^9 (0.1×10^9)	1.5	20	13×10^7 (12×10^7)

low levels of productivity in the less fertile offshore provinces of the ocean; in these areas microflagellates may have rapid specific growth rates. That is, the rate of synthesis of new cell material per unit of existing cellular material may be high, but since the biomass of phytoplankton in the offshore provinces is low, the absolute productivity is low when compared with more fertile areas.

Dinoflagellates are a functionally complex group and exploit a variety of marine niches. Some dinoflagellate species are slow-growing and persistent in tropical waters, while other species form concentrated but sporadic so-called red tide aggregations that may be poisonous to both marine life and humans; still others occur with diatoms, during nutrient-rich conditions. However, the niche most often associated with dinoflagellates is intermediate between that occupied by diatoms at the fertile end of the nutrient spectrum and that of coccolithophorids and microflagellates at the nutrient-poor end of the spectrum.

Highly fertile regions of the ocean have been studied frequently because important fisheries are concentrated there. The poisonous red tide dinoflagellate outbreaks have also received considerable study. While diatoms and dinoflagellates are major primary producers where the intensity of production is high, microflagellates and coccolithophorids are important in the large tropical and subtropical offshore province of the ocean where production is low (see table). The offshore province occupies 90% of the ocean area, so that the coccolithophorids and microflagellates make a significant contribution to the global carbon cycle even though they may not support important fisheries. *See* PHYTOPLANKTON.

Benthic plants. An advance in the understanding of ocean fertility has been recognition of the role played by benthic algae and sea grasses. These plants occur in dense stands in shallow waters and carry out primary production at extremely high rates throughout the year. Attached algae and grasses are efficient at nutrient scavenging since they remove dissolved nutrients from many volumes of water as it passes over them. Most of the production by attached algae and sea grasses is not grazed directly by herbivores but becomes detritus, and it is consumed in that form by a wide variety of deposit and suspension feeders. Examples of the important benthic primary producers are brown algae such as *Macrocystis*, *Nereocystis*, *Fucus*, or *Laminaria*, and the sea grasses *Zostera* and *Thallasia*.

Productivity measurement. Productivity is usually expressed as the increase in organic matter per unit area per unit time since light enters the marine environment as an areal function from the surface. This convention makes it possible to compare the productivity of the ocean directly with that of land. The units most frequently used are grams of carbon per square meter (g C/m^2). Primary production is measured over short time intervals of 1 day or less, producing data units of g C/(m^2) (day). For food web considerations these data are converted to annual production as g C/(m^2) (year), which introduces very large errors because oceanographers have not made daily measurements throughout annual cycles to determine the correct daily-to-annual conversion factors for different marine ecosystems.

Productivity conventionally is expressed in terms of carbon. However, there is no fundamental rationale for choosing carbon over nitrogen, phosphorus, or a cell property like protein. Carbon is convenient, however, because the most precise technique for measuring primary production uses the radioactive isotope of carbon. The excess inorganic carbon in seawater satisfies the requirement of the radioactive isotope of carbon. The excess inorganic carbon in seawater satisfies the requirement of the radiocarbon method that there be no detectable change in the precursor substrate during the measurement. With nitrogen, phosphorus, or silicon, the requirement that there be significant change in the substrate concentration is not often satisfied, so determination of the in-place rate of uptake of these elements is difficult or impossible with isotopic tracers. Conversely, production in terms of nitrogen, silicon, or phosphorus can be accurately measured by following the disappearance of these nutrients from seawater, a technique not possible with carbon.

In fertile areas primary production may be over 1.0 g C/(m^2)(day), but the sea grasses, kelp, or upwelling phytoplankton can exceed 10.0 g C/(m^2)(day). It is suspected that these high daily rates of production in fertile areas do not persist for long in the annual cycle, so oceanographers estimate that even highly fertile areas have an annual primary production of only about 300 g C/(m^2)(year). This crude estimate must suffice until time series studies are done. Provinces of the ocean that are permanently nutrient-poor due to strong thermal stratification have less than 0.1 g C/(m^2)(day) with little variation throughout the annual cycle. The yearly production is estimated to be 30–40 g C/(m^2)(year). While the coccolithophorids and microflagellates that are the dominant primary producers in this infertile area are important in the global flux of carbon, they are insignificant in the production of fish, as shown in the table. *See* ALGAE.

Patterns of fertility. The regions of the world's oceans differ dramatically in overall fertility. In the richest areas, the water is brown with diatom blooms, fish schools are abundant, birds darken the horizon, and the sediments are fine-grained black mud with a high organic content. In areas of low fertility, the water is blue and clear, fish are rare, and the bottom sediments are well-oxidized carbonate or clay. These extremes exist because the overall pattern of fertility is determined by the processes that transport nutrients to the sunlit upper layer of the ocean where there is energy for photosynthesis.

Upwelling. The fisheries of the world are concentrated in regions of enhanced fertility; it is not generally appreciated that the fertile food-producing areas of the ocean, like on land, are concentrated into a very small proportion of the total area. It has been estimated that 99% of the world's fish catch comes from

only 10% surface of the area of the ocean (see table), and of that, 50% comes from the 0.1% ocean area where upwelling carries nutrient-rich water to the surface. Upwelling on the west coasts of the Americas and Africa is a well-known process that enriches fertility, modifies the coastal climate, and has great economic impact because of the large upwelling-based fisheries. However, there are other processes that replenish nutrients to the depleted surface layer, but these processes are not nearly as well known as upwelling. *See* UPWELLING.

The annual cycle of heating and cooling of the Earth's surface in temperate and high latitudes drives the transport of nutrients into the upper layer of the ocean. When there is net negative heat flux to the ocean from the atmosphere, the surface water cools, becomes denser and unstable, and subsequently mixes with the underlying water. Mixing replaces nutrient-depleted surface water with nutrient-rich water and sets the stage for high productivity. When the heat flux becomes positive, the now nutrient-rich water is heated, becomes less dense, and forms a stable upper layer separated from the cooler deep water by a thermocline. In the stabilized upper layer, light for photosynthesis is available, and a rapid burst of phytoplankton growth occurs. This brief burst of high primary productivity by the diatom-dominated phytoplankton assemblages is the famous spring bloom of regions such as Long Island Sound, the Gulf of Maine, or the North Sea. The spring bloom in temperate zone seas depletes the upper layer of nitrate, phosphate, and silicate, causing a productivity decrease to a lower maintenance level until the next mixing event replenishes the nutrients.

Estuaries. Studies in lateral estuaries of the Chesapeake Bay and along the coast of Maine show that mixing during the strong tidal flows of the full and new moon generate short blooms in phase with the lunar cycle. In the waters around the British Isles, tidal mixing produces quasistationary blooms where the tidal currents hit subsurface banks or ledges and water mixes upward. Downstream from the wake-generating obstruction, heating restabilizes the upper layer, and blooms of phytoplankton exploit the available nutrients. Wake mixing over topographic features such as Georges Bank, Grand Banks, or Dogger Bank complements seasonal mixing and supports continuously high productivity in a manner analogous to the topographically fixed blooms of upwelling centers. An important portion of the world's fish catch comes from these banks. This yield is made possible by the physical transport of nutrients upward to replace the nitrogen, phosphorus, and silicon removed in the form of organic exports such as harvested fish. Physical processes drive both the seasonal heat-driven mixing and stabilization and the tidal mixing and stabilization; both processes are important to the continuing fertility of fishing areas such as Georges Bank.

Polar seas. In the polar seas a short burst of high production occurs in midsummer as increased sunlight becomes available for photosynthesis. At that time the water stabilizes slightly from solar heating, and the melting ice packs release fresh water that further stabilizes the surface layer. Phytoplankton blooms in both the Bering and Weddel Seas have some of the highest concentrations of phytoplankton ever recorded in the open ocean, but in both the Arctic and Antarctic the blooms are short-lived and very patchy in distribution. A brief burst of primary production supports a rich food web with large populations of higher-trophic-level organisms such as krill fish and marine mammals. The ecology of these highly fertile polar seas is not well understood. Because the valuable unexploited food resources of the polar areas merit careful management, it is important to have a basic understanding of these ecosystems.

A major difference in the productivity cycle of polar seas is that phytoplankton growth does not proceed to nutrient depletion, and productivity is regulated at the end of the brief summer by the lack of light, deep wind-driven mixing, and low temperatures. Polar and tropical waters are dramatic opposites in regulation of fertility. Polar waters always have an abundance of nutrients, so they are described as being continuously fertile; but there is light limitation of photosynthesis during most of the year. Tropical and subtropical offshore waters are permanently nutrient-depleted; light is never limiting in the upper layer, but the solar heating is so intense that thermal stratification is not broken down in the annual cycle. A biological desert is maintained because the subthermocline nutrient-rich water never mixes or circulates to the upper layer. (Dramatic exceptions to this are coastal and equatorial upwelling and the wake-stream mixing that occurs behind islands and atolls.) In the tropical and subtropical offshore waters that are permanently low in fertility, there is very efficient recycling of nutrients and high turnover rates. That is not true for the regenerated production of tropical and subtropical waters, which explains the fragile nature of tropical and subtropical ecosystems in comparison to the relatively resilient character of higher latitude marine ecosystems.

Richard T. Barber

Size of populations and fluctuations. In temperate waters, fish often spawn at the same place at the same time of year. The larvae drift in a current from spawning ground to nursery ground, and the adults migrate from the feeding ground to the spawning ground. Adolescents leave the nursery to join the adult stock on the feeding ground. This migration circuit is based on the track of the larval drift, and the population is isolated by its unique time of spawning. The larvae suffer intense mortality, and numbers in the population are perhaps regulated naturally during the period of larval drift. The same principles apply to the upwelling areas in a more diffuse way. The tuna spawn nearly all year round across the North Pacific, and their larvae and juveniles drift in the major currents.

There are three basic methods of measuring the sizes of marine populations. The first is by census based on samples, that together constitute a known

fraction of the whole population, as in the case of sessile species such as shellfish, or of a particular age range, as in the case of fish that have pelagic eggs whose total abundance can be measured by fine-meshed nets hauled vertically through the water column. The second is by marking or tagging, in which a known number of marked individuals are mixed into the population and the ratio of marked to unmarked fishes is subsequently estimated from samples. This method may survive in lakes, but in the sea it is no longer used, because better estimates of stock may be obtained by the third method. From the catches in numbers at age in a fish population, sequential population analysis is used to obtain estimates of fishing mortality (or stock) by age. From the oldest age group in a year class, with educated guesses of fishing and natural mortality, the analysis proceeds back in the year class to the youngest, correcting the estimate as it proceeds.

The largest measured populations in numbers are of pelagic fish, particularly of the families of herring and sardines. The Atlantic herring (*Clupea harengus*) comprises a number of distinct stocks, the largest of which comprises 1 billion mature individuals. Each population ranges over hundreds or even thousands of kilometers in the northeast Atlantic.

Populations of bottom-living fish tend to fluctuate slowly over periods of 50–75 years. But those of pelagic fish, like the Atlantic herring, may fluctuate dramatically. Fish that spawn at a fixed season are obviously vulnerable to climatic change. During long periods there are shifts in wind strength and direction, with consequent delays or advancements in the timing of the production cycle. If the cycle is progressively delayed, the fish larvae, hatched at a fixed season, become short of food, and the populations decline with time. The match or mismatch of larval production to that of their food is probably the agent by which variation in the magnitudes of the recruiting year classes are determined. Climatic change affects the population during the period of the larval drift when isolation is maintained and when numbers are probably regulated. Recruitment is determined by the growth and mortality of the fish larvae and the juveniles, variations in which may be modified by a range of matches and mismatches in the production of larvae and juveniles to production of their food. *See* CLIMATE HISTORY. D. H. Cushing

Biological species and water masses. Biogeographical regions in the ocean are related to the distribution of water masses. Their physical individuality and ecological individuality are derived from partly closed patterns of circulation and from amounts of incident solar radiation characteristic of latitudinal belts. Each region may be described in terms of its temperature-salinity property and of the biological species which are adapted to all or part of the relatively homogeneous physical-chemical environment.

Cosmopolitan species. The discrete distributions of many species are circumscribed by the regions of oceanic convergence bounding principal water masses. Other distributions are limited to current systems. Cosmopolitan species are distributed across several of the temperature-salinity water masses or oceans; their wider specific tolerances reflect adaptations to broadly defined water types. No pelagic distribution is fully understood in terms of the ecology of the species.

A habitat is integrated and maintained by a current system: oceanic gyral, eddy, or current, with associated countercurrents. This precludes species extinction that could occur if a stock were swept downstream into an alien environment. The positions of distribution boundaries may vary locally with seasonal or short-term changes in temperature, available food, transparency of the water, or direction and intensity of currents.

Phytoplankton species are distributed according to temperature tolerances in thermal water masses, but micronutrients (for example, vitamin B_{12}) are essential for growth in certain species. The cells of phytoplankton reproduce asexually and sometimes persist in unfavorable regions as resistant resting spores. New populations may develop in prompt response to local change in temperature or in nutrient content of the water. Such species are less useful in tracing source of water than are longer-lived, sexually reproducing zooplankton species.

Indicator organisms. The indicator organism concept recognizes a distinction between typical and atypical distributions of a species. The orgin of atypical water is indicated by the presumed affinity of the transported organisms with their established centers of distribution.

Zooplankton groups best understood with respect to their oceanic geography are crustaceans such as copepods, and euphausiids, chaetognaths (arrow worms), polychaetous annelids, pteropod mollusks, pelagic tunicates, foraminiferids, and radiolarians. Of these the euphausiids are the strongest diurnal vertical migrants (60–200 ft or 200–700 m). The vertical dimension of euphausiid habitat agrees with the thickness of temperature-salinity water masses, and many species distributions correspond with the positions of the masses. In the Pacific different species, some of which are endemic to their specific waters, occupy the subarctic mass (such as *Thysanoessa longipes*), the transition zone, a mixed mass lying between subarctic and central water in midocean and between subarctic and equatorial water in the California Current (for example, *Nematoscelis difficilis*), the barren North Pacific central (such as *Euphausia hemigibba*) and South Pacific central masses (for example, *E. gibba*), the Pacific equatorial mass (such as *E. diomediae*), a southern transition zone analogous to that of the Northern Hemisphere (represented by *Nematoscelis megalops*), and a circumglobal subantarctic belt south of the subantarctic convergence (such as *E. lucens*).

Epipelagic fishes and other strongly swimming vertebrates are believed to be distributed according to temperature tolerances of the species and availability of food. However, distributions of certain bathypelagic fishes (such as *Chauliodus*) have been related to water mass. *See* SEAWATER. Edward Brinton

Bibliography. C. Lalli and T. Parson, *Biological Oceanography: An Introduction*, 2d ed., 1995; J. S. Levinton, *Marine Biology: Function, Biodiversity, Ecology*, 2d ed., 2001; K. H. Mann and J. R. N. Lazier, *Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans*, 2d ed., 1996; T. R. Parsons, M. Takahashi, and B. Hargrave, *Biological Oceanographic Processes*, 3d ed., 1984; A. C. Redfield, The biological control of chemical factors in the environment, *Amer. Sci.*, 46:205–221, 1958.

Sebaceous gland

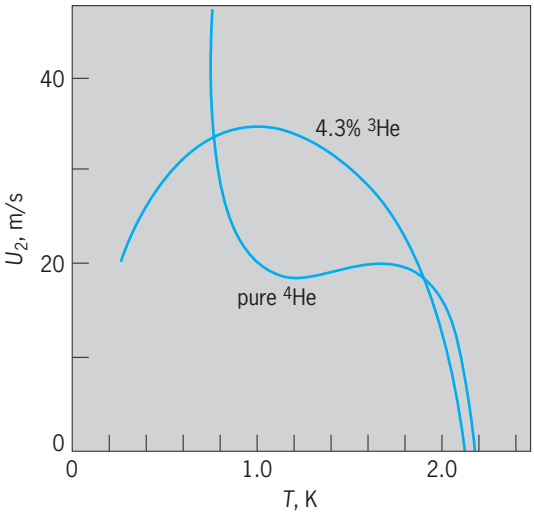
A gland which produces and liberates sebum, a mixture composed of fat, cellular debris, and keratin. When the gland arises in association with a hair follicle, it forms a thickened outpushing from the side of the developing follicle near the epidermis. Central cells in these sebaceous glands form oil droplets within the cytoplasm. These cells disintegrate to liberate the sebaceous substance and are therefore of the holocrine type. The Meibomian or tarsal glands, within the tarsus or supporting plate at the edge of the eyelids, are sebaceous and complex tubuloacinous structures. The numerous separate glands open along the entire edge of the upper and lower lids. Retained secretions of the tarsal glands produce a cyst termed a chalozion or Meibomian cyst. *See* EPITHELIUM; GLAND.

Olin E. Nelsen

Second sound

A type of wave propagated in the superfluid phase of liquid helium (helium II) and in certain other substances under special conditions. Predicted independently by L. Tisza in 1938 and L. Landau in 1941, such waves were observed first by V. Peshkov in 1944 and have subsequently provided a rich source of information for the study and understanding of the superfluid state. The name is misleading since second sound is not in any sense a sound wave, but a temperature or entropy wave. In ordinary or first sound, pressure and density variations propagate with very small accompanying variations in temperature; in second sound, temperature variations propagate with no appreciable variation in density or pressure.

Two-fluid model. The two-fluid model of helium II provides further insight into the nature of second sound. In this model the liquid can be described as consisting of superfluid and normal components of densities ρ_s and ρ_n, respectively, such that the total density $\rho = \rho_s + \rho_n$. The superfluid component is frictionless and devoid of entropy; the normal component has a normal viscosity and contains the entropy and thermal energy of the system. As the temperature goes from 0 K to T_λ (the superfluid transition temperature), ρ_s goes from ρ to 0, and ρ_n from 0 to ρ. In a temperature or second-sound wave, the normal and superfluid flows are oppositely directed so that $\rho_s \mathbf{V}_s + \rho_n \mathbf{V}_n = 0$, where $\mathbf{V}_s$ and $\mathbf{V}_n$ are the superfluid and normal flow velocities. Thus a variation in relative densities of the two components, and hence a temperature fluctuation, propagates with no change in total density or pressure. In a first-sound wave, the two components move in phase, that is, $\mathbf{V}_n \cong \mathbf{V}_s$.



Velocity of second sound as a function of temperature for pure ^{4}He and a 4.3% ^{3}He mixture. (*After J. C. King and H. A. Fairbank, Second sound in ^{3}He-^{4}He mixtures below 1 K, Phys. Rev., 93:21, 1954*)

Second sound can be observed by methods similar to those used with first sound except that a heater and thermometer replace the sound transmitter and receiver. Thus an electrical resistor in the form of a thin metallic film or wire is the most common type of transducer for second sound.

Velocity. The velocity of second sound goes from 0 at T_λ to a value approaching that of first sound at 0 K as seen in the **illustration**. This is in excellent agreement with the theoretical relation in the equation shown below deduced from the two-fluid

$$u_2^2 = \frac{\rho_s}{\rho_n}\frac{T}{C}S^2$$

model, where T is the temperature, C the specific heat, and S the entropy of the liquid. At temperatures below about 0.5 K, Landau showed that the only excitations present in superfluid helium are longitudinal phonons (quantized sound waves), and hence heat pulses would be propagated as a fluctuation in the density of these phonons. As the temperature approaches 0, the equation shown for second-sound velocity reduces to $u_2 = u_1\sqrt{3}$, where u_1 is the phonon or first-sound velocity. Experimentally this value of velocity is approached at about 0.5 K, but another effect interferes—the interactions between phonons is very weak and the phonon mean free path goes rapidly toward infinity as the temperature and the density of phonons are reduced. Thus, below 0.5 K second sound is no longer observed in the true sense but the phonons propagate ballistically, interacting only with the walls of the container. This behavior is indeed found, as evidenced by large spreading of temperature pulses and the arrival of the leading edge with the phonon velocity.

In a variety of experiments second sound displays

the properties expected in wave phenomena such as diffraction, interference, and shock effects for large amplitude pulses.

^{3}He-^{4}He mixtures. Even a small addition of the lighter isotope helium-3 significantly affects many of the superfluid properties including second sound, especially in the very low temperature range. At the lowest temperatures where the ^{3}He excitations dominate the entropy and specific heat, the second-sound velocity is proportional to the square root of the temperature as expected for first sound in an ideal gas. Study of the second sound in this region provides an excellent means of measuring the details of the ^{3}He interactions.

Liquid ^{3}He. Liquid ^{3}He is a superfluid only at very low temperatures (below about 0.002 K). Second sound should exist in the principal phases (A and B) of this superfluid, but the expected velocity is quite small and the attenuation so large as to make direct detection unlikely. However, a new condensed form of second-sound wave and spin density wave can be propagated in a restricted narrow region of the ^{3}He phase diagram, the A_1 phase. This phase lies between the normal phase and the superfluid A phase and exists only in a strong magnetic field.

The superfluid component of the A_1 phase is apparently made up of pairs of ^{3}He atoms with their nuclear magnetic moments lined up parallel to the strong applied magnetic field. Counterflow of the normal and superfluid components, therefore, involves both the temperature and spin density changes in this normal second-sound wave/spin density wave combination. The velocity is directly proportional to the reduced temperature $(1 - T/T_{A1})$ and reaches a maximum value of about 1.4 m/s at the lower temperature limit of the A_1 phase. Here T_{A1} is the transition temperature between the A_1 phase and the normal phase. *See* LIQUID HELIUM; SUPERFLUIDITY.

Solid dielectric crystals. Theoretical predictions that second sound should exist in certain solid dielectric crystals under suitable conditions have been confirmed experimentally for solid helium single crystals at temperatures between 0.4 and 1.0 K. In a dielectric solid, heat pulses are carried only by phonons, as for liquid helium II, at very low temperatures. If resistive phonon scattering processes dominate, such as phonon-phonon umklapp processes, scattering by isotope or chemical impurities, or scattering by boundaries or imperfections in the crystal, the first-order diffusive flow equation would govern heat or temperature propagation. However, if the mean free path for such processes is large and the mean free path for nondissipative phonon-phonon normal processes is small, the second-order wave equation is obeyed, and heat pulses are propagated as second sound. That second sound has not yet been seen in most other dielectric crystals is probably due to the diffusive scattering of the phonons by impurities. However, evidence of second-sound propagation has also been found in pure sodium fluoride (NaF) crystals, and fully developed second sound has been found in pure semimetal bismuth crystals. Second sound has provided a good method of measuring directly the mean free path for phonon normal scattering processes. *See* LOW-TEMPERATURE ACOUSTICS.

Liquid crystals. Another quite different class of materials can exhibit second sound. In smectic A liquid crystals, when the wave vector is oblique with respect to the layers of these ordered structures, a modulation of the interlayer spacing can propagate at nearly constant density. It is likely that second sound will continue to be a valuable probe for studying the properties of interesting ordered materials.

Henry A. Fairbank

Bibliography. K. H. Bennemann and J. B. Ketterson, *The Physics of Liquid and Solid Helium*, 1976; J. Wilks and D. S. Betts, *An Introduction to Liquid Helium*, 2d ed., 1987; P. Wolf and P. Volhardt (eds.), *The Superfluid Phases of Helium-3*, 1990.

Secondary emission

The emission of electrons from the surface of a solid into vacuum caused by bombardment with charged particles, in particular, with electrons. The mechanism of secondary emission under ion bombardment is quite different from that under electron bombardment; the discussion here is limited to the latter case because it is in this sense that the term secondary emission is generally used.

The bombarding electrons and the emitted electrons are referred to, respectively, as primaries and secondaries. Secondary emission has important practical applications because the secondary yield, that is, the number of secondaries emitted per incident primary, may exceed unity. Thus, secondary emitters are used in electron multipliers, especially in photomultipliers, and in other electronic devices such as

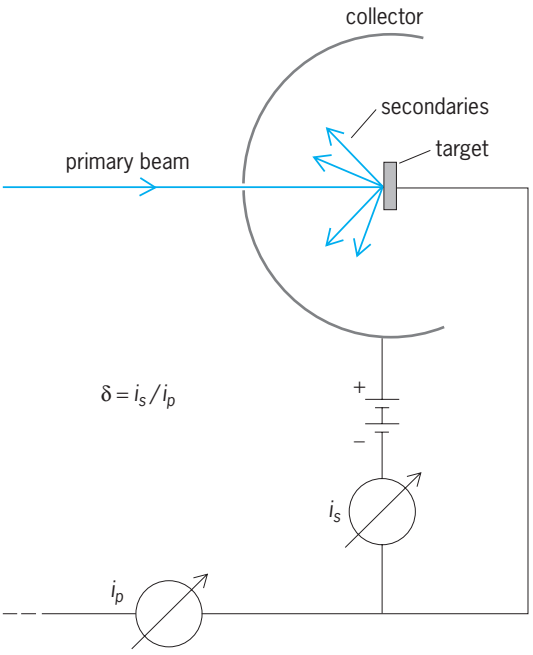


Fig. 1. Schematic circuit for measuring secondary yield; i_p and i_s represent the primary and secondary currents.

television pickup tubes, storage tubes for electronic computers, and so on.

Secondary yield. The most thoroughly investigated property of secondary emission is the yield as a function of the energy of the primaries. The yield may be measured by means of the circuit shown schematically in **Fig. 1**. A beam of primary electrons strikes a target with an energy determined by the potential difference between the target and the cathode. The primary beam passes through a hole in the collector, which has been made positive with respect to the target. The secondaries emitted by the target then flow to the collector, and the yield is obtained as the ratio of the secondary current i_s to the primary current i_p.

Mechanism of the process. The emission of secondary electrons can be described as the result of three processes: (1) excitation of electrons in the solid into high-energy states by the impact of high-energy primary electrons, (2) transport of these secondary electrons to the solid-vacuum interface, and (3) escape of the electrons over the surface barrier into the vacuum. The efficiency of each of these three processes, and hence the magnitude of the secondary emission yield δ, varies greatly for different materials.

Taking into account the material characteristics that, in addition to the value of the primary energy E_p determine the yield δ, one arrives at the equation

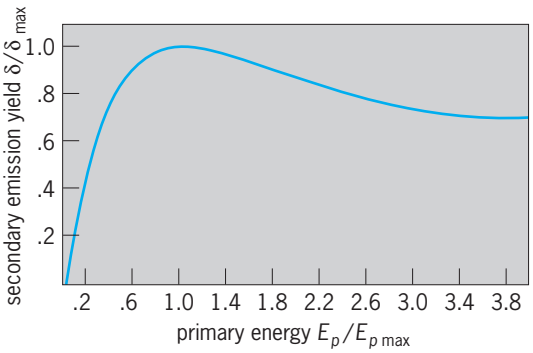


Fig. 2. Theoretical curve of secondary emission yield as a function of primary energy in normalized coordinates.

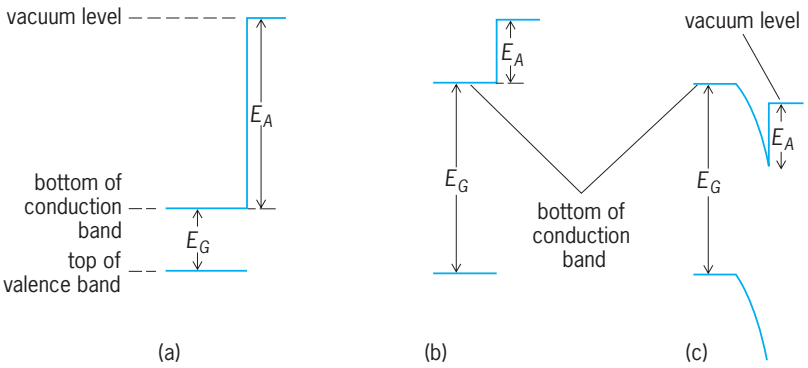


Fig. 3. Typical energy band models for semiconductors (or insulators). (*a*) $E_G \ll E_A$. (*b*) $E_G \gg E_A$. (*c*) Negative electron affinity.

below for the dependence of δ upon E_p, where ϵ

$$\delta = \frac{B_1 B_2 E_p}{\epsilon R}(1 - e^{-R/L})$$

is the energy required to produce a secondary electron; R is the range of primary electrons; B_1 is the coefficient, taking into account that only a fraction of the excited electrons diffuse toward the surface; L is the mean free path of the secondary electron; and B_2 is the probability that an electron reaching the solid-vacuum interface can escape over the surface barrier. Here ϵ and R are associated with process 1 above, B_1 and L with process 2, and B_2 with process 3. Without giving a detailed derivation of the equation above, it is qualitatively plausible that δ increases with increasing B_1, B_2, E_p, and L and decreases with increasing ϵ and R.

On the basis of the equation, a universal curve can be derived (**Fig. 2**) in which δ/δ_{max} is plotted versus E_p/E_{pmax}, where δ_{max} is the maximum yield and E_{pmax} is the corresponding primary energy. Whereas curves for the absolute values of δ versus E_p vary over a wide range for different materials, experiments have generally confirmed the validity of the curve in Fig. 2. The peak in the curve can be interpreted as follows: With increasing E_p, the number of secondaries produced within the solid increases, but at the same time the primaries penetrate to a greater depth in the material. Because of energy loss processes, the "escape depth" of the secondaries has a finite value which is determined by some of the parameters entering the equation above. Thus the peak of the curve represents the point beyond which the number of secondaries produced at a depth greater than the escape depth exceeds the number of additional secondaries produced due to the higher primary energy. Because the escape depth is predominantly determined by L and B_2 in the equation, the variations in these two parameters are the main reasons why the δ values for different materials vary over such a wide range.

Experimental yield curves. In a discussion of measured δ and E_p curves, it is useful to consider metals and semiconductors (or insulators) separately. In metals the secondaries lose their energy rapidly by electron-electron scattering. As a result, L and B_2 in the equation are small, and the escape depth is of the order of, at most, nanometers. Hence, δ_{max}, and consequently E_{pmax}, have low values, typically well below 2.

For semiconductors and insulators, the situation is more complicated and is best understood in terms of energy-band models. Referring to **Fig. 3**, the highest δ values that can be obtained depend on the relative position of the top of the valence band (where the secondary electrons originate), the bottom of the conduction band, and the vacuum level. Three typical band models are shown in Fig. 3. The model shown in Fig. 3*a* is characterized by a small ratio of band-gap energy E_G to electron affinity E_G to electron affinity E_A. In the model shown in Fig. 3*b* the E_G to E_A ratio is large. In the model shown in Fig. 3*c* the bands are bent downward to such an extent that the

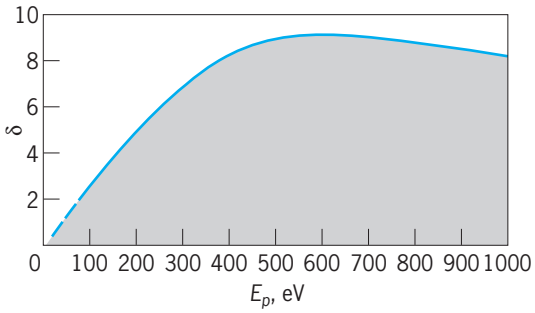


Fig. 4. Secondary emission yield versus primary energy for MgO.

vacuum level lies below the bottom of the conduction band in the bulk. A material with this characteristic is said to have negative effective electron affinity. This concept is also of great importance in photoelectric emission. The differences in secondary emission yields associated with each of the three band models can be qualitatively summarized as follows. *See* PHOTOEMISSION.

$E_G \ll E_A$ model. Secondary electrons excited from the valence band to levels above the vacuum level tend to lose their energy by exciting additional electrons from the valence band into the conduction band and thus arrive at the solid–vacuum interface with insufficient energy to overcome the surface barrier. In other words, the escape depth is very small, and the maximum δ values are below 2, similar to those of metals. Examples of this model are germanium and silicon.

$E_G \gg E_A$ model. Whereas secondary electrons excited from the valence band gradually lose energy by phonon-phonon scattering, an appreciable number of secondaries reach the solid–vacuum interface with sufficient energy to overcome the surface barrier. In other words, the escape depth is larger than in the case where $E_G \ll E_A$, of the order of tens of nanometers, and maximum δ values in the 8–15 range are typically obtained. Most of the materials used in practical devices fall into this category. Examples are MgO (see **Fig. 4**), BeO, Cs_3Sb (cesium antimonide), and KCl.

Negative effective electron affinity. Here the vacuum level is below the bottom of the conduction band. This case differs drastically from that shown in Fig. 3*b* because electrons that have dropped to the bottom of the conduction band as a result of phonon-phonon scattering still have enough energy to escape into the vacuum. Because the lifetime of electrons in the bottom of the conduction band is orders of magnitude longer than in states above this level, the escape depth of the secondaries is orders of magnitude greater than in the case represented in Fig. 3*b*. The most important material in this category is cesium-activated gallium phosphide, GaP(Cs). **Figure 5** shows the δ versus E_p curve for GaP(Cs) by comparison with MgO. Because of the much greater escape depth, δ values exceeding 100 are readily obtained. The curve for GaP(Cs) still follows quite closely the universal curve (Fig. 2), but the $E_{p\max}$ value is now in the 5–10-kV region compared with several hundred volts for materials represented in Fig. 3*a* or *b*.

Materials of the GaP(Cs) type represent a major breakthrough in the use of secondary emission for practical devices. In photomultipliers, for example, GaP(Cs) is superior to the conventional materials represented in Fig. 3*b* for a number of reasons, the greatest advantage being the improved signal-to-noise ratio. Negative effective electron affinity materials are not in universal use because of the more complex activation procedure and the associated higher cost. *See* BAND THEORY OF SOLIDS; SEMICONDUCTOR.

Dependence on angle of primary beam. The secondary yield for a given primary energy increases as the angle θ between the primary beam and the normal to the surface increases; the secondaries are then produced closer to the surface and consequently have a larger escape probability. At the same time,

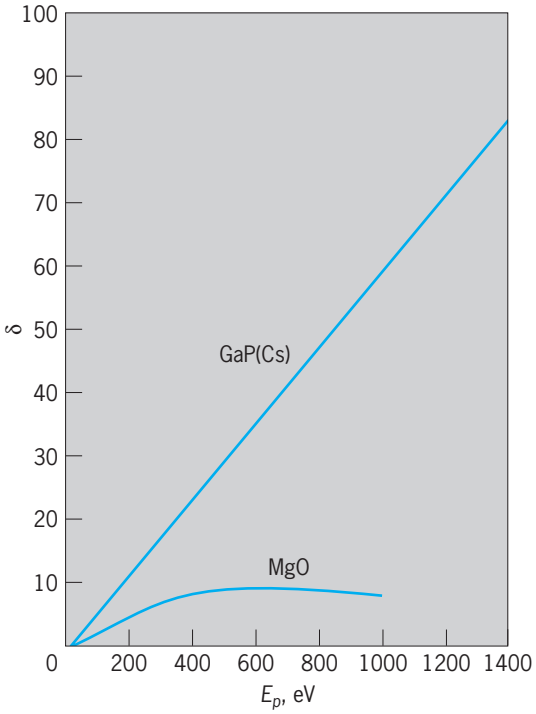


Fig. 5. Secondary emission yield versus primary energy for GaP(Cs). The curve for MgO is shown for comparison.

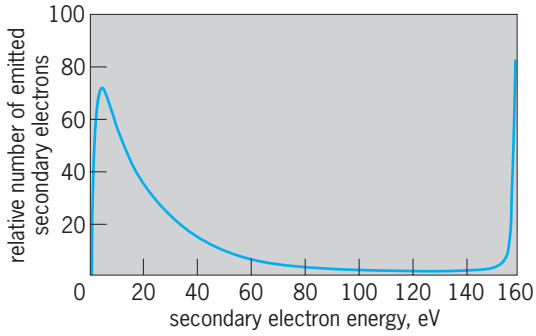


Fig. 6. Relative number of secondary electrons as a function of secondary energy for $E_p = 160$ V.

the energy for which the yield reaches its maximum value increases with increasing θ.

Secondary energies. A typical energy distribution of secondary electrons emitted by a silver target bombarded with primaries of 160-eV energy is given in **Fig. 6**. Note that most of the secondaries have relatively low energies. A small fraction of the emitted electrons have the same energy as the incident primaries and are called reflected primaries.

Alfred H. Sommer

Bibliography. R. L. Bell, *Negative Electron Affinity Devices*, 1973; A. J. Dekker, Secondary electron emission, in *Solid State Physics*, vol. 6, p. 251, 1958; A. Modinos, *Field, Thermionic, and Secondary Emission*, 1984.

Secondary ion mass spectrometry (SIMS)

An instrumental technique that measures the elemental and molecular composition of solid materials. Secondary ion mass spectrometry also provides methods of visualizing the two- and three-dimensional composition of solids at lateral resolutions approaching several hundred nanometers and depth resolutions of 1–10 nm. This technique employs an energetic ion beam to remove or sputter the atomic and molecular constituents from a surface in a very controlled manner. The sputtered products include atoms, molecules, and molecular fragments that are characteristic of the surface composition within each volume element sputtered by the ion beam. A small fraction of the sputtered atoms and molecules are ionized as either positive or negative ions, and a measurement by secondary ion mass spectrometry determines the mass and intensity of these secondary ions by using various mass analysis or mass spectrometry techniques. In this technique, the sputtering ions are referred to as the primary ions or the primary ion beam, while the ions produced in sputtering the solid are the secondary ions. Most elements in the periodic table produce secondary ions, and secondary ion mass spectrometry can quantitatively detect elemental concentrations in the part per million to part per billion range by using appropriate standards and specialized analytical conditions. Molecular detection limits by secondary ion mass spectrometry range between 0.1% and part-per-million concentration levels.

Dynamic and static SIMS. Secondary ion mass spectrometry analyses are divided into two broad categories known as dynamic and static. In the dynamic type, a relatively intense primary ion beam sputters the sample surface at sputter rates ranging 1–20 atomic layers/s (0.5–10 nm/s). This high sputter rate provides a very useful method for determining the in-depth concentration of different elements in a solid and is the most common method of secondary ion mass spectrometry analysis. Since the primary ion beam is eroding away the surface at a high rate, most molecular or chemical bonding information is rapidly destroyed, and hence the most common secondary ions detected in a dynamic secondary ion mass spectrometry analysis are elemental ions or clusters of elemental ions.

Static or molecular secondary ion mass spectrometry utilizes a very low intensity primary ion beam, and static analyses are typically completed before a single monolayer has been removed from the surface. Most static analyses are stopped before 1% of the top surface layer has been chemically damaged or eroded; under these conditions, molecular and molecular fragment ions characteristic of the chemical structure of the surface are often detected. Thus, static secondary ion mass spectrometry is best suited for near-surface analysis of molecular composition or chemical structure information, while dynamic secondary ion mass spectrometry provides the best technique for bulk and in-depth elemental analysis. *See* SPUTTERING.

Ion production. Secondary ions are formed by kinetic and chemical ionization processes in which sputtering is achieved by energy transfer from the primary ions to the solid surface. Typically, SIMS primary ions impact the surface with kinetic energies of 5–20 keV, and this energy is ultimately transferred to the sample atoms and molecules. The energy transfer initiates a collision cascade within the solid that ejects atoms and molecules at the solid surface-vacuum interface. A fraction of the ejected or sputtered species are in metastable or ionized states; however, most secondary ions formed within or near the surface are neutralized by unbound electrons in the solid before the ions can escape from the surface. However, metastable neutral atoms, that is, those neutral atoms having an internal energy above the ionization potential of the atom or molecule, can eject an electron above the sample surface and form a positive ion. Similarly, neutral atoms or molecules can abstract an electron from the solid, forming a negative secondary ion if the energy gained in the electron attachment exceeds the energy required to remove the electron from the surface. Kinetic sputtering processes are those in which the primary ion does not alter the electronic state of the solid. Kinetic sputtering is most commonly observed using primary ions of noble gases such as argon (Ar^+), krypton (Kr^+), or xenon (Xe^+). Chemical ionization processes occur in primary ion sputtering when the primary species alters the electronic state of the solid.

Two types of surface alterations are most commonly used in secondary ion mass spectrometry, and these modifications either reduce or increase the number of unbound electrons in the solid. For example, bombarding the surface with various oxygen ions (O_2^+, O^+, or O^-) reduces the number of conduction band electrons and favors the formation of positive ions. Similarly, bombarding a surface with cesium ions (Cs^+) increases the number of conduction band electrons and favors the formation of negative secondary ions. The use of either oxygen or cesium ion beams in dynamic secondary ion mass spectrometry is referred to a reactive-ion sputtering, because secondary ion yields are typically enhanced by one or more orders of magnitude over yields achieved with

unreactive primary species such as Ar^+. *See* CONDUCTION BAND; INERT GASES; ION.

Instrumentation. Instrument designs for secondary ion mass spectrometry range from custom-built, laboratory instruments to general-purpose, commercially available units. All instruments require a primary ion gun or column to generate and transport the primary ion beam to the sample surface, a sample chamber with sample mounting facilities, a mass spectrometer which performs mass-to-charge separation of the different secondary ions, and an ion-detection system. The complete instrument is typically housed in an ultrahigh vacuum chamber. *See* ION SOURCES; MASS SPECTROSCOPE.

Secondary ion mass spectrometry instrumentation is often distinguished by the method used to form secondary ion images. The ion microscope (see **illus.**) is used in a direct imaging technique that utilizes a relatively large primary ion beam ranging between 10 and 50 micrometers in diameter to sputter the sample surface. The secondary ions produced by this beam are extracted from the sample surface, mass analyzed, and detected. Secondary ion images mapping the lateral distribution of selected ions formed at the surface are acquired by transporting a two-dimensional image of the ions formed at the sample surface through the mass spectrometer and projecting a magnified image of these ions onto a two-dimensional ion detector. Typical image resolutions range from 1 to 5 μm from analysis of spots 100 μm in diameter. The lateral resolution of ion images formed by the ion microscope is determined by the quality of the secondary ion optics and the topography of the sample surface. Thus, smooth or flat samples typically produce higher-resolution images than rough surfaces.

The other mode of SIMS imaging is achieved by using an ion microprobe technique in which a microfocused primary ion beam is rastered about the sample surface, producing secondary ions at each raster position. Mass analysis of the secondary ions at each point in the raster, coupled with synchronization of the detection circuit with primary beam position, provides a method of generating secondary ion images at high spatial resolutions. The lateral resolutions of ion microprobe images are limited only by the spot size of the primary ion beam, and innovative source designs using liquid-metal ion guns routinely produce 50-nm spot sizes. Ion microprobe imaging thus provides the highest spatial resolution in secondary ion mass spectrometry analyses; however, ion microscope imaging is a faster imaging technique, since many pixels are irradiated per unit time.

A variety of mass spectrometers are used in secondary ion mass spectrometry instrumentation, and the most common include single- and double-focusing magnetic sector designs, quadrupole mass filters, and time-of-flight mass spectrometers. Magnetic sector and quadrupole mass filters are used most commonly in dynamic secondary ion mass spectrometry instruments, while time-of-flight spectrometers provide optimal performance for static secondary ion mass spectrometry applications. Both

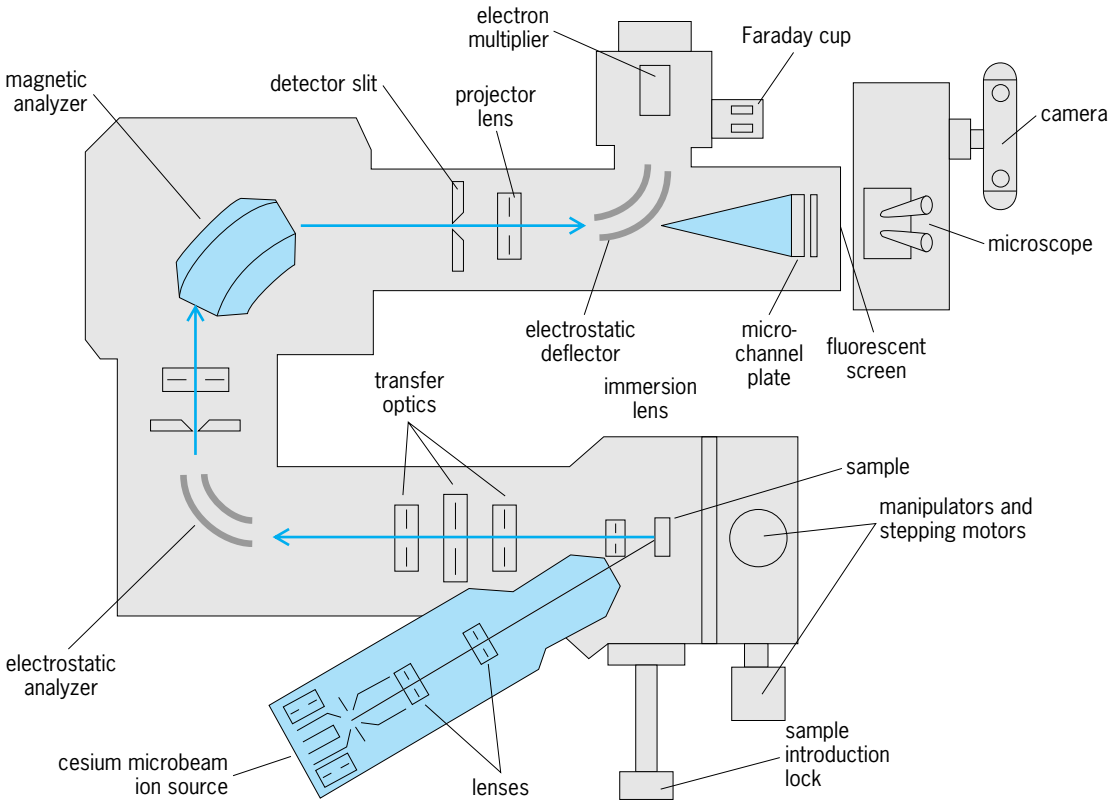


Diagram of an ion microscope.

magnetic sector and time-of-flight spectrometers can operate as either ion microscopes or ion microprobes, while instruments using quadrupole mass filters image almost exclusively with ion microprobe techniques. *See* TIME-OF-FLIGHT SPECTROMETERS.

Analytical applications. Both dynamic and static secondary ion mass spectrometry are used for analysis.

Dynamic. Dynamic secondary ion mass spectrometry has been successfully utilized in diverse applications, including the characterization of semiconductor materials, metals and metal alloys, catalysts, and geological and biological samples. The analytical issues are the detection and localization of specific elements on the surface or in the bulk of the materials. The most common application area of dynamic SIMS is in semiconductor analysis. Another important contribution of secondary ion mass spectrometry to materials science is the identification and localization of trace elements in metal grain boundaries, providing detailed insight into the chemistry of welds and alloys. Dynamic secondary ion mass spectrometry also has the unique ability to detect specific catalyst poisons on the surface and in the bulk of used or spent catalysts. In geological sciences, SIMS has been used to detect isotopic anomalies in the composition of various geological and meteoritic samples. Sensitive isotope ratio data have played a very important role in a more complete understanding of the age of the universe. Finally, dynamic secondary ion mass spectrometry has found extensive applications to the characterization of both hard and soft biological tissue.

The most important application of dynamic SIMS is the bulk and in-depth analysis of semiconductor materials. Specific analyses include detection of surface contaminants such as alkalis, alkaline-earth metals, and transition metals [titanium (Ti), chromium (Cr), iron (Fe), nickel (Ni), and zinc (Zn)] on silicon (Si) and gallium arsenide (GaAs) surfaces, high-sensitivity (bulk) analysis of boron (B) and oxygen in silicon, and in-depth analysis of semiconductor dopants such as B, phosphorus (P), and As. Essentially all of these studies employ reactive primary ion sputtering to achieve high detection sensitivity. Bulk analysis for specific impurities such as B or O in silicon are performed at high primary currents in order to maximize sputter rate and, hence, detection sensitivity. Bulk analyses generally do not provide useful image information since the erosion or sputter rates are high ($\sim$10 nm/s).

The most common dynamic SIMS application is depth profiling. In this technique the intensities of one or more secondary ions are determined as a function of sputter depth into the solid. Knowledge of the relative sensitivity factor of the implanted species and the total analysis depth permits converting the data from raw signal intensity versus sputter time into an elemental concentration versus depth. Depth profiling is used to determine the distribution and peak depths of ion implants into semiconductor solids. Ion implantation is a very common process for doping semiconductor materials with low levels of electrically active species such as B, P, and As. Doping via ion implantation is achieved by forming an ion of the dopant species, accelerating this ion to high energies (50–250 keV), and implanting the dopant ion into the solid at known doses (number of dopant atoms/cm^2). The most critical parameters in ion implantation are the peak depth, peak concentration, and shape of the implant profile. Since modern semiconductor devices require low implant peak concentrations ($<10^{19}$ dopant atoms/cm^3 or 0.1% concentration), dynamic secondary ion mass spectrometry is an ideal method for measuring and quantifying these ion implants.

Accurate depth profiling does not detect ions produced at the edges or walls of the sputter crater, since these signals could be generated over a range of depths. Crater wall discrimination is achieved either by mechanical or electronic aperturing of the secondary ion signal.

A typical example of such an analysis is determination of the C, Si, Al, and O distributions in a GaAs/Al-GaAs laser diode structure. A depth profile for this type of analysis can be achieved by using a Cs^+ primary ion beam and negative secondary ion detection. The raw data (secondary ion intensity versus sputter time) are converted to atomic concentration (atoms/cm^3) versus sputter depth (μm) by using ion implant standards and accurate crater-depth measurement. *See* SEMICONDUCTOR.

Static. Applications of static secondary ion mass spectrometry to analyses of the near-surface region of solids has become increasingly important as ever more sophisticated materials are developed. The near-surface region is generally defined as the top five monolayers (2–3 nm) of a solid. Examples of technology areas in which the chemistry of the near-surface region plays a critical role include high-performance glass coatings, liquid-crystal displays, manufacturing, semiconductor processing, biopolymer and biocompatible materials development, and polymer adhesives and coatings. The development of high-performance time-of-flight mass spectrometry techniques coupled with ion beam sputtering revolutionized the applications of static secondary ion mass spectrometry in many of these areas. The techniques combine the sensitivity of static secondary ion mass spectrometry with a high-transmission, high-mass-resolution instrument. In addition, time-of-flight secondary ion mass spectrometry performs high lateral resolution imaging by using either ion microscope or ion microprobe techniques. *See* MASS SPECTROMETRY; MATERIALS SCIENCE AND ENGINEERING; SEMICONDUCTOR DIODE.

Robert W. Odom

Bibliography. A. Benninghoven, F. G. Rudenaur, and H. W. Werner, *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications and Trends*, 1987; E. de Hoffmann and V. Stroobant, *Mass Spectrometry: Principles and Applications*, 2001; J. C. Vickerman, A. Brown, and N. M. Reed, *Secondary Ion Mass Spectrometry: Principles and Applications*, 1989.

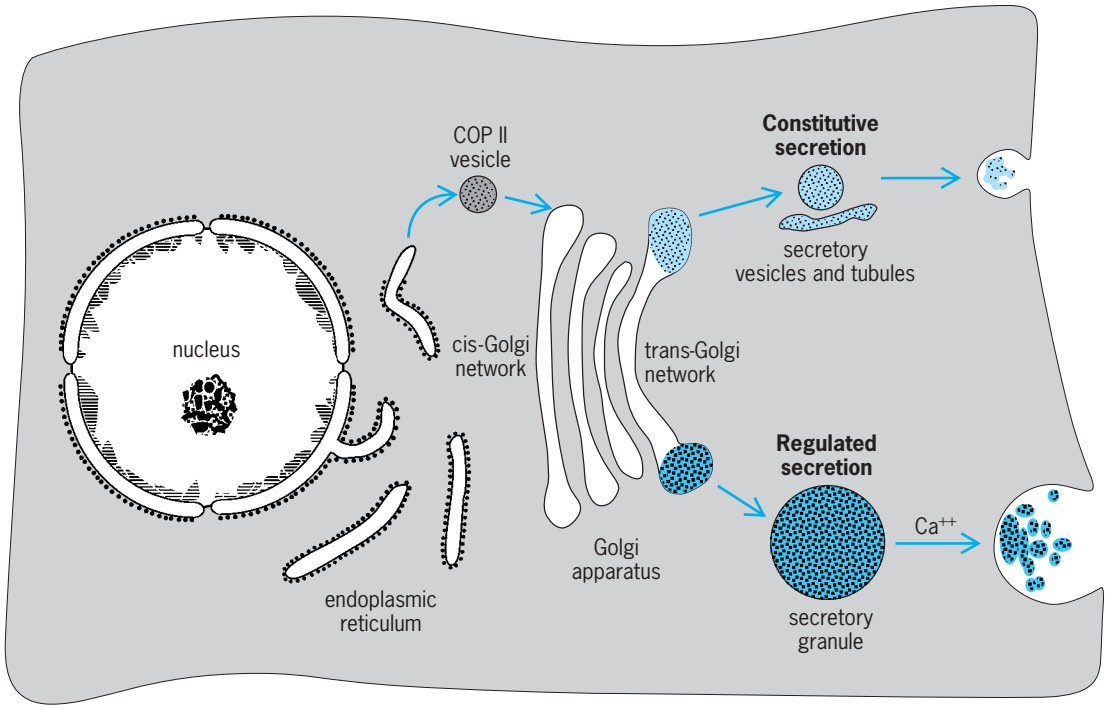
Secretion

The export of proteins by cells. With few exceptions, in eukaryotic cells proteins are exported via the secretory pathway, which includes the endoplasmic reticulum and the Golgi apparatus. Secreted proteins are important in many physiological processes, from the transport of lipids and nutrients in the blood, to the digestion of food in the intestine, to the regulation of metabolic processes by hormones. *See* CELL (BIOLOGY); CELL ORGANIZATION; ENDOCRINE MECHANISMS.

Synthesis on endoplasmic reticulum. Proteins destined for export or for insertion into the plasma membrane are synthesized initially on ribosomes attached to the outside of the rough endoplasmic reticulum, a portion of the membrane-bounded endoplasmic reticulum that is specialized for the synthesis of secretory proteins and most of the cell's membrane proteins. These proteins have a unique address tag called a signal sequence, usually located at or near the amino terminus, that ensures their recognition by a complex of proteins in the cytoplasm, known as the signal recognition particle, during their synthesis on the ribosomes. The signal recognition particle mediates the delivery of the growing polypeptides to the translocation apparatus. The translocation apparatus in turn mediates the transfer of the proteins across the membrane of the endoplasmic reticulum and into its interior space (lumen) or, in the case of membrane proteins, into the endoplasmic reticulum membrane itself. Proteins in the lumen of the endoplasmic reticulum or its membrane assume their appropriate three-dimensional structure, a process known as folding that is accelerated by chaperone proteins in the endoplasmic reticulum. *See* ENDOPLASMIC RETICULUM; PROTEIN; RIBOSOMES.

Export to Golgi apparatus. After they are folded, proteins to be exported enter small vesicles for transport to the Golgi apparatus (see **illus.**). The vesicles are coated with coat protein complex II (COPII) on their cytoplasmic surface, and they emerge from unique sites of the endoplasmic reticulum. The assem bly of the coat is thought to deform the membrane and facilitate the process of vesicle budding. Although the proteins making up the coat have been identified, it is not understood how secretory and membrane protein cargoes are concentrated within the COPII transport vesicles. Some glycoproteins, including the coagulation factors V and VIII, bind to a specific receptor protein, ERGIC-53, which acts to promote their incorporation into transport vesicles. Genetic mutations in ERGIC-53 lead to hemophilia (poor blood coagulation following injury) due to deficient secretion of factors V and VIII. *See* GLYCOPROTEIN; GOLGI APPARATUS; HEMOPHILIA; MUTATION.

The COPII vesicles are transported toward the Golgi apparatus in a process involving movement along microtubular tracks. As they approach the Golgi apparatus, they fuse together to form the *cis*-Golgi network, a tubulated structure on the *cis* side of the Golgi stack (the side closest to the endoplasmic reticulum). In the *cis*-Golgi network, proteins destined for further transport are segregated from



Secretory pathway of eukaryotic cells. Proteins are synthesized on ribosomes bound to the endoplasmic reticulum and subsequently transferred to the *cis*-Golgi network in COPII-coated vesicles. After traversing the Golgi apparatus, proteins are packaged into transport vesicles and tubules forming in the *trans*-Golgi network. In constitutive secretion, these vesicles and tubules subsequently fuse with the plasma membrane of the cell. In regulated secretion, newly synthesized proteins exit the *trans*-Golgi network and are routed to storage organelles called secretory granules. Secretory granule exocytosis occurs when an external signal is received by the cell and cellular calcium (Ca^{2+}) levels increase.

those whose predominant site of action is in the endoplasmic reticulum, such as ERGIC-53. ERGIC-53 and other endoplasmic reticulum proteins are recycled back to the endoplasmic reticulum by vesicular transport. The cargo is then transported across the cisternae of the Golgi stack, from the *cis* to the *trans* side, in a process that is dependent upon the formation of COPI-coated vesicles.

Sorting and transport to plasma membrane. When proteins reach the last cisterna of the Golgi, a highly tubulated region known as the *trans*-Golgi network, they are sorted and packaged again into transport vesicles, some of which are in the form of elongated tubules. From here, there are two pathways that proteins can take to the cell surface, depending on the cell type. Proteins can be transported directly to the plasma membrane (constitutive secretion) or to secretory granules (regulated secretion).

Constitutive secretion. In all cells, there exists a constitutive secretion pathway whereby vesicles and tubules emerging from the *trans*-Golgi network fuse rapidly with the plasma membrane (see illus.). The emerging vesicles and tubules can attach to microtubules, cytoskeletal elements emanating from the Golgi region, that accelerate their transport to the plasma membrane. In polarized epithelial cells, there can be two types of constitutive transport carriers that form in the *trans*-Golgi network.

Polarized epithelial cells have two plasma membrane domains, the apical and basolateral, that are separated by tight junctional complexes. The apical membrane domain faces the external milieu, lining the lumen of glands and the inside of the intestinal tract, for example. The basolateral surface is oriented toward the tissue interior and is in proximity to blood vessels. The transport carriers (vesicles and tubules), which contain distinct subsets of cargo molecules, will fuse with either the apical or basolateral plasma membrane domain. Signals in the cargo molecules themselves determine which type of transport vesicles, apical or basolateral, they enter. However, the receptors in the *trans*-Golgi network that sort the proteins into different vesicle populations have not been identified, nor have coat protein complexes been identified in the *trans*-Golgi network that might function as COPI and COPII do in the early secretory pathway. In some epithelial cell types, such as liver hepatocytes, nearly all of the protein traffic from the *trans*-Golgi network is routed to vesicles that fuse with the basolateral (sinusoidal) membrane. A subset of the membrane-bound proteins is subsequently redistributed to the apical (canilicular) plasma membrane in a process resembling endocytosis, known as transcytosis. *See* ABSORPTION (BIOLOGY); CELL MEMBRANES; ENDOCYTOSIS.

Regulated secretion. In cells that secrete large amounts of hormones, digestive enzymes, or enzymes that digest immune complexes and pathogens, most secretory and membrane proteins emerging from the *trans*-Golgi network are not immediately secreted, but are stored in membrane-bounded secretory granules (see illus.). Secretory granules forming in the *trans*-Golgi network contain large aggregates of content proteins. The aggregates are segregated away from proteins transported by constitutive transport vesicles. During the formation of secretory granules, hormone precursor proteins may be cleaved by enzymes to generate mature hormone molecules. Insulin, for example, is generated from its precursor, proinsulin, during the formation of secretory granules from the *trans*-Golgi network. Secretory granules release their contents into the extracellular space in a process known as exocytosis, when their membranes fuse with the plasma membrane. Exocytosis occurs only after the cell receives a signal, usually initiated by the binding of a hormone or neurotransmitter to a receptor on the cell surface. The receptor triggers a signal transduction cascade that results in increased concentrations of second messengers such as cyclic adenosine 3′,5′-monophosphate and phosphatidylinositol triphosphate. In most secretory cells, the second messengers or the hormone receptors themselves trigger the opening of calcium channels. Calcium ions stream into the cytoplasm through these channels, entering either from the extracellular milieu or from intracellular storage sites in the endoplasmic reticulum and secretory granules. Calcium initiates the docking of the secretory granules with the plasma membrane and the activation of the fusion apparatus (see illus.). *See* ENZYME; HORMONE; SIGNAL TRANSDUCTION.

Proteins destined for lysosomes also follow the secretory pathway (although they are not ultimately secreted). However, after reaching the *trans*-Golgi network, most are diverted into clathrin-coated vesicles. The clathrin-coated vesicles forming in the *trans*-Golgi network are similar to those formed during endocytosis at the cell surface, but contain a different set of the accessory proteins known as adaptors. Clathrin-coated vesicles fuse with endosomes after budding from the *trans*-Golgi network, and their cargo is subsequently delivered to lysosomes.

Fusion of transport vesicles with target membranes. Each step of transport along the secretory pathway requires that vesicles forming from one membrane-bounded organelle fuse with their correct target membranes. A specialized set of fusion proteins exists in the membrane of every vesicle population (v-SNAREs) and every target membrane (t-SNAREs). v-SNAREs form complexes with t-SNAREs, and these complexes mediate the fusion of the membranes of the vesicle and the target. Each type of vesicle has its own unique v-SNARE, and different target membranes contain different sets of t-SNAREs. The interaction between sets of v-SNAREs and t-SNAREs, a highly selective process, contributes to the specificity of vesicle fusion along the secretory pathway. The ability of the v-SNAREs and t-SNAREs to participate in membrane fusion events depends, in turn, on the activity of accessory proteins. One family of these proteins, which hydrolyze guanosine 5′-triphosphate, is known as the *rab* protein family. Different members of the *rab* family are responsible for activating SNARE interactions at each step during the transport of proteins from the endoplasmic reticulum to the plasma membrane. Thus, *rabs*

and other organelle-specific accessory proteins supplement the SNAREs in regulating the specificity of transport-vesicle targeting and fusion along the secretory pathway.

Unconventional secretion. In exceptional cases, proteins can be exported directly from the cytoplasm without using the secretory pathway. One such protein is fibroblast growth factor, a hormone involved in the growth and development of tissues such as bone and endothelium. Several interleukins, proteins that regulate the immune response, are also released via an unconventional route that may involve transport across the plasma membrane through channel proteins. These channels have adenosine 5′-triphosphatase (ATPase) enzyme activity and use the energy derived from the hydrolysis of ATP to catalyze transport. *See* ADENOSINE TRIPHOSPHATE (ATP); CELLULAR IMMUNOLOGY.

Secretion in prokaryotes. Bacteria also have the ability to export proteins, including toxins that cause illness in humans. Proteins exported via this route have signal sequences that are recognized by protein complexes in a manner analogous to recognition by signal recognition particle in eukaryotes. Proteins are directed during translation (or in some cases, posttranslationally) for insertion into pores in the plasma membrane in a process resembling that occurring in the endoplasmic reticulum of eukaryotic cells. In gram-negative bacteria, such as *Escherichia coli*, the exported proteins are found primarily in the periplasmic space between the inner and outer membranes. A similar pathway is used in bacteria by membrane-bound proteins and proteins that form cell walls. *See* BACTERIA.

Michael Rindler

Bibliography. Y. A. Chen and R. H. Scheller, SNARE-mediated membrane fusion, *Nat. Rev. Mol. Cell Biol.*, 2:98–106, 2001; J. Klumperman, Transport between ER and Golgi, *Curr. Opin. Cell Biol.*, 12:445–449, 2000; H. Lodish et al., *Molecular Cell Biology*, W. H. Freeman, 2000; G. Palade, Intracellular aspects of the process of protein synthesis, *Science*, 189:347–358, 1975; M. J. Rindler, Biogenesis of storage granules and vesicles, *Curr. Opin. Cell Biol.*, 4:616–622, 1992.

Secretory structures (plant)

Cells or organizations of cells which produce a variety of secretions. The process of secretion is a separation of a substance from the protoplast of a cell. The secreted substance may remain deposited within the secretory cell itself or may be released from the cell. Substances may be secreted to the surface of the plant or into intercellular cavities or canals. Some of the many substances contained in the secretions are not further utilized by the plant (resins, rubber, tannins, and various crystals), while others take part in the functions of the plant (enzymes and hormones). Secretory structures range from single cells scattered among other kinds of cells to complex structures involving many cells; the latter are often called glands.

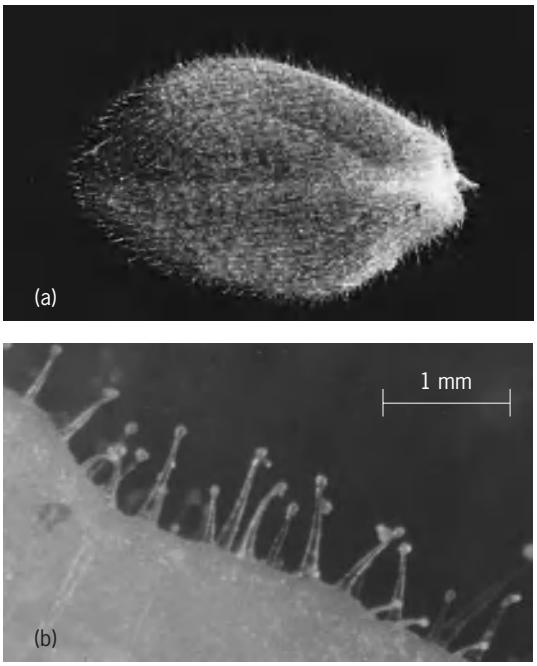


Fig. 1. *Petunia*. (*a*) Young leaf covered with glandular hairs. (*b*) Higher magnification of the hairs.

Glandular hairs. Epidermal hairs of many plants are secretory or glandular. Such hairs commonly have a head composed of one or more secretory cells borne on a stalk (**Fig. 1***a* and *b*). The shaggy hairs of winter buds of many trees produce sticky secretions that permeate and cover the buds with a protective film. The hair of a stinging nettle is bulbous below and extends into a long, fine process above (**Fig. 2**). If one touches the hair, its tip breaks off, the sharp edge penetrates the skin, and the poisonous secretion is released.

Nectaries. Glands secreting a sugary liquid—the nectar—in flowers pollinated by insects are called nectaries (**Fig. 3**). Nectaries may occur on the floral stalk or on any floral organ: sepal, petal, stamen, or ovary. The nectary may be flat, depressed, or padlike. Its surface layer, one or more cells deep, consists of

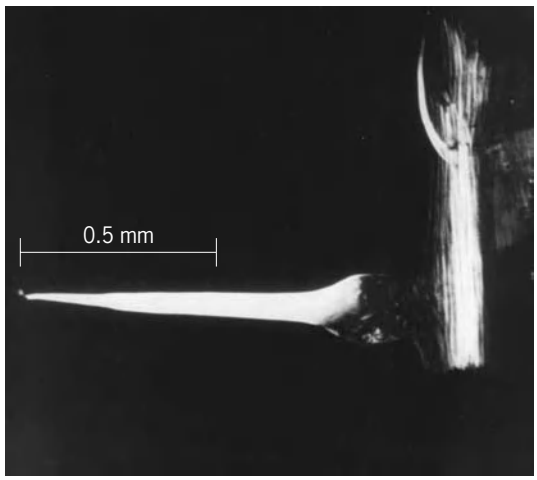


Fig. 2. Hair of stinging nettle.

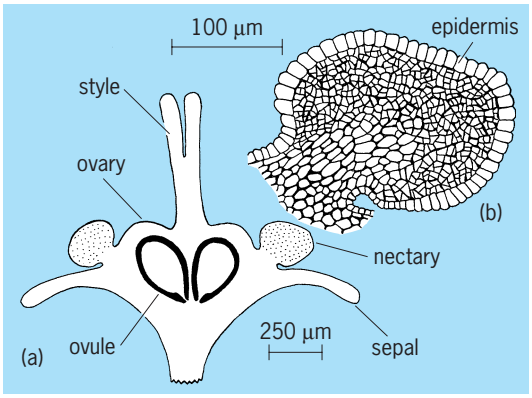


Fig. 3. *Ceanothus*. (*a*) Longitudinal section of flower. The nectary is ringlike and surrounds the ovary. (*b*) Enlarged view of a section of the nectary. The epidermis and the small cells beneath it are secretory.

secretory cells which usually have dense cytoplasm. Under the electron microscope these cells show few small vacuoles, a dense ground substance, numerous mitochondria, and a strongly developed endoplasmic reticulum, but the dictyosomes are not conspicuous. In the nectaries of monocotyledonous flowers, which occur in the partitions in the ovary (septal nectaries), the outer walls of the secretory cells develop numerous protuberances toward the interior of the cell. Since the plasmalemma lines these protuberances, its surface area is greatly increased. Thus a large secretory surface is provided. The vascular tissue, especially the sugar-conducting phloem, occurs close to the secretory tissue. The nectar may contain as much as 60% sugar.

Hydathodes. The hydathode structures discharge water—a phenomenon called guttation—through openings in margins or tips of leaves. The water flows through the xylem to its endings in the leaf and then through the intercellular spaces of the hydathode tissue toward the openings in the epidermis. Strictly speaking, such hydathodes are not glands because they are passive with regard to the flow of water. The pressure forcing the water to be discharged originates in the root. Some hydathodes, however, are like nectaries in having a layer of actively secreting cells.

Digestive glands. Some carnivorous plants have glands that produce secretions capable of digesting insects and small animals. These glands occur on leaf parts modified as insect-trapping structures. In the sundews (*Drosera*) the traps bear stalked glands, called tentacles. Each gland consists of four layers of cells, the innermost being in contact with the conducting tissue of the stalk. When an insect lights on the leaf, the tentacles bend down and cover the victim with a mucilaginous secretion, the enzymes of which digest the insect. In the Venus' flytrap (*Dionaea*) the insect is trapped by the sudden folding of the leaf. The upper surface of the leaf bears digestive glands and hairs acting as triggers for the folding mechanism. The traps in the pitcher plants are pitcher-shaped leaves that bear nectaries on the surface and inside the trap near its top; the digestive glands are deeper inside. Both consist of a layer of secretory columnar cells, a layer of rounded cells, and a layer of suberized cells. If an insect, attracted by the nectar, falls into the pitcher, it is prevented from escaping by rigid hairs pointing downward and is digested within the trap. The digested insects provide nitrogen that is essential for the plant's growth. Under the electron microscope the mucilage-producing cells are seen to contain large numbers of dictyosomes. The mucilage, a polysaccharide stainable with ruthenium red, occurs in the Golgi vesicles, which are derived from the dictyosomes. The vesicles separate from the dictyosomes and migrate toward the cell wall, frequently the anticlinal wall. The vesicle membrane and the plasmalemma open and join, and the contents of the vesicle are released into the space between the plasmalemma and the wall. The material then migrates along the wall toward the outside wall, where it is secreted by passing through the wall and cuticle. Dictyosomes do not appear to be involved in the production of digestive (proteolytic) enzymes. As judged by the scutellar cells of grass embryos, cells forming such enzymes show a strong development of the endoplasmic reticulum.

Resin ducts. Resin ducts are canals lined with secretory cells that release resins into the canal (**Fig. 4**). The canals are intercellular spaces that originate by separation of cells. Resin ducts are common in gymnosperms and occur in various tissues of roots, stems, leaves, and reproductive structures.

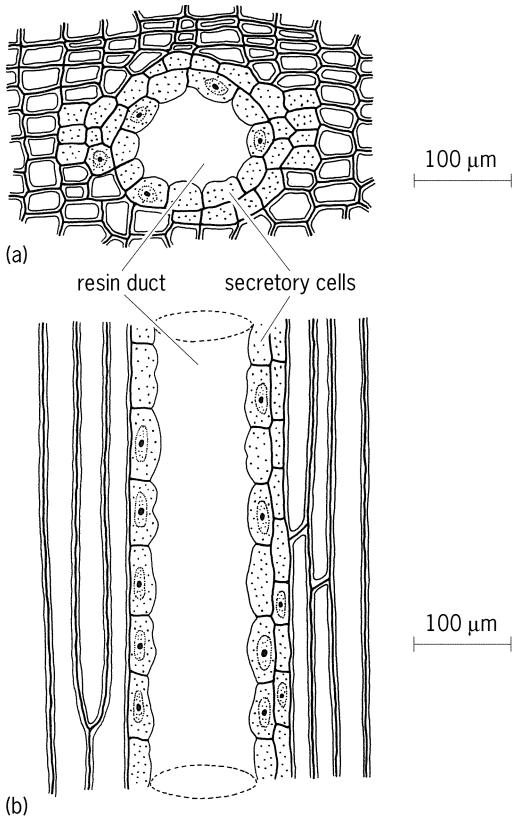


Fig. 4. Resin duct from pinewood in (*a*) transverse section and (*b*) longitudinal section.

They may arise during normal development or as a result of injury.

Gum ducts. These ducts are similar to resin ducts and may contain resins, oils, and gums. Usually, the term gum duct is used with reference to the dicotyledons, although gum ducts also may occur in the gymnosperms.

Oil ducts. Oil ducts are intercellular canals whose secretory cells produce oils or similar substances. Such ducts may be seen, for example, in various parts of the plant of the carrot family (Umbelliferae). In contrast to the oil ducts, the oil cavities of the kind found in the fruit rind and other plant parts of citrus result from a breakdown of oil-containing cells.

Laticifers. These are cells or systems of cells containing latex, a milky or clear, colored or colorless liquid. Latex occurs under pressure and exudes from the plant when the latter is cut. The single-cell laticifer is often much branched (**Fig. 5**); a few laticifer cells originating in the embryo may branch and invade all newly produced plant parts, forming the entire laticifer system of the mature plant. Laticifers derived from more than one cell result from breakdown of walls between adjacent cells; they are sometimes called laticiferous tubes. Both kinds of laticifers have multinucleate protoplasts. The latex is the product of this protoplast, but its relation to the cytoplasm seems to be more complex than that between the vacuole and cytoplasm in ordinary cells. The latex of some plants has a high content of rubber and is used as the source of natural commercial rubber. Examples of rubber-yielding plants are the Brazilian (*Hevea*) and Indian (*Ficus*) rubber trees. *See* INSECTIVOROUS PLANTS; PLANT METABOLISM; PLANT TISSUE SYSTEMS.

Katherine Esau

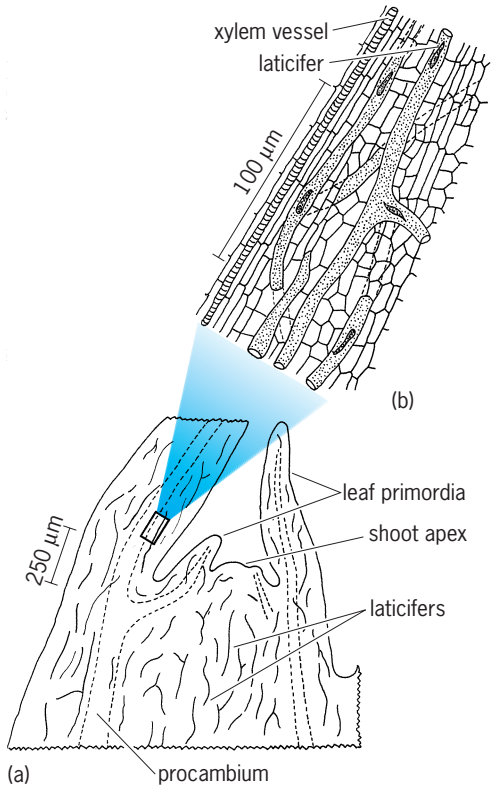


Fig. 5. Laticifers of *Nerium oleander*. (*a*) Longitudinal section of shoot tip. The wavy solid lines represent parts of laticiferous cells, which are arranged irregularly. (*b*) Enlarged view of the area in the rectangle in *a* showing parts of laticifers with cytoplasm and nuclei.

Sedative

A medication capable of producing a mild state of inhibition of the central nervous system (CNS) associated with reduced awareness of external stimuli. Numerous pharmacologic agents can induce different degrees of sedation, depending on the following variables: dosage; route of administration; absorption, metabolism, and excretion rates of the compound; specific receptor sites in the central nervous system that are affected by the agent; environmental setting; and state of the patient. *See* CENTRAL NERVOUS SYSTEM.

Types. Ethanol was probably the first sedative compound and was widely used for its analgesic and hypnotic properties, as well as for its ability to decrease inhibitory anxiety with resultant relaxation and occasional euphoria. Subsequent sedative compounds such as the barbiturates (for example, phenobarbital and secobarbital) and the benzodiazepines (for example, diazepam and alprazolam) also affect the same γ-aminobutyric acid (GABA) complex inhibitory receptor area in the central nervous system as ethanol does; however, they occupy different sites of this complex. Although these sedative compounds possess properties of tolerance and habituation, they vary in their addictive potential according to specific receptor sites and the particular type of patient. For example, in one controlled study, alprazolam was shown to be more likely to produce euphoric responses in alcoholics and their male offspring than a placebo or buspirone (an anxiety-reducing drug that is not a benzodiazepine). In other clinical studies, it has been noted that drug abusers and alcoholics appear to be more likely than other populations to abuse benzodiazepines. *See* ADDICTIVE DISORDERS; BARBITURATES.

In addition to those compounds that affect the GABA complex, other classes of chemical agents are used as sedatives. The antihistamines as well as some antidepressant drugs possess sedative side effects in addition to their primary pharmacologic properties. Since these classes of compounds are not addicting, they can be safely used as hypnotics. *See* TRANQUILIZER.

Effects. The sedative drugs that act on the GABA receptor sites, including GABA, are effective anticonvulsants and show cross-tolerance. Thus, they can be interchanged to treat their respective withdrawal syndromes. The benzodiazepines are useful in treating patients with the psychiatric diagnosis of generalized anxiety disorder. However, sedative side effects and possible interference with the ability to sustain attention and manual dexterity are possible. The duration of the side effects and withdrawal symptoms is related to the half-life of each of these sedative

compounds. Those sedative agents with short half-lives that are used to induce sleep can produce rebound insomnia when discontinued, and the long-half-life hypnotics are more likely to cause cognitive deficits on the day following administration. Cognitive impairment and drug dependence with subsequent withdrawal symptoms are more likely to occur if the dosage of the benzodiazepine is high, if use is of more than 4 months' duration, and if the patient is elderly or has had a prior history of alcohol or drug dependence. High-potency, short-half-life compounds appear to be more capable of producing pharmacologic dependence. *See* ANXIETY DISORDERS.

Donald M. Gallant

Bibliography. American Psychiatric Association Task Force Report on Benzodiazepine Dependency, Benzodiazepine dependence, toxicity, and abuse, *Amer. J. Psychiat.*, 1990; D. A. Ciraulo et al., Parental alcoholism as a risk factor in benzodiazepine abuse, *Amer. J. Psychiat.*, 146:1333–1335, 1989; H. Y. Meltzer (ed.), *Psychopharmacology: The Third Generation of Progress*, 1987.

Sedentaria

A group of 28 or more families of polychaete annelids in which the anterior, or cephalic, region is more or less completely concealed by overhanging peristomial structures, or the body is divided into an anterior thoracic and a posterior abdominal region; the pharynx or proboscis is usually soft and epithelial, lacking hard jaws or paragnaths. *See* POLYCHAETA.

Orbiniidae. This family has 13 genera and about 80 species. They are transitional between Errantia and Sedentaria in that the prostomium is exposed (**Fig. 1**); the thorax and abdomen are weakly separable; and the pharynx is muscular in the more primitive members and epithelial in the more developed members. There are two subfamilies: the Protoariciinae with 6 genera, chiefly from the Southern Hemisphere; and the Orbiniinae with 7 genera, including large, well-known species from worldwide areas in shallow to great ocean depths. The best known are the species in *Naineris*, *Orbinia*, *Scoloplos*, and *Phylo*. *See* ERRANTIA.

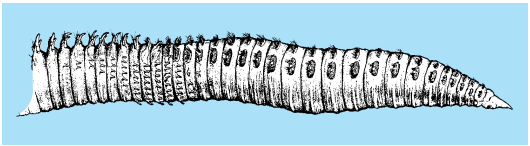


Fig. 1. *Orbinia* (Orbiniidae) in right lateral view.

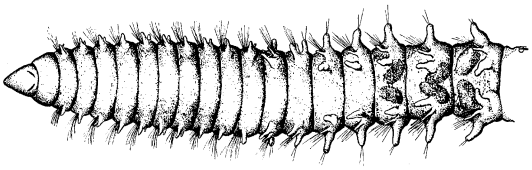


Fig. 2. *Paraonis* (Paraonidae) in dorsal view.

Paraonidae. This family has 7 genera and about 30 species, most members being small to minute and slender and inconspicuous (**Fig. 2**). They occur in all seas, often to great depths, and sometimes in tremendous numbers. Most have been named in recent years and come from deep seas, for example, the genera *Paraonis* and *Aricidea*.

Spioniform worms. The next 9 families make up the spioniform worms, characterized by the presence of a pair of short to long, grooved palpi near the mouth.

Apistobranchidae. This family has 1 genus and 2 species; both are rare and come from the North Atlantic Ocean and southern California.

Spionidae. This family has about 24 genera and more than 180 species. Some members are economically important; for example, *Polydora* drills in mollusk shells and is an oyster pest, and others foul bottoms of ships and ocean piers. The best-known genera are *Boccardia*, *Laonice*, *Nerine* (**Fig. 3**), *Polydora*, *Prionospio*, *Spiophanes*, and *Spio*. Most of these occur in intertidal or shallow seas and a few in abyssal depths.

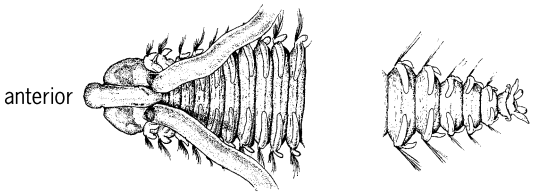


Fig. 3. *Nerinides* (Spionidae) showing the anterior portion and the tail section, in dorsal view.

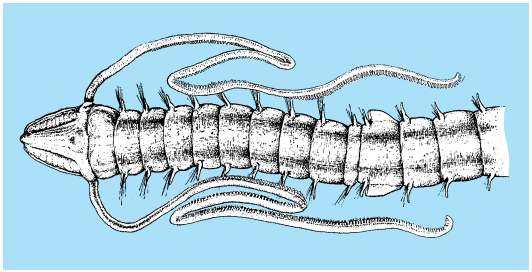


Fig. 4. *Magelona* (Magelonidae) showing the thorax and three abdominal segments, in dorsal view.

Magelonidae. This family has 1 genus (*Magelona*) and 14 species; they have a worldwide distribution, usually in shallow depths (**Fig. 4**).

Disomidae. This family has 2 genera and 11 species, occurring in cosmopolitan areas in the shelf and at great ocean depths; they include the *Disoma* and *Poecilochaetus*, the latter placed in a separate family by some authorities.

Heterospionidae. This family has 1 genus and 3 species, living in shallow and abyssal depths in the Atlantic and Pacific oceans.

Chaetopteridae. This family has 6 genera and about 31 species (**Fig. 5**). The best-known genus, *Chaetopterus*, is cosmopolitan, occurring in polar and tropical latitudes; its pelagic larvae survive long transport in oceanic currents. Other genera are *Telepsavus*, *Phyllochaetopterus*, and *Mesochaetopterus*.

Flabelligeridae. This family, the cage worms (**Fig. 6**), has at least 15 genera and 110 species. The anterior part of the body is often concealed by a setal cage, formed by setae of the first few segments. The body surface is papillated or covered with mud, and the blood is green. Common shallow-water genera are *Pherusa* and *Flabelligera*. Deep-water genera include *Ilyphagus* and *Brada*. *Flota* is pelagic throughout its life; it is covered by a thick mucoid sheath and has two kinds of setae. *Fauveliopsis* is an aberrant form which lacks the characteristic cephalic structures; it may therefore represent another family.

Poeobiidae. This family has a single, pelagic genus (*Poeobius*) and species from the North Pacific Ocean (**Fig. 7**). It lacks setae but in other respects resembles the flabelligerids.

Psammodrilidae. This family has 2 genera and 2 species, both occurring in the littoral sands in France. The length is 0.04–0.24 in. (1–6 mm) and the characters are aberrant in that the body resembles that of Protodrilidae, but the neuropodial hooks resemble those of Sedentaria. *See* ARCHIANNELIDA.

Fringe worms. This group includes 3 families: the Cirratulidae, Cossuridae, and Ctenodrilidae; they are characterized by long, tentacular structures along the side of the body. The Cirratulidae (**Fig. 8**) has 13 genera and about 120 species. They occur in all seas and at all depths and are important detritus feeders in coastal waters. *Dodecaceria* is associated with the construction of calcareous reefs or drilling in calcareous shells. Other genera are *Cirratulus*, *Chaetozone*, *Cirriformia*, and *Tharyx*. The Cossuridae (**Fig. 9**) have 1 genus (*Cossura*) and 8 species, occurring in worldwide areas. The Ctenodrilidae have 3 genera and 5 species; among the genera are *Ctenodrilus* and *Zeppelina*.

Limivorous (mud-swallowing) worms. Seven families make up this group, which is characterized by an eversible, pouchlike proboscis; some are tubicolous, and others occur in burrows or wander over the surface muds.

Scalibregmidae. This family has about 12 genera and 35 or more species; they are found chiefly in sublittoral and great depths, from slope to deep-ocean bottoms. Included in the genera are *Oncoscolex*, *Polyphysia*, *Scalibregma*, and *Sclerocheilus*.

Opheliidae. This family has 10 or more genera and more than 100 species; all are errantiate in habit or live in burrows without tubes. The most abundant and ubiquitous genus in intertidal zones is *Polyophthalmus*, which resembles a large nematode. Other well-known genera are *Ophelia*, the malodorous *Travisia*, *Armandia*, and *Ammotrypane*. An intertidal population of *Thoracophelia* in southern California was estimated to number about 158,000,000 individuals in a mile of beach sand and to total 7 tons (6.3 metric tons), wet weight; 14,600 tons (13,140 metric tons), of organic matter in sand has been cycled annually. Other limivorous annelids perform similar functions in recycling the sediments. *See* NEMATA.

Sternaspidae. This family has 1 genus, *Sternaspis*, sometimes called the gooseberry worm; it has several named species which are not clearly identifiable.

Capitellidae. Sometimes called bloodworms because of the dark-red color, this family has about 24 genera and more than 70 species (**Fig. 10**). They abound in intertidal mud flats and estuaries, and some are tolerant to varying salinities. Although chiefly mud-swallowers, some invade egg capsules of sea squids,

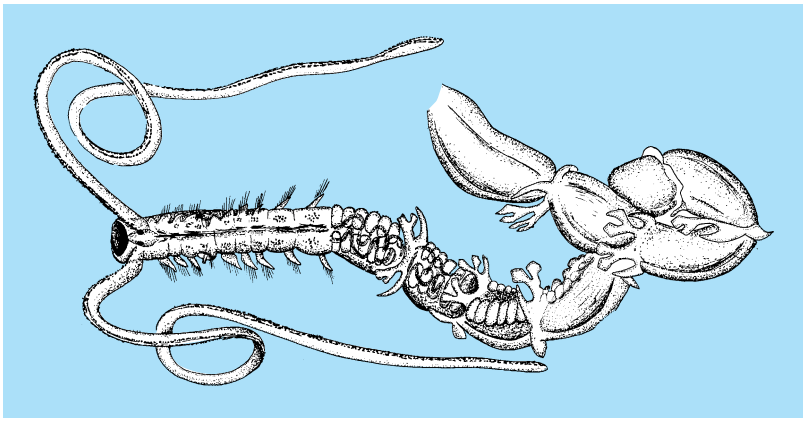


Fig. 5. ***Telepsavus*** **(Chaetopteridae) in dorsal view.**

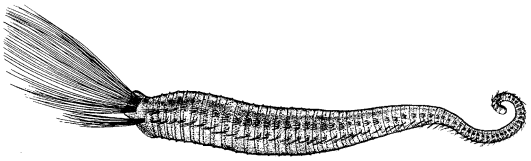


Fig. 6. ***Pherusa*** **(Flabelligeridae) or cage worm, shown in right lateral view.**

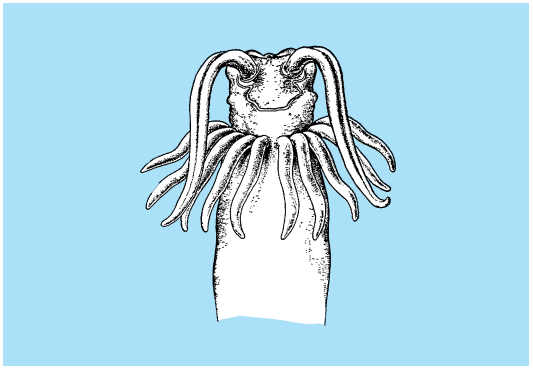


Fig. 7. Anterior end of ***Poeobius*** **(Poeobiidae) in dorsal view.**

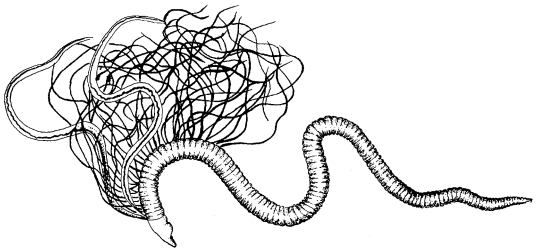


Fig. 8. ***Chaetozone*** **(Cirratulidae) in left lateral view.**

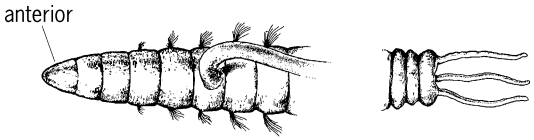


Fig. 9. ***Cossura*** **(Cossuridae) anterior and posterior ends in dorsal view.**

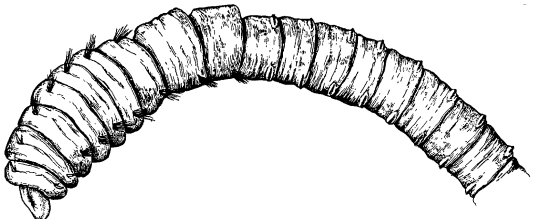


Fig. 10. ***Capitomastus*** **(Capitellidae) in left lateral view.**

and others occur in brackish and fresh water. One member is host to a monstrillid copepod. *See* COPEPODA.

Stypocapitellidae. This family takes its name from a monotypic genus for a small species from western Germany.

Arenicolidae. Known as the lugworms, this family has 4 genera and about 16 species. They are chiefly littoral, in sand and mud flats, where they form far-reaching colonies.

Maldanidae. Commonly called bamboo worms, this family has about 30 genera and 200 species, living in worldwide seas and in all oceanic depths (**Fig. 11**). They sometimes form beds of great extent in shelf and slope depths. Some are large and multisegmented; others have definite (often 19) setigerous segments and short bodies. Five subfamilies are generally recognized: Maldaninae have cephalic and anal plaques with the anal aperture located dorsally; common genera are *Asychis* and *Maldane*. Euclymeninae have well-developed plaques and an anal pore within the plaque; common genera are *Axiothella*, *Clymenella*, *Clymenopsis*, and *Clymenura*. The Lumbriclymeninae are less modified and primitive in some respects; they include *Lumbriclymene*, *Praxillura*, and *Notoproctus*. The Nicomachinae include *Nicomache*, and the Rhodininae have *Rhodine*, often with segmental flanges.

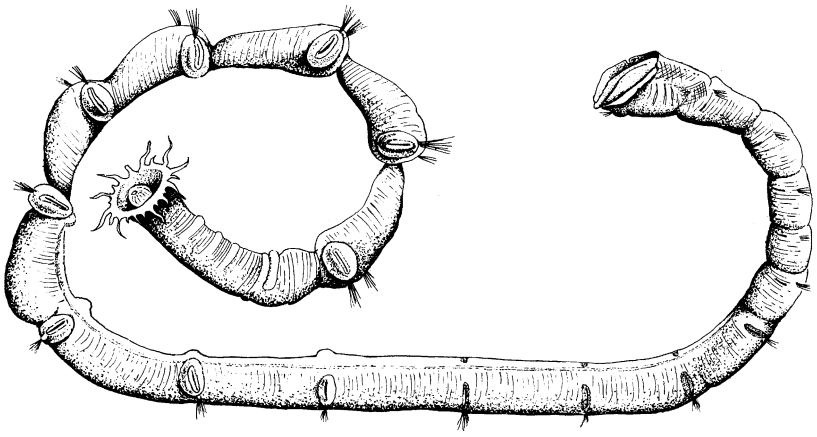


Fig. 11. ***Axiothella*** **(Maldanidae), in left lateral view, showing cephalic and anal plaques.**

Oweniidae. This family has 5 genera and about 20 species. *Owenia* is a shallow-water form, sometimes abundant in tropical inlets of sand. *Myriochele* occupies tapering tubes (**Fig. 12**) of characteristic structure, in which the animal is securely contained. *Myriowenia* (**Fig. 13**) is modified with a pair of long palpi. The remaining families are characterized by an opercular or tentacular crown which conceals head structures.

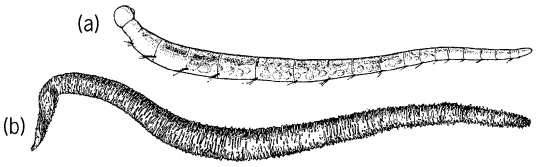


Fig. 12. ***Myriochele*** **(Oweniidae) with (*a*) entire ovigerous individual and (*b*) tube in right lateral view.**

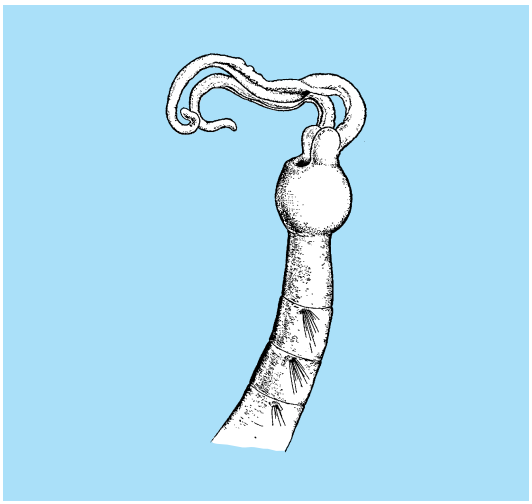


Fig. 13. ***Myriowenia*** **(Oweniidae) anterior end with long, grooved palpi, in left lateral view.**

Sabellariidae. Members of this family, called sand-cementing worms, have a compact operculum formed of setae of the first several segments (**Fig. 14**). The family has 7 genera and about 57 species, living at intertidal to abyssal depths in worldwide areas. Common genera are *Idanthyrsus*, chiefly tropical; *Phragmatopoma*, concentrated in the eastern Pacific; *Sabellaria*, well known from worldwide areas; and *Gunnarea* in South Africa. *Lygdamis* and *Phalacrostemma* are abyssal genera.

Pectinariidae. Commonly called cone worms, this family has 5 genera and about 46 species, occurring from intertidal sandy beaches to great oceanic depths. Common genera are *Cistenides* and *Pectinaria* from shallow bottoms and *Petta* from great depths.

Ampharetidae. This is a large, chiefly deep-water group, which includes more than 50 genera and at least 160 species. Some members occur in intertidal sands, and one group (*Hypania* and *Hypaniola*) occurs in the inland seas of Europe. They

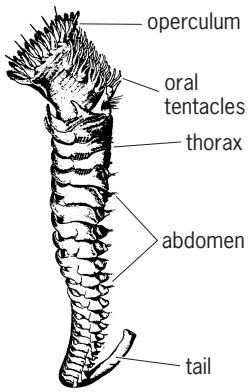


Fig. 14. ***Sabellaria*** **(Sabellariidae) in right lateral view.**

are characterized by retractile oral tentacles, unique cephalic structures, and highly modified parapodial parts. Most members construct tubes of specific design. The blood is green, and the body sometimes is highly colored. The family has 2 or 3 subfamilies: the Melinninae and Samythinae have a conspicuous dorsal membrane, with or without dorsal spines; and the Ampharetinae lack a dorsal membrane. Most members occur in slope and abyssal depths of the ocean; a few are intertidal, and *Hypania* and its allies are limited to the fresh waters of European inland seas. Included in the genera are *Amage*, *Ampharete*, and *Sabellides* in the Ampharetinae, and *Melinna* and *Melinnexis* in the Melinninae.

Terebellidae. This family has about 48 genera and more than 320 species. They are chiefly large, thick-bodied, tubicolous, rarely free-living worms, with the anterior end covered by a matted mass of tentacular cirri. Some harbor commensal polynoids in their tubes; others have commensal copepods. The family is usually divided into 4 subfamilies: the Amphitritinae with the genera *Amphitrite*, *Eupolymnia*, *Loima*, *Pista*, and *Terebella*; Polycirrinae with *Amaeana*, *Enoplobranchus*, and *Polycirrus*; Thelepinae with *Euthelepus*, *Thelepus*, and *Streblosoma*; and Artacaminae with *Artacama* and *Artacamella* (**Fig. 15**).

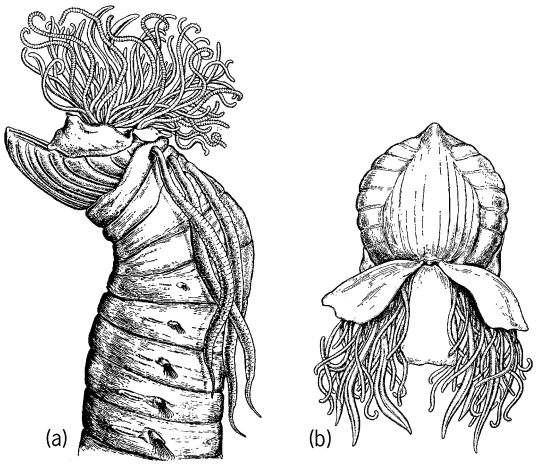


Fig. 15. ***Artacamella*** **(Terebellidae, Artacaminae). (*a*) Left lateral view. (*b*) Dorsal view.**

Trichobranchidae. This family has 5 genera and about 24 species, chiefly occurring at considerable ocean depths. *Terebellides* is the best known; all the other members of this family are rare and live at great ocean depths.

Sabellidae and Serpulidae. These 2 families make up the feather-duster worms. The Sabellidae have 31 genera and more than 257 species, often occurring in intertidal depths but descending to great abyssal depths. They are sometimes brilliantly colored, in conspicuous massed beds; some attain gigantic proportions, with a body length of 12–16 in. (30–40 cm); most are much smaller. The family has 3 subfamilies: The Sabellinae are the most numerous and the largest and include the genera *Eudistylia*, *Sabella* (**Fig. 16**), *Bispira*, *Hypsicomus*, and *Megalomma*. The Fabriciinae are chiefly small to minute, colonial, and intertidal or estuarine, but sometimes occur in deep water; common genera are *Chone*, *Euchone*, *Fabricia*, *Jasmineira*, and a fresh-water genus, *Manayunkia*, which is abundant in the Great Lakes district and its tributaries. The Myxicolinae are best known for *Myxicola*, which has been studied for its great neural fiber system.

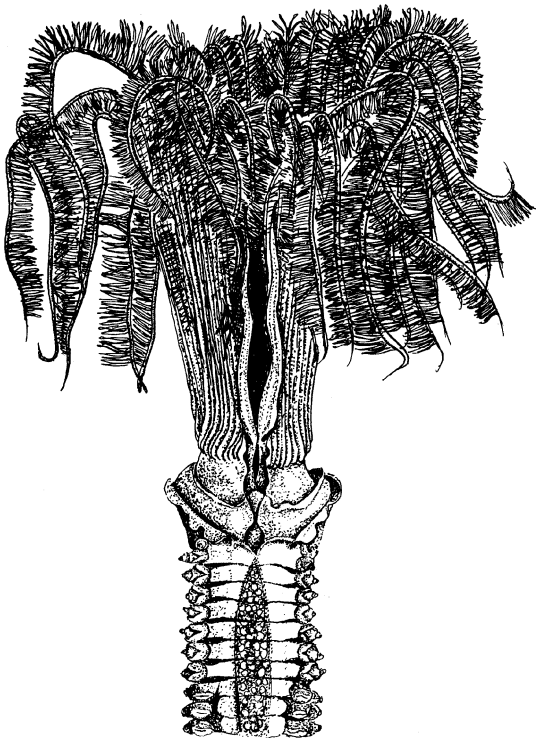


Fig. 16. ***Sabella*** **(Sabellidae) in dorsal view.**

The Serpulidae have more than 55 genera and 340 species, occurring in cosmopolitan areas and at all depths. They construct calcareous tubes, which may be so abundant as to clog drains and waterways. The peristomium is modified as a tentacular crown, with one filament modified as an operculum (**Fig. 17**). Common genera are *Crucigenia*, *Eupomatus*, *Hydroides* (a fouling form with cosmopolitan distribution), *Merciérella* (a euryhaline or estuarine form of

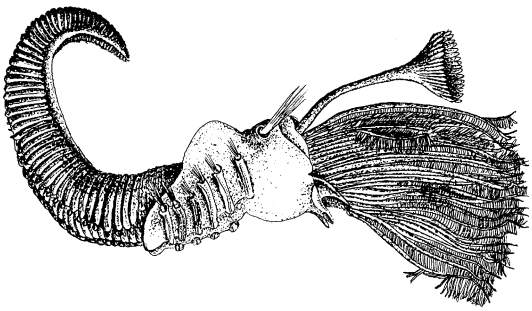


Fig. 17. *Serpula* **(Serpulidae) in right lateral view.**

economic importance), *Protula*, and *Serpula*. The Filograninae include *Filograna* and *Salmacina*. *See* ANNELIDA.

Olga Hartman

Bibliography. O. Hartman, *Catalogue and Index of the Polychaetous Annelids of the World*, Hancock Foundation, University of Southern California, 1959–1965; B. H. McConnaughey and L. F. Fox, The anatomy and biology of the marine polychaete *Thoracophelia mucronata* (Treadwell), Opheliidae, *Univ. Calif. (Berkeley) Publ. Zool.*, 47:319–328, 1949; S. P. Parker (ed.), *Synopsis and Classification of Living Organisms*, 2 vols., 1982; D. P. Wilson, On the Mitraria larva of *Owenia fusiformis delle Chiaje*, *Phil. Trans. Roy. Soc. London*, ser. B, 221:231–334, 1932.

Sedimentary rocks

Rocks formed of any of the materials that accumulate at Earth's solid surface at ambient temperatures, regardless of their consistency. Together with lava flows, sedimentary rocks form a cover of stratified material, known as the stratisphere, that, lying on igneous and metamorphic rocks, covers over two-thirds of the lands and nearly all of the ocean floor. Generally measured in hundreds to a few thousand meters, it locally reaches up to 15 km (9 mi) in thickness. *See* EARTH; IGNEOUS ROCKS; METAMORPHIC ROCKS; ROCK.

Most sediments accumulate as sand or mud deposited by air or water under the influence of gravity. Because of changes in the material supplied and in the conditions of deposition, the resulting accumulations are generally layered (stratified) in a near-horizontal position, with each stratum representing an event or episode (**Fig. 1***a*).

Derivation of sediment. In theory, the Earth's surface forms an oblate spheroid (geoid). The oceanic surface comes close to this. However, the brittle shell (lithosphere) of the solid Earth, floating on a viscous asthenosphere, rises to different elevations, thus deviating from this ideal. Its surface includes the vast deep-sea floor with an average elevation of 3800 m (12,464 ft) below sea level and a continental platform with an average elevation of 840 m (2755 ft) above sea level. The plane of reference or base level to which erosion is reducing the lands is sea level. Most sediment is transported to the oceans, where it is deposited, along with lava flows, building up the bottoms toward this base level. Other forces—tectonics and terrestrial volcanism—keep erosion and deposition from achieving these ends. Plate tectonics is always renewing parts of the ocean floor and subducting others into the Earth's interior, and is always raising new mountain ranges. Two other factors complicate the process. Parts of the continents sink progressively over long periods, trapping sediments before they reach the sea. Also, the sea level oscillates, flooding continents to deposit shallow-water marine sediments, and then withdrawing and exposing them to erosion. *See* ASTHENOSPHERE; DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; EROSION; LITHOSPHERE; PLATE TECTONICS; VOLCANO.

Weathering. The formation of sediment begins with weathering, the interaction of rock with the atmosphere. Some minerals are dissolved by rainwater. Others, such as the crystals of feldspar in granites, are decomposed by carbonic acid (H_2CO_3) in rainwater. These minerals break down into hydrous aluminum silicates (clay minerals), and release other metals and silicon dioxide (silica) to solution. This, along with heating and cooling, wetting, and drying, and frost action breaks up the rock. *See* CLAY MINERALS; GRANITE; SILICA MINERALS; WEATHERING PROCESSES.

This resulting residue of detritus is generally a mixture of clay and coarser fragments (clasts) subject to further weathering, and is known as the regolith. Organic compounds, such as wood and pollen grains, generally become incorporated. Clasts are also provided by surf tearing at sea cliffs, by the plucking and grinding of glaciers, and, rarely, by meteorite impact that blasts craters and ejects clasts, glass droplets, and vapor. *See* REGOLITH; SEDIMENTOLOGY.

Detrital transport and deposition. The regolith moves down slopes. At the same time, it becomes eroded by water, wind, and ice feeding dust to the air. Mostly the particles follow the gravipotential gradient to lower elevations. Some are trapped in subsiding continental basins, but most ultimately reach the ocean. Skeletal elements of organisms, added in transit and at depositional sites, may become incorporated and permanently preserved as fossils.

Some of this transport occurs en masse in mudflows or glaciers and in the bulk flow of aerial and aqueous suspensions (turbidity currents). However, most involves the movement of individual particles whose properties, above all, size, interact with fluid motion to sort them into separate populations. *See* TURBIDITY CURRENT.

Particles are classified primarily by their mean diameter (**Fig. 2**). Sand grains, pebbles, cobbles, and boulders are moved by traction—dragged, rolled, or bounced along the substrate—while the transport of clay is limited to suspension in the air, water, or ice. Deposition occurs as the wind or water slackens. Much of the clay begins in colloidal sizes, which coagulate into larger sizes. These coagulates tend to be filtered out by animals and generally are packaged with the animals' digestive waste into sand-grade pellets that, even in the deep ocean, reach the bottom in days or weeks. Dust particles in the air are caught by raindrops and delivered to the land or water surface.

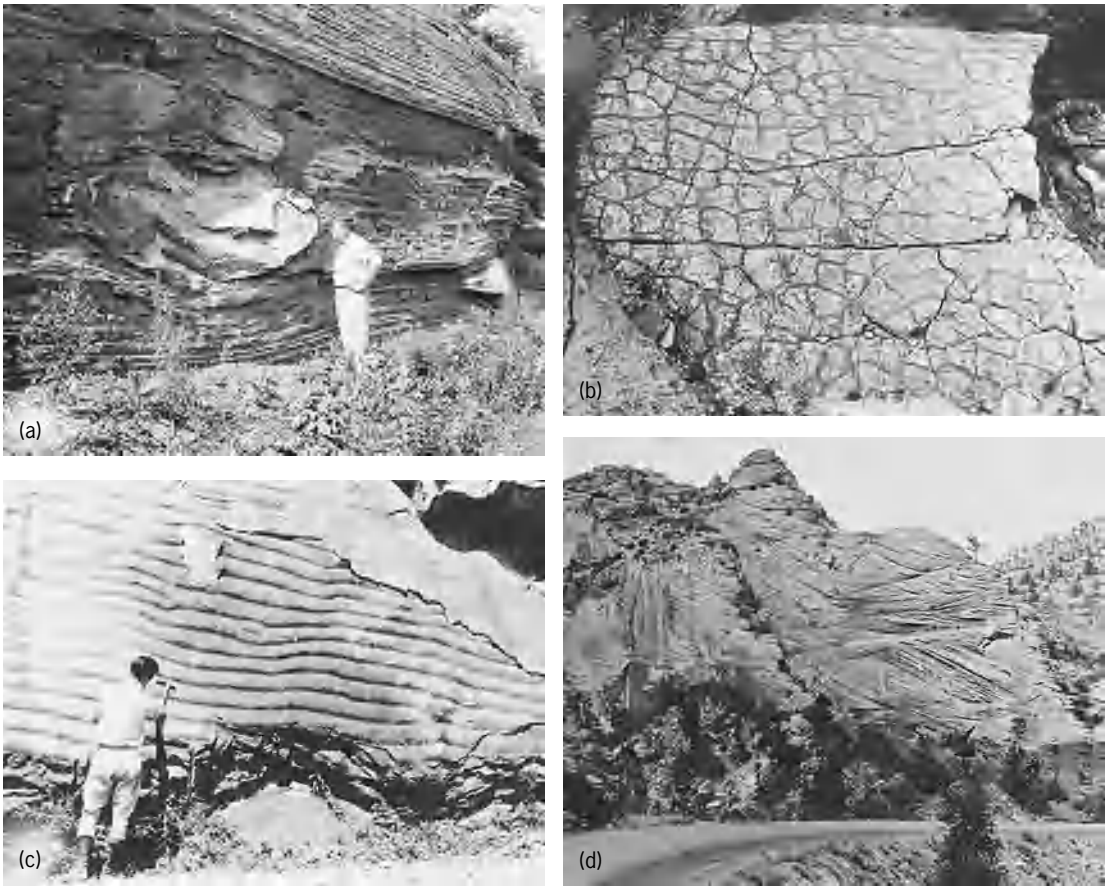


Fig. 1. Structures in sedimentary rocks. (*a*) Stratification in Carboniferous shales and sandstones near Montgomery, West Virginia, and a concretion. (*b*) Mud cracks in Lower Carboniferous limestone northeast of Bluefield, West Virginia. (*c*) Current ripples in Cambrian dolomite northeast of Bluefield, Virginia (*courtesy of Carl O. Dunbar*). (*d*) Dune cross-bedding in Jurassic Navajo Sandstone near entrance to Zion Canyon, Utah (*Virginia Division of Natural Resources*).

Hydrodynamics are not limited to the interaction between individual particles and the suspending fluid, but can involve bodies of water loaded with suspended sediment that travel en masse down subaqueous gradients as turbidity currents. These erode the sea floor as they gather speed and deposit sediment as they slow and spread out over deep-water plains. The resulting bed (turbidite) is texturally graded with the coarser particles, which settle out most rapidly, at the base and progressively finer particles above these. Millimeter-scale turbidites show a gradation from silt to clay, thus recording the settling-out of suspensions caused by a storm in a lake or marine basin. Centimeter-to-decimeter-scale beds, grading from sand to clay and traceable for thousands of km^2, record centurial or millennial land- or mud-stirring events such as earthquakes that triggered landslides or tsunamis. In addition, these beds characterize large bodies of sediment deposited in abyssal plains or trenches fronting continents or island arcs such as the flysch of the Northern Alps and similar deposits in most major mountain ranges. Events of truly extraordinary violence are recorded in meter-scale deposits, such as the Miocene turbidite with boulders in its base that has been traced for hundreds of kilometers in the Apennine Mountains, Italy. *See* MIOCENE; TURBIDITE.

Deposition may be temporary, with resumption of transport when the velocity of the transporting fluid increases. Thus detrital matter is moved in stages; for example, by streams into lakes and the sea, and then by waves and currents. Deposits along the way are temporary unless subsidence takes them below base level. Weathering continues throughout, and wear and tear reduces the clasts in size, rounding them and grinding up the softer minerals to gradually

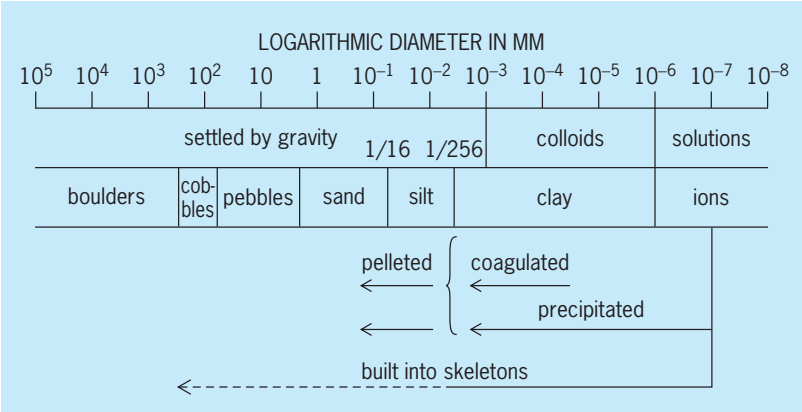


Fig. 2. Range of sedimentary particle sizes (Wentworth logarithmic scale). (*Adapted from W. Krumbein and L. L. Sloss*)

diminishing the less stable constituents in favor of quartz and clay minerals. *See* QUARTZ.

Transport and deposition of chemical sediments. Chemical sediments may derive from direct precipitation from sea water or from physical accumulation of skeletal matter precipitated by organisms. Solutes move with the water cycle. While dissolution occurs largely within the rock or regolith, the waters emerge in springs, flow as streams, and rest in lakes. Solutes are precipitated (Fig. 2) when concentrated to saturation to grow isolated crystals ranging in size from clay to sand or coarser grades, or to form aggregates of varying size, which may undergo mechanical transport or form solid crusts. In the last several hundred million years, most chemical sediments have derived from marine organisms, which extract calcium carbonate, silica, and calcium phosphate from water to build skeletons. After the organisms' death these skeletons may remain in place or become mechanically transported in the manner of detrital sediment.

Pyroclastic transport and deposition. Lumps of magma ejected from volcanoes freeze into ballistic volcanic "bombs" and help to build cones of sediment. Sand-grade crystal tuffs reach a wider distribution, and finer volcanic ash (microscopic shards) will travel over long atmospheric distances. Nuées ardentes, suspensions of glowing shards in hot gases, rush down the land surface to deposit a layer of consolidated ash (ignimbrite). Submarine eruptions yield the frozen glassy froth, called pumice, which contributes to deep-sea sediments and floats to beaches. *See* IGNIMBRITE; MAGMA; PUMICE; PYROCLASTIC ROCKS; TUFF.

Sedimentary structures. Bodies of sediment may be planar (sheetlike) as in Fig. 1*a* or pod-shaped as in dunes or reefs. Strata at the centimeter-to-meter scale are generally called beds, while those at the millimeter and micrometer level are distinguished as laminae. Typical dunes retain their pod shape, moving by transfer of sand from the windward to the lee side, where it accumulates in steeply inclined cross-laminae (Fig. 1*d*). Sand in water commonly forms ripples, which may grow into cross-laminated sheets (Fig. 1*c*). Most strata are formed by vertical accretion, resulting from a rain of suspended particles, and derive their stratification from changes in the particle size or composition of their components. *See* DUNE.

Polygonal shrinkage cracks (Fig. 1*b*) and large pores document the desiccation of ponds and intertidal flats. Animals leave tracks and trails on the surface of strata or may burrow for purposes of shelter or feeding in a process of bioturbation. In contrast to actual remains of organisms (body fossils), such structures are known as trace fossils. *See* FOSSIL.

Diagenesis. All sediments are overprinted by postdepositional (diagenetic) changes. Compaction results from the compression of the sediment under load, expelling pore water. It also tends to bring the platelike (platy) shapes of clay minerals into alignment with the bedding. Most mudstones show fissility parallel to bedding and are called shales. With burial to several kilometers, some clay minerals will give up some of their water, leading to further compaction. *See* DIAGENESIS; SHALE.

One of the first chemical overprints suffered by most sediments is chemical reduction. Weathering and transport generally occur in oxygen-rich environments, as does most deposition. But in most aqueous deposits (excepting the deep ocean floors where sediment supply is limited to wind-blown dust) the organic matter in the sediment exceeds the available oxidizing capacity. Bacteria use up the oxygen in the pore waters to decompose organic matter and decompose nitrate and sulfate to obtain more. The resulting oxygen deficit breaks down the ferric oxides, leading to the formation of sulfides and turning red and brown shades to green and black. *See* OXIDATION-REDUCTION.

Slow interaction between sediment and pore waters will gradually dissolve the less stable minerals, such as aragonite ($CaCO_3$) or magnesian calcite, to precipitate purer calcite or dolomite [$CaMg(CO_3)_2$] crystals. While these may replace the original grains, solution transfer may leave cavities to precipitate this solute to other sites, where the growth of crystals occludes pores and cements loose particles into solid rock. Solution transfer also dissolves load-bearing grains at points of contact (**Fig. 3**), precipitating on the walls of pores to convert porous and friable sandstone into hard, dense rock (solution welding). *See* ARAGONITE; CALCITE; DOLOMITE.

Local chemical gradients, such as caused by the decay of organisms, may induce the precipitation of calcite or siderite ($FeCO_3$) to form the dense nodules termed concretions (Fig. 1*a*). Large-scale chemical changes may dissolve disseminated fossils of opaline silica (sponge spicules, diatoms, radiolaria) to reprecipitate it as quartz in discrete nodules or beds of chert. *See* CHERT; CONCRETION.

In evaporite settings, magnesium ions may react with aragonitic calcium carbonate to form dolomite before burial. Large-scale conversion of calcite to dolomite occurs where evaporitic solutions penetrate buried limestone. In the Florida-Bahama platform, much of a Tertiary carbonate sequence up to 2 km (1.2 mi) thick has been converted to more coarsely crystalline dolomite. *See* LIMESTONE.

Of particular economic importance is the diagenesis of organic matter. Peat accumulated in bogs becomes compacted by burial and loses hydrogen in the form of methane to become first lignite, then bituminous coal, and, under high pressure, anthracite. *See* PEAT; COAL.

In lakes and the sea where phytoplankton are the main contributors of organic matter, bacteria in the sediment consume it with the available oxygen, including that obtained from nitrates and sulfates. Continuing anaerobically with release of methane, they convert the remaining organic matter into the waxy-appearing kerogen that is insoluble in most solvents. Burial to greater depths and higher temperatures results in the chemical cracking of kerogen, with the liberation of gases (such as methane, butane, and pentane) and liquid petroleum. *See* KEROGEN; PETROLEUM.

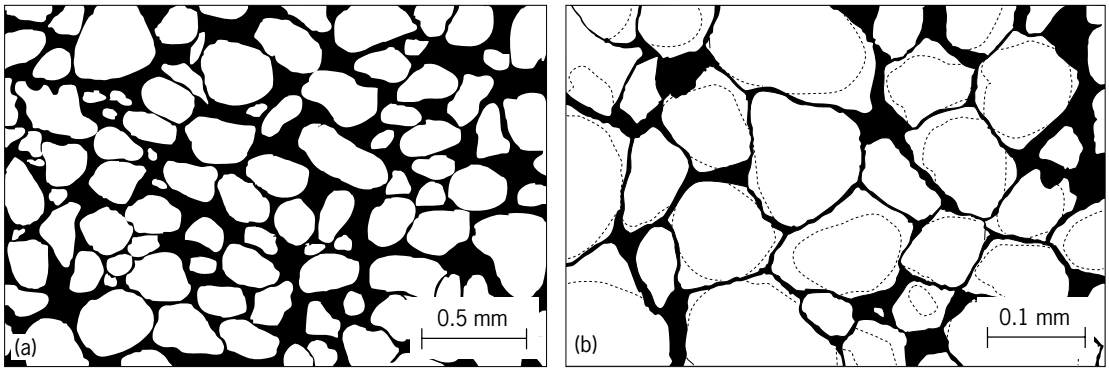


Fig. 3. Solution welding in sandstones. (*a*) Thin section of Jurassic Entrada Sandstone, Utah. The grains (white) are well sorted and rounded, and porosity (black) has been largely preserved. (*b*) Sandstone from Devonian Cedar Valley Formation, Iowa. Outlines of the original rounded grains, revealed by cathodoluminescence, are shown by thin lines. Dissolution of quartz at grain contacts has welded them together, while the silica, thus dissolved, was precipitated on the free surfaces in optical continuity with the sand grains to form a hard sandstone with reduced pore space. (*After P. A. Scholle, A Color Illustrated Guide to Constituents, Textures, Cements and Porosities of Sanstones and Associated Rocks, Amer. Assoc. Petrol. Geol., Mem. 28, 1979*)

Classification of sedimentary rocks. Sedimentary rocks are mixtures of various components, which in theory could be divided into detrital, chemical, and biogenic fractions. However, the breakup of skeletal matter and the diagenetic overprint make it difficult to assess the chemical versus biogenic components (making it necessary to lump them), while division of the detrital components into sand-silt and clay fractions remains apparent. Most sedimentary rocks are therefore assignable to a four-component system in which the end members are sand and silt (mostly quartz), carbonate, and nondetrital silica (chert). The mixtures of these can be visualized within a tetrahedron in which the great majority of rocks fall close to the sand-clay-carbonate triangle (**Fig. 4**). Individual strata that fall outside this scheme, mainly chemical precipitates, are dealt with as miscellaneous sediments.

Detrital sedimentary rocks. Rudites are detrital rocks dominated by clasts in the pebble-cobble-boulder range (Fig. 2). They include the loose gravels of stream or beach action, or their cemented equivalents (conglomerates). Rudites with sharply angular clasts, distinctive of short travel from their points of origin, are distinguished as sedimentary breccia, as distinct from breccias of tectonic origin that, cutting across strata, are not part of the stratisphere. *See* BRECCIA; CONGLOMERATE; GRAVEL.

Gravels are sought for road building and coarse aggregates in concrete. Deposited in settings of high water velocity, their finer accessories tend to be dense minerals, such as gold, silver, compounds of tin and uranium, and diamonds, which may form economically significant placer deposits.

Arenites or sandstones are abundant on the continental platforms. Here they may accumulate as dune sands, in streams, as beach sands, and in the shallows of lakes and seas. Turbidity currents carry sands to the depths of lakes and to the abyssal plains and trenches fronting the continents. *See* SANDSTONE.

Sandstones derived from immaturely weathered granites, retaining much feldspar, are termed arkose; those derived from metamorphic terranes, containing fragments of schist, slate, and greenstone are termed graywacke; while mature weathering preserving mostly the grains of quartz are called orthoquartzite. *See* ARKOSE; GRAYWACKE.

Texturally, sandstones may be divided into well-sorted grainstones whose open pores lend them porosity and permeability (not necessarily retained, as shown in Fig. 3*a*). In packstone, the presence of much fine material fills the interspaces between sand grains, while in wackestone the sand grains are floating in the finer matrix.

Porous and permeable sandstones are major sources of water and petroleum (gas, oil). Sandstone was widely used for building, and sand is an important aggregate for concrete and tarmac. Like rudites, sandstones serve as ores where they have concentrated silt-grade grains of heavy minerals such as gold and silver and minerals of tin and titanium.

Siltstone is a term generally reserved for silt-grade sedimentary rock that lack appreciable admixture

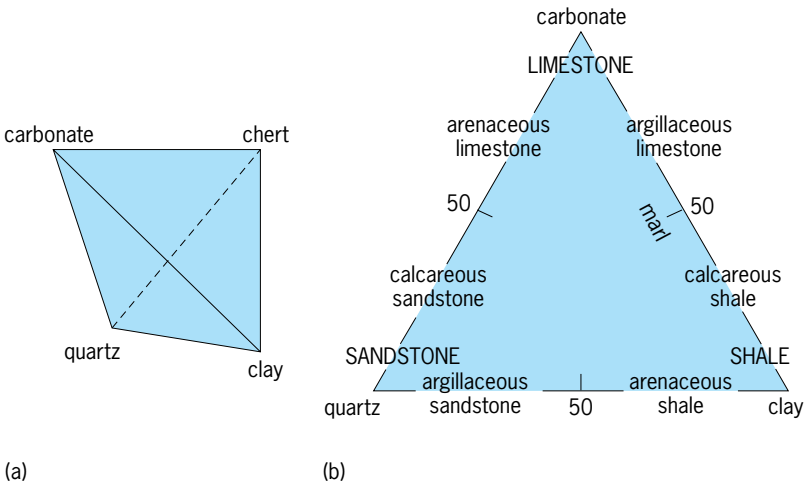


Fig. 4. Classification of the common sedimentary rocks. (*a*) Tetrahedron showing the main end members: detrital sand and silt (mainly quartz), clay fraction (mainly clay monerals), carbonates (generally a mixture of chemical and biogenic components), and siliceous chemical and biogenic components (chert). (*b*) Carbonate-quartz-clay face shown in detail. (*After W. Krumbein and L. L. Sloss*)

of clay. A notable example is loess, a deposit of wind-blown dust that accumulates on the lee side of deserts and peripheral to ice sheets and is prized for its agricultural properties. *See* LOESS.

Argillaceous or clay-bearing rocks (mudstones, shales) are the most abundant of all sedimentary rocks. They are initially plastic, but compaction turns them brittle and generally aligns their platy clay minerals to lend them a bed-parallel fissility, for which they are generally referred to as shales. Most are silty, and many are mixed with carbonate and then termed marls. Most extensive are the red clays of abyssal oceanic depths, consisting largely of the finest of wind-carried dust. *See* ARGILLACEOUS ROCKS; MARL.

Clays are an essential ingredient of portland cement, and are the raw materials for ceramics, from bricks to pottery. Some are mined as fuller's earth for filtering and other industrial purposes. Others, rich in zeolites, are mined for their base-exchange properties. Oil shales are rich in kerogen, and will yield petroleum on thermal cracking. Clays derived from regolith reduced to bauxite by intense tropical weathering are our source of aluminum. *See* BAUXITE; FULLER'S EARTH; OIL SHALE; ZEOLITE.

Carbonate sediments. Carbonates form more than 20% of sedimentary rocks. Rain contains dilute carbonic acid; hence waters becomes calcium-enriched in weathering. The air in soils, enriched in carbon dioxide (CO_2) due to organic respiration, adds to the calcium-carrying capacity of ground water. Where calcium carbonate–saturated waters reequilibrate with the atmosphere in caves, calcite is precipitated in hanging stalactites, standing stalagmites, and the gravity-defying helictites—all included in the term speleothems. In spring-fed water bodies, this reequilibration, aided by photosynthesis of plants, may precipitate calcite in porous but coherent (boundstone) aggregates. These calcareous tufas or travertines may dam streams and form waterfalls. Cyanobacteria and algae may coat pebbles with calcite and in lakes may build calcitic mounds (stromatolites). The photosynthesis of lacustrine phytoplankton may precipitate calcite mud (micrite or calcilutite) or mix with clay or marl. Skeletal elements of calcite include the egg capsules of the charophyte algae and the shells of mussels and snails. Buried whole or in fragments, these may accumulate as limestone, which may have been sorted into calcarenites divisible, like detrital sandstones, into carbonate grainstones, packstones, or wackestones. *See* CARBON DIOXIDE; STALACTITES AND STALAGMITES; STROMATOLITE; TRAVERTINE; TUFA.

Similar processes operate in the oceans. Due to the interference by magnesium, chemical precipitation at times, such as the present, yields the mineral aragonite rather than calcite. To depths of 4–5 km (2.5–3 mi), oceanic waters are generally carbonate-saturated. But as this saturation varies directly with pressure and inversely with temperature, water masses adjust their carbonate content by precipitation or dissolution. At greater depths (below the calcite compensation depth), the water is undersaturated and the carbonate constituents are dissolved. *See* MARINE SEDIMENTS.

Upwelling waters warmed on the Bahama Banks precipitate micritic aragonite muds, with or without aid from photosynthesis, or form round sand-size pellets (ooids) which, sorted by waves and currents, accumulate to form oolite, a rock found in all geologic ages. *See* OOLITE.

On tropical intertidal flats free of detrital matter, surficial mats of cyanobacteria and algae precipitate laminated and commonly crinkly laminae of micritic aragonite with characteristic large pores (fenestra), while in shallow water they produce mounds (stromatolites) in some restricted areas. Stromatolites were far more widespread in the geological past and dominated the carbonate deposits in the Precambrian seas. *See* PRECAMBRIAN.

Beginning in Cambrian time, some 550 million years ago, protists (for example, algae) and metazoans began to build skeletons out of calcite or aragonite and now dominate carbonate deposition. At present, the largest depositors are the prymnesiophyte algae (coccoliths) and the planktonic foraminifera of the oceanic plankton. *See* CAMBRIAN.

Deposition by organisms commonly produces a wide range of particle sizes. Storms in shallow waters commonly sort out shell beds (coquinas). *See* COQUINA.

Carbonate boundstones are chiefly produced in coral reefs and mainly held together by the colonies of red algae, corals, and other invertebrates. Surf and tsunamis break the reef rock into blocks that may accumulate as massive talus deposits of calcirudite. Mountain ranges, such as the Dolomites (in the southern Alps) or the Guadalupe Mountains of Texas-New Mexico, owe their massive cliff faces to ancient reefs bodies of boundstone and rudite, hundreds of meters thick. *See* TALUS.

Calcarenites include oolite, concentrations of sand-size skeletons such as the tests of foraminifera, the ossicles of echinoderms, or fragments of larger shells, and may form carbonate beach sands, as in the Florida Keys. *See* CALCARENITE.

Chalks are marine carbonate rocks of mudstone grade, consisting largely of coccoliths—complex plates of calcite secreted by prymnesiophyte algae—and of the tests of planktonic globigerinacean protozoans. Since the Late Jurassic, they have been the most widespread sedimentary rock on Earth, but have played a relatively small role on the continental shelves, except during Cretaceous times. *See* CHALK; CRETACEOUS; JURASSIC.

Carbonate rocks have great importance as reservoirs for water and petroleum. They are crushed for road gravel and smelting flux, and burnt for quicklime and portland cement. Many of the great buildings of England are made of the Jurassic limestones of Bath, and in the United States of the Carboniferous limestones of Indiana. Italian travertine is widely used as a building material. The Jurassic Solnhofen limestone of Germany provides lithographic stone. Much of the ornamental marble used worldwide is not true metamorphic marble but dense limestone, often fossiliferous such as the Devonian reef rocks of Belgium and the Jurassic Rosso Ammonitico of Italy. *See* MARBLE.

Siliceous chemical and biogenic sediments. Silica (SiO_2) is second only to carbonate in the solutes derived from weathering. In early times, it was presumably precipitated as silica gel and in silicates, but skeletal organisms since Cambrian time have extracted most of it to produce skeletal elements of opal, an amorphous and hydrated form of silica. These, in the form of sponge spicules and the microscopic tests of radiolarians and diatoms, cover large parts of the ocean floor with radiolarian and diatomic ooze. Diagenesis can transform these deposits of silt-grade opal into the lightly cemented diatomites, such as those of California's Miocene, with alteration to opal-CT (platy, dehydrated and crystalline), but ultimately converts them into dense, hard chert or flint, consisting of quartz. Diatomite is mined for filters, abrasives, and insulation, while flint was long vital to the lighting of fires and discharge of guns. *See* DIATOMACEOUS EARTH; OPAL.

Evaporites. In arid regions, evaporation of constricted water bodies raises solute contents to the level of precipitation. Evaporation of seawater sequentially yields a little carbonate, then much calcium sulfate (gypsum or anhydrite), then sodium chloride (halite), and then salts, mainly chlorides of potassium and magnesium. Gypsum is the source of plaster of paris, halite is the source of table salt and ice-clearing brines. Potash salts, such as mined in the Permian of Germany and New Mexico, are a major source of fertilizer. *See* ANHYDRITE; GYPSUM; HALITE; SALINE EVAPORITES.

Evaporites formed from continental salt lakes produce strata of different minerals such as the sodium carbonate mineral, trona, mined in the Eocene of Wyoming or the borax deposits in the Eocene of California.

Carbon-hydrogen sediments. In marshes, the remains of higher plants, such as Sphagnum moss and woody materials rich in cellulose and lignin, may accumulate where bacteria are inhibited by acidity. Diagenesis converts the resulting peat first into lignite, then into bituminous coal, and at high pressures into anthracite. These most abundant fossil fuels are mined in strata that range up to tens of meters thick, while the coal-associated methane is now produced as well.

While organic matter in marine settings rarely became a dominant rock component, the kerogen formed by diagenetic processes is sufficiently enriched in oil shales to constitute a potential source of petroleum for the future.

Miscellaneous chemical sediments. In phosphorites, apatite skeletons of certain brachiopods or vertebrate teeth and bones are generally associated with phosphatic pellets of sand to granule grade. They appear to be by-products of bacterial decomposition, which may enrich the phosphate content of pore waters to saturation. Such deposits form chiefly at sites of marine upwelling that localize organic productivity. Deposits of Late Cretaceous to early Tertiary times off North Africa or in Miocene times in Florida have given rise to large fertilizer industries. *See* APATITE.

Sedimentary iron ores are of several types. Bog iron ores of ochreous limonite form where reduced acid waters carrying dissolved ferrous iron meet the atmosphere, as in the bogs of New Jersey, and where bacteria make their living by catalyzing the oxidation reaction. Such ores were formerly widely smelted and served American colonists. *See* LIMONITE.

Oolitic iron ores are rocks of sand to granule-grade pellets of clay minerals of the chamosite group, variously altered to red hematite. Such ores have been widely mined for iron in the Silurian sedimentary rocks of North America, Jurassic sedimentary rocks of western Europe, the Cretaceous sedimentary rocks of North Africa, and the Oligocene sedimentary rocks of Colombia. *See* OLIGOCENE.

The largest sources of iron are the banded-iron formations of Precambrian age. At the time, the ocean and atmosphere lacked free oxygen, under which conditions much iron remained dissolved in the oceans. Photosynthesis by cyanobacteria provided an aerobic atmosphere, which then aerated the oceans and lead to precipitation of ferric iron oxides. Thick sequences of the banded-iron formation, consisting of alternating layers of hematite and chert, are now mined on all of the inhabited continents. *See* BANDED IRON FORMATION.

Manganese oxides occur in ores that parallel those of the oolitic and the banded iron ores, as well as in crusts and nodules formed on the deep-sea floor. *See* MANGANESE NODULES.

Related to the oolitic iron ores are the pelleted rocks composed of green glauconite. Normally accessory to marine sands, they may come to represent the bulk of the sedimentary rock, as in the Paleocene of New Jersey. Their texture and high potassium content have led to their use as a textural and fertilizing additive to soils. *See* GLAUCONITE; PALEOCENE.

Diamictites and intrusive sediments. Diamictites are doughlike mixtures of particles, ranging from clay to cobbles or boulders. As particles melt out of glaciers, they are kneaded under the moving ice to form deposits of glacial till. Diamictites can also be formed by landslides and in the ejecta blankets of large meteorite impacts. *See* GLACIAL GEOLOGY AND LANDFORMS.

The rapid burial of watery clays, especially if rich in hydrocarbons and thereby light and mobile, can lead to their penetration into the overlying sediments in domes that rise to the surface as mud diapirs. Extreme examples of these form in subduction zones, where sediments scraped off the underthrusting ocean plate become tectonically mixed with muds contributed from the continent to form melanges. These may erupt and flow out to form layers in the stratigraphic succession, layers that in Timor, Indonesia, can reach thicknesses of a kilometer and which represent diamictites containing exotic masses of stratigraphic successions of all sizes, up to hundreds of meters. *See* DIAPIR; SUBDUCTION ZONES.

The most common intrusive sediment bodies are salt domes. Having low density and yielding readily to pressure, halite can penetrate overlying rocks in the form of salt domes in cylindroid or tabular bodies with dimensions from hundreds to thousands of

meters. The updoming of the strata around salt domes has provided many highly productive oil pools. *See* SALT DOME.

Historical content. Each stratum represents an event or episode. Its composition and structure record physical and chemical happenings, and its fossils may reveal much more about the site of deposition. In the succession of strata, each is younger than the one it covers and older than the one that succeeds it, thus forming a link in a chain of history. These chains have recorded the evolution of life, the movements of the crustal plates, the numerous reversals in the Earth's magnetic field, changes in the fractionation of various isotopes in the oceanic reservoir, and climatic oscillations related to Earth's orbital variations. The layers of the stratisphere thus represent a vast archive of Earth history, the domain of stratigraphy. *See* GEOCHEMISTRY; GEOCHRONOMETRY; PALEONTOLOGY; ROCK MAGNETISM; STRATIGRAPHY.

A. G. Fischer

Bibliography. H. Blatt, *Sedimentary Petrology*, 2d ed., 1992; S. Boggs, *Principles of Sedimentology & Stratigraphy*, 3d ed., 2000; W. Krumbein and L. L. Sloss, *Stratigraphy and Sedimentation*, 1963; G. Middleton (ed.), *Encyclopedia of Sediments and Sedimentary Rocks*, Kluwer, Dordrecht, 2004; F. J. Pettijohn, *Sedimentary Rocks*, 3d ed., 1983; F. J. Pettijohn and P. E. Potter, *Atlas and Glossary of Sedimentary Structures*, 1964; M. E. Tucker, *Sedimentary Rocks in the Field* (*Geological Field Guide*), 2003.

Sedimentation (industry)

The separation of a dilute suspension of solid particles into a supernatant liquid and a concentrated slurry. If the purpose of the process is to concentrate the solids, it is termed thickening; and if the goal is the removal of the solid particles to produce clear liquid, it is called clarification. Thickening is the common operation for separating fine solids from slurries. Examples are magnesia, alumina red mud, copper middlings and concentrates, china clay (kaolin), coal tailings, phosphate slimes, and pulp-mill and other industrial wastes. Clarification is prominent in the treatment of municipal water supplies.

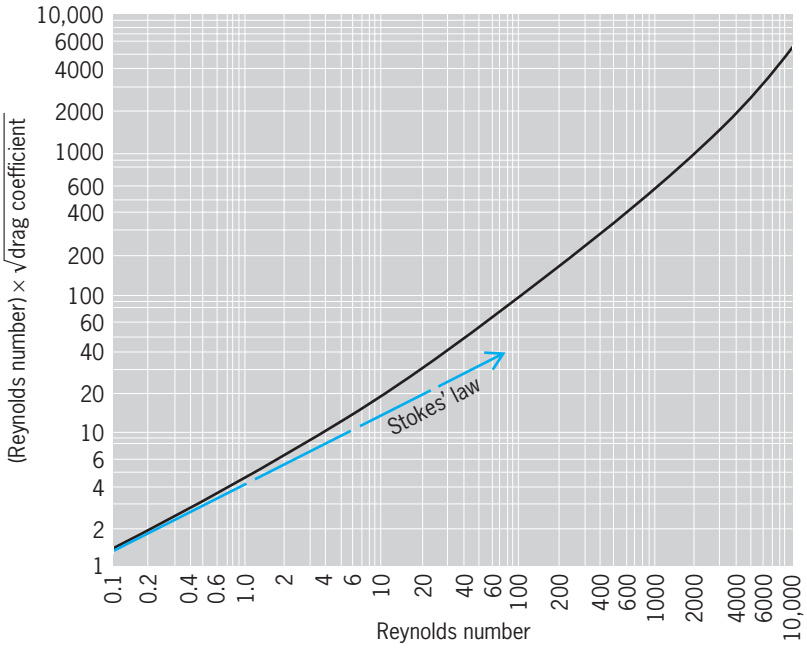


Fig. 1. Relation between the drag coefficient and the Reynolds number for spheres settling in fluids.

The driving force for separation is the difference in density between the solid and the liquid. Ordinarily, sedimentation is effected by the force of gravity, and the liquid is water or an aqueous solution. For a given density difference, the solid settling process proceeds more rapidly for larger-sized particles. For fine particles or small density differences, gravity settling may be too slow to be practical; then centrifugal force rather than gravity can be used. Further, when centrifugal force is inadequate, the more positive method of filtration may be employed. All those methods of separating solids and liquids belong to the generic group of mechanical separations. *See* CENTRIFUGATION; CLARIFICATION; FILTRATION; THICKENING.

Settling of spheres through fluids. The basic laws of physics that apply to sedimentation describe the resistance exerted by a liquid to the motion of a solid particle flowing through it. When a particle is released in a fluid in a force field, such as gravity, initially it accelerates. But the motion of the particle generates a frictional resistance which acts to attenuate the acceleration. Quickly, a balance is reached between the force moving the particle and the resisting force; then by Newton's law of motion, the acceleration drops to zero, and the particle proceeds at a constant velocity as long as the active force can act. This steady velocity, called the terminal velocity, prevails in most sedimentation processes.

The law of resistance of a solid particle moving through a fluid is complicated by many variables: the particle diameter and shape, the fluid density and viscosity, the particle velocity (generally its acceleration has a negligible effect), and the proximity of the subject particle to other particles and to equipment surfaces. The ideal situation is considered to be described by a single, smooth, spherical particle, sufficiently far from other particles, and from equipment surfaces, that it is unaffected by their presence. **Figure 1** shows how, for this situation, the terminal velocity u_t is influenced by the acceleration of gravity g, the diameter D_p and density ρ_p of the particle, and the density ρ and viscosity μ of the fluid. These variables are related by dimensionless groups (1) and (2). Figure 1 presents a particular relation between

$$\text{Reynolds number} = \frac{D_p u_t \rho}{\mu} \tag{1}$$

$$\text{Drag coefficient} = \frac{g D_p (\rho_p - \rho)}{\rho} \tag{2}$$

these two dimensionless groups which permits the terminal velocity to be calculated directly. The use of such a relation requires that the units used to

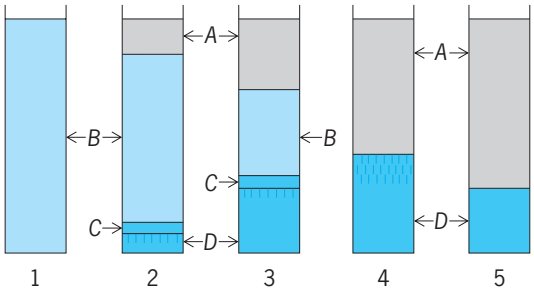


Fig. 2. Stages 1–5 in batch sedimentation of flocculated particles. Layers A–D are explained in the text. (*After W. L. McCabe, J. C. Smith, and P. Harriott, Unit Operations of Chemical Engineering, 4th ed., McGraw-Hill, 1985*)

describe each variable be self-consistent so that the two terms are indeed dimensionless. *See* DIMENSIONLESS GROUPS.

To use Fig. 1, for example, to predict a terminal velocity, the magnitude for the ordinate is calculated from known values of g, D_p, μ, ρ, and ρ_p. The corresponding value of the abscissa is read from the solid curve in Fig. 1, and the terminal velocity is calculated from this abscissa value.

Stokes' law. For small particles, or those moving in a very viscous liquid, the law of settling is governed by Stokes' law, which is Eq. (3). This equation

$$u_t = \frac{gD_p^2(\rho_p - \rho)}{18\mu} \tag{3}$$

is represented by the straight, broken line in Fig. 1; it applies for Reynolds numbers less than 0.3.

Sedimentation rates in practice. Particles too minute to settle at practical rates may form flocs by adding agents such as sodium silicate, alum, lime, and alumina. Because the agglomerated particles act like a single large particle, they settle at a feasible rate and leave a clear liquid behind.

A batch of flocculated pulp passes through several stages during sedimentation. **Figure 2** represents the original homogeneous flocculated pulp ready to settle. In stages 1–5 shown in Fig. 2, layer B is a uniform suspension of the same solid concentration as that of the original pulp, and layer A is clear supernatant liquid. Layer D consists of flocs resting lightly on one another, with liquid filling the voids between the flocs. Layer C is a transition layer, whose solid concentration varies continuously from that in layer B to that in layer D. As settling continues, layers C and B decrease and finally disappear. The end of this process is shown as stage 4 in Fig. 2. Subsequently, a new effect, called compression, begins. During compression, the weight of the deposit collapses the structure of the flocs, and some of the liquid in the flocs of layer D is expelled as small geysers. The thickness of layer D decreases to an equilibrium height called the ultimate height (Fig. 2, stage 5), and the process stops. *See* MECHANICAL SEPARATION TECHNIQUES; REYNOLDS NUMBER.

Vincent W. Uhl

Bibliography. W. L. McCabe, J. C. Smith, and P. Harriott, *Unit Operations in Chemical Engineering*, 6th ed., 2000; R. H. Perry and D. Green (eds.), *Perry's Chemical Engineers' Handbook*, 7th ed., 1997; J. F. Richardson, J. H. Harker, and J. Backhurst, *Chemical Engineering Volume 2 (Coulson & Richardson's Chemical Engineering)*, 5th ed., 2002.

Sedimentology

The study of natural sediments, both lithified (sedimentary rocks) and unlithified, and of the processes by which they are formed. Branches of this discipline have been known as sedimentation or sedimentary petrology, or they have been included in stratigraphy. Sedimentology includes all those processes that give rise to sediment or modify it after deposition: weathering, which breaks up or dissolves preexisting rocks so that sediment may form from them; mechanical transportation; deposition; and diagenesis, which modifies sediment after deposition and burial within a sedimentary basin and converts it into sedimentary rock. Sediments deposited by mechanical processes (gravels, sands, muds) are known as clastic sediments, and those deposited predominantly by chemical or biological processes (limestones, dolomites, rock salt, chert) are known as chemical sediments. **Figure 1** shows a diagram of the processes that are involved in the sedimentary cycle. *See* STRATIGRAPHY; WEATHERING PROCESSES.

Sedimentary cycle. The raw materials of sedimentation are the products of weathering of previously formed igneous, metamorphic, or sedimentary rocks. In the present geological era, 66% of the continents and almost all of the ocean basins are covered by sedimentary rocks. Therefore, most of the sediment now forming has been derived by recycling previously formed sediment. Identification of the oldest rocks in the Earth's crust, formed more than 3×10^9 years ago, has shown that this process has been going on at least since then. Old sedimentary rocks tend to be eroded away or converted into metamorphic rocks, so that very ancient sedimentary rocks are seen at only a few places on Earth. The area (or volume) of sedimentary rocks declines exponentially with age, and the half-life of a sedimentary rock is about 2×10^8 years. *See* EARTH CRUST; IGNEOUS ROCKS; METAMORPHIC ROCKS.

Major controls. The major controls on the sedimentary cycle are tectonics, climate, worldwide (eustatic) changes in sea level, the evolution of environments with geological time, and the effect of rare events.

Tectonics. These are the large-scale motions (both horizontal and vertical) of the Earth's crust. Tectonics are driven by forces within the interior of the Earth but have a large effect on sedimentation. These crustal movements largely determine which areas of the Earth's crust undergo uplift and erosion, thus acting as sources of sediment, and which areas undergo prolonged subsidence, thus acting as sedimentary basins. Rates of uplift may be very high (over 10 m or 33 ft per 1000 years) locally, but probably such rates prevail only for short periods of time. Over

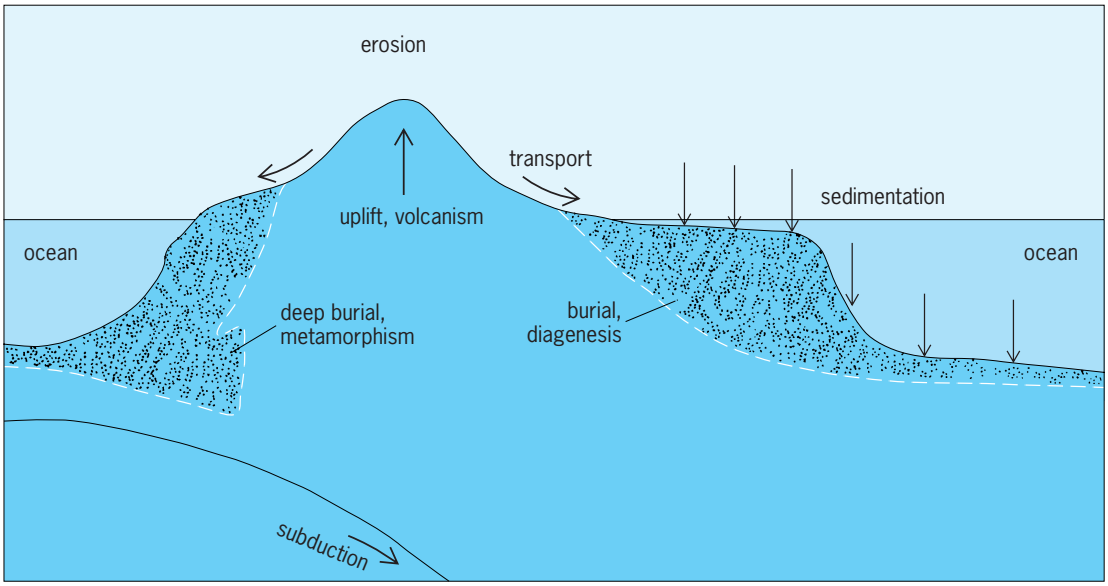


Fig. 1. Diagram showing processes that constitute the sedimentary cycle. Heavy arrows indicate mass fluxes of material. Light arrows indicate where deposition is taking place.

millions of years, uplift even in mountainous regions is about 1 m (3.3 ft) per 1000 years, and it is closely balanced by rates of erosion. Rates of erosion, estimated from measured rates of sediment transport in rivers and from various other techniques, range from a few meters per 1000 years in mountainous areas to a few millimeters per 1000 years averaged over entire continents. Though these rates are very slow by human standards, over periods of millions of years they produce immense quantities of sediment; all the continents would soon be reduced to sea level if it were not for the renewed uplift produced by tectonics. It is estimated that the total volume of all sediment and sedimentary rocks is over 6.5×10^{10} km^3 (1.55×10^8 mi^3), a volume that would cover the Earth to a depth of 1.3 km (0.82 mi) if spread evenly over its surface. Recent studies of the Moon and planets show that the Earth is very active tectonically compared with other bodies of similar size in the solar system. *See* BASIN; PLATE TECTONICS.

Climate. This plays a secondary but important role in controlling the rate of weathering and sediment production. The more humid the climate, the higher these rates are; evidence exists, however, that in some continental climates (for example, in the southwestern United States) the rate of sediment production reaches a maximum in semiarid regions (250–750 mm or 30–90 in. of rainfall per year). A combination of hot, humid climate and low relief permits extensive chemical weathering, so that a larger percentage of source rocks goes into solution, and the clastic sediment produced consists mainly of those minerals that are chemically inert (such as quartz) or that are produced by weathering itself (clays). Cold climates and high relief favor physical over chemical processes. Source rocks disintegrate to form sediment that includes little clay and many minerals that are chemically unstable in the Earth's surface environments. *See* CLIMATOLOGY.

Sea level. Tectonics and climate together control the relative level of the sea. In cold periods, water is stored as ice at the poles, which can produce a worldwide (eustatic) lowering of sea level by more than 100 m (330 ft). Changes in the rate of sea-floor spreading can produce large alterations in sea level by changing the average depth of the ocean basins. Widespread, but not worldwide, uplift or subsidence can also result in emergence or flooding of substantial areas of the continents. Changes in sea level, whether local or worldwide, strongly influence sedimentation in shallow seas and along coastlines; sea-level changes also affect sedimentation in rivers by changing the base level below which a stream cannot erode its bed.

Evolution of environments. One of the major conclusions from the study of ancient sediments has been that the general nature and rates of sedimentation have been essentially unchanged during the last billion years of geological history. However, this conclusion, uniformitarianism, must be qualified to take into account progressive changes in the Earth's environment through geological time, and the operation of rare but locally or even globally important catastrophic events. The most important progressive changes have been in tectonics and atmospheric chemistry early in Precambrian times, and in the nature of life on the Earth, particularly since the beginning of the Cambrian. *See* CAMBRIAN.

Rare events. Throughout geological time, events that are rare by human standards but common on a geological time scale, such as earthquakes, volcanic eruptions, and storms, produced widespread sediment deposits. There is increasing evidence for a few truly rare but significant events, such as the rapid drying up of large seas (parts of the Mediterranean) and collisions between the Earth and large meteoric or cometary bodies (bolides).

Sediment movement and deposition. Sediment is moved either by gravity acting on the sediment particles or by the motions of fluids (air, water, flowing ice), which are themselves produced by gravity. Deposition takes place when the rate of sediment movement decreases in the direction of sediment movement; deposition may be so abrupt that an entire moving mass of sediment and fluid comes to a halt (mass deposition, for example, by a debris flow), or so slow that the moving fluid (which may contain only a few parts per thousand of sediment) leaves only a few grains of sediment behind.

Fluids transport sediment mainly by traction (sliding, rolling, bouncing on the bed) and suspension caused by turbulence. Though the irregular, turbulent motion of fluids such as air and water are the most important way that sediment is held up within a mass of flowing fluid, other mechanisms such as buoyancy and impacts between the grains may also operate. *See* TURBULENT FLOW.

A fundamental dynamic property of sediment grains is their settling velocity. This property is the constant velocity of fall that a grain attains in a fluid at rest, when the force of gravity is balanced by the forces of buoyancy and fluid resistance. The settling velocity depends on the density and viscosity of the fluid, as well as on the size, shape, and density of the grains. As measured under standard conditions (in water at 20°C or 68°F), it can be used directly to characterize sediment or be converted into an equivalent diameter (the diameter of a quartz sphere with the same settling velocity as the measured value). Settling velocities in water vary from less than 0.1 mm (0.004 in.) per second for silt, to 5–10 cm (2–4 in.) per second for average sand, to meters per second for large boulders.

Transport and deposition by flowing water. The different ways in which sediment is transported by water flows can be illustrated by describing the changes that are observed in a laboratory channel (flume) as the discharge of water is gradually increased over a flat bed of sand. The flow strength just necessary to start movement of grains on the bed defines the competence of the flow for a particular type (size, shape, density) of sediment. The first grain motion is by sliding or rolling of the grains in contact with the bed, and this is soon followed by a leaping motion known as saltation. In saltation, grains rise up at a steep angle from the bed, are caught by the flow, and drop back to the bed; they follow a so-called ballistic trajectory. Sliding, rolling, and saltation generally are grouped together as traction. Sediment moved in this way is described as bed load. At low flow strengths, the grains do not rise far above the bed (only a few grain diameters in water) and they move much more slowly than the fluid. Rates of sediment movement (as measured by the sediment discharge, or mass rate of transport of sediment past a particular cross section of the channel) depend on a high power of flow strength. Bed load is supported by direct contact with the bed (and by buoyancy). Its motion responds to the fluctuating shear stress on the bed produced by turbulence in the flow rate,

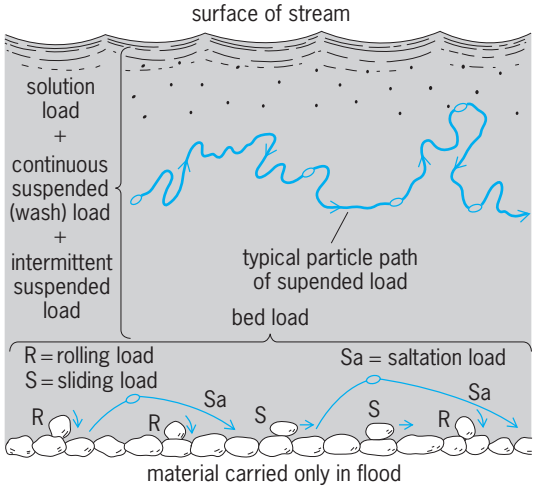


Fig. 2. Diagram of types of load carried by a stream. (*After R. M. Garrels, A Textbook of Geology, Harper and Row, 1951*)

but it is otherwise not much affected by turbulent fluid motions (**Fig. 2**).

As flow strength is increased, however, the upward velocity components of turbulent eddies become strong enough to overcome the settling velocity of the sediment, and the grains are taken into suspension. A part of the sediment load is still moved by traction, but a larger and larger part is moved by suspension (suspended load). Suspended sediment, at first concentrated near the bed, becomes more uniformly distributed in the flow as the flow strength increases. The total rate at which sediment can be moved (capacity) depends upon the availability of erodible sediment in the banks or bed and upon the sediment properties and flow strength. Sediment discharge continues to increase with flow strength, but not as rapidly as in the early stages of sediment transport. There is almost no limit to the amount of fine sediment that can be carried in suspension: concentrations as high as 30% by weight have been recorded. As concentration increases, so does the likelihood of collisions between the coarser grains. It has been suggested that such collisions become an important support mechanism (dispersive pressure) for sand and gravel at volume concentrations higher than about 10%, and such concentrations are certainly present close to the bed at high flow strengths.

As the rates of sediment transport increase, the nature of the bed also changes. Even if the bed is flat at first, it develops asymmetrical waves or bedforms migrating downstream (**Fig. 3**). Small current ripples (10–30 cm or 4–12 in. long) form soon after sediment movement begins, but at higher flow strengths they are replaced by large dunes on the bed; in deep flows, these may reach heights of several meters and lengths of tens of meters. At very high flow strengths, however, the dunes are washed out and the bed becomes almost flat again. If the flow becomes supercritical, nearly symmetrical bedforms (antidunes) migrating upstream may be

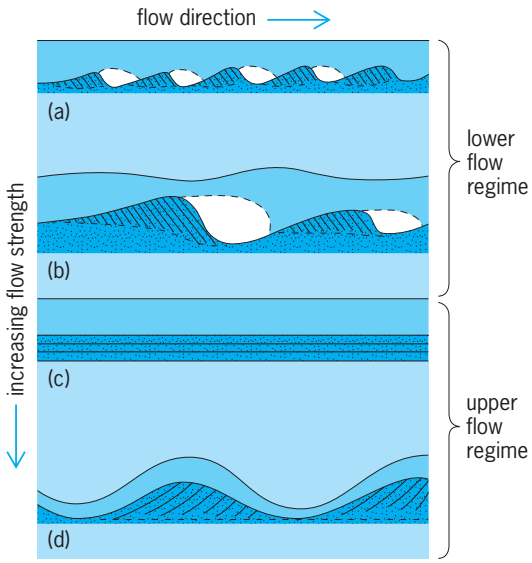


Fig. 3. Bedforms *a–d* produced by flow over a sand bed. (*After H. Blatt, G. V. Middleton, and R. Murray, Origin of Sedimentary Rocks, 2d ed., Prentice-Hall, 1980*)

formed. *See* DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; FROUDE NUMBER; STREAM TRANSPORT AND DEPOSITION.

Transport by waves. Water waves produce an oscillating motion of water close to the bed. Therefore, movement of sediment by waves differs in detail from that produced by unidirectional flows. Net sediment movement generally still results, however, because of the asymmetrical nature of the oscillating flow close to the bed. Waves produce stages of sediment transport similar to those resulting from unidirectional flows. Suspension replaces traction at high levels of wave activity. Bedforms show a transition from small to larger ripples, and a plane bed reappears at the highest wave intensities. Breaking of waves greatly enhances the ability of waves to carry sediment into suspension. After breaking, waves form a bore (swash) that washes sediment up the beach, and down again in the backwash. The approach of waves is generally oblique to the shore and produces a movement of sand along the beach that is known as longshore drift. *See* NEARSHORE PROCESSES.

Transport by wind. Sediment transport by wind is similar to transport by water, but different forms of transport predominate because of the low density of air compared with water, and the great thickness and high speed attained by moving air masses. Whereas sand is easily transported in suspension by water flowing at speeds of the order of 1 m (3.3 ft) per second, it can be suspended only rarely by winds. Most sand is moved by wind in saltation, which is a far more effective mechanism of sediment transport in air than in water. Saltating sand grains rise to heights of several centimeters and return to the bed with enough momentum to drive forward other grains, a process known as surface creep. Ripples produced by saltating grains are smaller and more regular than those produced by flow of water, but dunes develop to much larger sizes and show a wide variety of forms because of the thickness of the wind boundary layer and the variability of wind directions. *See* AIR MASS.

Only clay and silt are easily carried into suspension by the wind, but once suspended, these grains may be transported for hundreds of kilometers. Fine sediment (ash) is produced by explosive volcanic eruptions and carried high into the upper atmosphere, where it is transported great distances by the wind. Deposits of silt (loess) several meters thick were formed at the end of the last ice age by winds that blew sediment derived from rivers that drained the ice sheets. *See* DUNE; DUST STORM; LOESS; WIND.

Transport by ice. Glaciers and ice sheets incorporate sediment that falls onto the ice surface or is eroded by grinding or plucking from material beneath the ice. Such material is transported en masse to the point where the glacier melts. Sediment deposited by melting directly out of ice is almost completely unsorted and is known as till. Some ice sheets melt directly into lakes or the sea, producing debris that is modified as it settles to the bottom. Much of the sediment that is derived from glaciers is reworked by mass flow (flow till) or by meltwater (outwash). Some of the sediment eroded by glaciers consists of finely ground material known as rock flour. The abundance of this material in lakes fed by glaciers is at least partly responsible for their intense green color. *See* GLACIOLOGY; TILL.

Sediment gravity flows. These are flows whose motion results from gravity acting on the sediment grains rather than on the fluid in the flow (indeed, no fluid is necessary for certain types of gravity flow, and so these flows can operate in fluid-free environments, such as the Moon). Four classes of sediment gravity flow have been recognized, differentiated by the mechanisms that support sediment above the bed. If the mechanism is grain impacts, the flow is a grain flow; if it is turbulence, the flow is a turbidity current; if it is the upward escape of interstitial fluid, the flow is a liquefied flow; and if it is the strength of the matrix between the grains (for example, a mud matrix), the flow is a plastic debris flow. Such flows are generally short-lived, but they may be very large. Some recorded turbidity currents have transported as much as 100 km^3 (24 mi^3) of sediment and have flowed for several hundred kilometers over the ocean floor. *See* TURBIDITY CURRENT.

Modification by transport. Sediment is modified during transport by chemical action (continuation of the weathering process begun in the source) and by physical breakage, abrasion, and sorting. Physical impact is most effective on the larger grains, and it has little effect on grains small enough to be transported in suspension. Boulders and pebbles are rapidly rounded by abrasion of the corners and edges of the grain, and prolonged transport in rivers may lead to a nearly spherical shape. The more friable minerals found in sand (for example, feldspar) may be broken, but the main mineral, quartz, is very resistant to abrasion. Nevertheless, studies using the scanning electron microscope have shown evidence

of rounding and submicroscopic impact marks, and reduction in size by flaking of small chips from the surface. Transport by wind is far more effective than transport by water in rounding quartz grains, and reworking by waves on beaches is more effective than transport down rivers—even when the transport distance is hundreds of kilometers down major rivers.

Sediment weathered from an igneous or metamorphic source rock, or produced from it by mechanical breakage, tends to have a wide range of grain sizes that follow a statistical distribution known as Rosin's distribution. Though a similar distribution may be detected in some cases after several kilometers of river transport, generally it is rapidly modified by hydraulic sorting, which tends to segregate the variously sized fractions by their different mechanisms of transport. In a river, for example, the coarsest grains are confined to the deepest part of the channel; the sand is confined to bars within the channel or to the natural levees; and either the mud is transported in suspension rapidly through the system to the sea, or it is deposited during floods in overbank deposits on the floodplain. Because each of these deposits consists mainly on only one fraction of the original source material, they are better sorted and the size distribution is changed. Most naturally occurring size distributions appear to be composed of mixtures of more than one (lognormal) distribution, resulting from different sources or mechanisms of deposition. *See* FLUVIAL SEDIMENTS.

Flocculation. Gravel and sand grains are generally chemically inert and noncohesive. This is not true of the finest, clay-sized particles. Colloid-sized particles (finer than 0.2 micrometer) generally have electrically charged surfaces, even if composed of inert minerals. Many clay minerals, such as montmorillonite, readily react with dissolved ions. The surface charge depends in part on the nature and concentration of ions in the surrounding solution. Though the grains of clay minerals are very small (less than 1 μm), electrochemical effects may cause them to aggregate into large flocs whose settling velocity (though still relatively low, because of the low bulk density of the floc) may be many times higher than that of the individual grains. Flocculation is very common as clay is delivered by rivers into the sea; thus it is an important process in estuaries and deltas. *See* CLAY MINERALS; DELTA; ESTUARINE OCEANOGRAPHY; GRAVEL; SAND.

Chemical sedimentation. Chemical weathering dissolves rock materials and delivers ions in solution to lakes and the ocean. The concentrations of ions in river and ocean water are quite different, showing that some ions must be removed by sedimentation. Comparison of the modern rate of delivery of ions to the ocean, with their concentration in the oceans, shows that some are removed very rapidly (residence times of only a few thousand years) whereas others, such as chlorine and sodium, are removed very slowly (residence times of hundreds of millions of years). The main mechanisms by which ions are removed from solution are biochemical precipitation, evaporation, adsorption or exchange with clays and other rock materials (including glass in fine volcanic ash), and reaction with hot rock in hydrothermal systems associated with mid-oceanic ridges. The most important chemical sediments are calcareous (limestones and dolomites), followed by siliceous chert deposits and evaporites, but there are a very large variety of chemical sediments whose origins are not yet fully understood. Chemical dissolution and precipitation are not restricted to weathering and primary sedimentation from lake water or seawater, but are equally important in diagenesis, that is, in transforming sediment after deposition into sedimentary rock. *See* MARINE SEDIMENTS; MID-OCEANIC RIDGE.

Carbonate sediments. Organisms began to secrete skeletons composed of calcium carbonate during the Cambrian Period, and since then most carbonate sediment has been precipitated biochemically. Though the warm surface waters of the oceans are known to be supersaturated with respect to calcite (the thermodynamically stable form of calcium carbonate in most natural waters), direct chemical precipitation of calcite is known in only a few environments (for example, limestone caves). The common carbonate minerals in modern sediments are biochemically precipitated aragonite and calcite. Aragonite is metastable under surface conditions and is generally dissolved away within a few thousand years (rarely millions) to be replaced by diagenetic calcite or dolomite. *See* LIMESTONE.

The best-known biochemical accumulations of carbonate at present are the tropical coral-algal reefs, but a wide variety of organisms precipitate carbonate from seawater now or they have done so in the past. Besides reef organisms, pelecypods, algae, and foraminifera are important in cooler marine waters; and in the surface layers of the oceans, planktonic foraminifera, pteropods, and algae (coccolithophorids) are important today, and were so at other periods of the Cenozoic and Mesozoic. Cretaceous chalks, for example, are composed largely of microscopic coccoliths and foraminifera, not very different from modern deep-sea calcareous oozes (though most ancient chalks were deposited in water only a few hundred meters deep). In other geological periods, reefs were built by different types of corals, by specialized pelecypods (rudists), by algae, and by now-extinct groups such as stromatoporoids. *See* CENOZOIC; CHALK; CRETACEOUS; MESOZOIC; REEF.

The most important differences between ancient carbonate rocks and modern carbonate sediments, however, are due to diagenesis. Modern carbonate sediments are composed mainly of aragonite and a form of calcite containing several percent of magnesium. Ancient carbonates consist largely of calcite low in magnesium and of dolomite, a calcium-magnesium carbonate only rarely present in modern sediments. The mineralogical changes necessary to produce typical ancient carbonates take place rapidly, particularly under the action of meteoric water. Even the formation of dolomite, which generally requires the wholesale dissolution of aragonite and calcite and the exchange of large quantities of magnesium from water passing through the porous

rock, has been a common and frequently rapid process throughout geological history. *See* ARAGONITE; AUTHIGENIC MINERALS; CALCITE; CARBONATE MINERALS; DIAGENESIS; DOLOMITE; MARL.

Siliceous sediment. Sediments composed largely of fine-grained quartz (chert) are common in the geological record, extending back to the earliest Precambrian. Since the beginning of the Cambrian, most cherts have been formed by biochemical precipitation by organisms that secrete an opaline skeleton, such as radiolarians, diatoms, and some sponges, followed by diagenetic dissolution and reprecipitation as microcrystalline quartz and chalcedony. *See* CHALCEDONY; CHERT.

Evaporites. Evaporation of seawater in the laboratory gives rise to a sequence of different precipitates as evaporation proceeds. Small amounts of calcareous sediments are first precipitated, followed by gypsum, anhydrite, halite, and finally by complex salts containing sodium, potassium, magnesium, sulfate ions, and chloride ions. A close match with this experimentally determined sequence is known to exist in several ancient deposits, though the relative abundances of different minerals vary considerably. This indicates that these deposits were formed by evaporation of seawater, and that the composition of the ancient seawater was not very different from that of modern seawater. Evaporites can presently be observed forming in several shallow-water environments (lakes and marine lagoons), though none is comparable in magnitude to the evaporite basins inferred for some periods in the past. Formation of large evaporite deposits was probably a rare event, but once conditions were appropriate, large volumes could be formed in only a few tens of thousands of years. *See* SALINE EVAPORITES; SEAWATER.

Other chemical sediments. Other chemical sediments are rich in iron, phosphorus, manganese, and carbon. The carbonaceous sediments include marine muds rich in organic matter, which are important as the source rocks of petroleum, and nonmarine peats, some of which give rise to commercial coals. Many sedimentary rocks of unusual character are economically important; they serve as ores of iron, copper, manganese, phosphate, gold, diamonds, and so forth. Equally important economically are more common sedimentary rocks, which serve as sources of building stone, crushed rock for construction, limestone for smelting, pure quartz sands, and so forth. *See* COAL; ORE AND MINERAL DEPOSITS; PETROLEUM.

Biological effects. Many so-called chemical sediments are actually produced by biochemical action. Much is then reworked by waves and currents, so that the chemical sediment shows clastic textures and consists of grains rounded and sorted by transport. Depositional and diagenetic processes, however, are often strongly affected by organic action, no matter what the origin of the sediment. Plants in both terrestrial and marine environments tend to trap sediment, enhancing deposition and slowing erosion. Organisms burrow through the sediment, producing distinctive structures (bioturbation) and modifying or destroying original clastic structures and textures. Bacteria promote decay of organic particles, changing the chemical environment within the sediment, and producing gases such as carbon dioxide and methane. At deeper levels of burial within a sedimentary basin, the residual organic matter in the sediments (kerogen or sapropel) is broken down by heat to produce liquid and gaseous hydrocarbons and carbon dioxide, which may migrate many tens of kilometers upward along permeable strata within the basin. *See* KEROGEN; SAPROPEL.

Sedimentary environments and facies. Sedimentary rocks preserve the main direct evidence about the nature of the surface environments of the ancient Earth and the way they have changed through geological time. Thus, besides trying to understand the basic principles of sedimentation, sedimentologists have studied modern and ancient sediments as records of ancient environments. For this purpose, fossils and primary sedimentary structures are the best guide. These structures are those formed at the time of deposition, as opposed to those formed after deposition by diagenesis, or by deformation. In describing sequences of sedimentary rocks in the field (stratigraphic sections), sedimentologists recognize compositional, structural, and organic aspects of rocks that can be used to distinguish one unit of rocks from another. Such units are known as sedimentary facies, and they can generally be interpreted as having formed in different environments of deposition. Though there are a large number of different sedimentary environments, they can be classified in a number of general classes, and their characteristic facies are known from studies of modern environments. Furthermore, the transition from one environment to another is not a random process, because only those environments that originally existed side by side can succeed each other without a break (Walther's law). Thus, it is possible to produce plausible reconstructions of environments that existed many millions of years ago. Like all historical reconstructions based on fragmentary evidence, these depend upon the imaginative skill of the investigator, as well as upon reliable deductions from data. *See* FACIES (GEOLOGY); TRACE FOSSILS.

Basin analysis. This describes studies of the larger-scale aspects of sediment accumulation, such as the different types of sedimentary basins, and how they form, and the types of source areas and how they are linked to particular types of basins. Clastic sediments derived from preexisting rocks (terrigenous sediments) have textures and mineralogy that preserve clues about the nature of the source from which they were derived (their provenance). Sedimentary structures and textures can be used to indicate the direction of sediment transport at the place of deposition (paleocurrents). By carefully dating the deposits, or at least determining which deposits are the same age (stratigraphic correlation), by studying their provenance and paleocurrents, and by reconstructing their depositional environments, it is possible to reconstruct how a sedimentary basin developed and where the sediment came from. In

addition, the long-term rates of deposition can be determined, and by studying mineralogical changes produced by diagenesis it is possible to reconstruct the sequence of pressures and temperatures that have affected the sediments in the basin. Tectonic theories suggest models that simulate the patterns of deposition and diagenesis within sedimentary basins. *See* SEDIMENTARY ROCKS. G. V. Middleton

Bibliography. P. A. Allen, *Earth Surface Processes*, 1997; H. Blatt, G. V. Middleton, and R. Murray, *Origin of Sedimentary Rocks*, 2d ed., 1980; M. R. Leeder, *Sedimentology*: *Process and Product*, 1982; M. R. Leeder, *Sedimentology and Sedimentary Basins*: *From Turbulence to Tectonics*, 1999; R. K. Matthews, *Dynamic Stratigraphy*: *An Introduction to Sedimentation and Stratigraphy*, 2d ed., 1984; A. D. Miall, *Principles of Sedimentary Basin Analysis*, 2d ed., 1990.

Seebeck effect

The generation of a temperature-dependent electromotive force (emf) at the junction of two dissimilar metals. This phenomenon provides the physical basis for the thermocouple. In 1821, T. J. Seebeck discovered that near a closed circuit composed of two linear conductors of two different metals a magnetic needle would be deflected if, and only if, the two junctions were at different temperatures, and that if the temperatures of the two junctions were reversed the direction of deflection would also be reversed. He investigated 35 different metals and arranged them in a series such that at a hot junction, current flows from a metal earlier in the series to a later one. *See* ELECTROMOTIVE FORCE (EMF).

A thermocouple consists of a pair of wires of dissimilar metals, joined at the ends. One junction is kept at an accurately known cold temperature, usually that of melting ice, and the other is used for the measurement of an unknown temperature, by measuring the emf generated as a result of the Seebeck effect. *See* THERMOCOUPLE; THERMOELECTRICITY. A. Earle Bailey

Seed

A fertilized ovule containing an embryo which forms a new plant upon germination. Seed-bearing characterizes the higher plants—the gymnosperms (conifers and allies) and the angiosperms (flowering plants). Gymnosperm (naked) seeds arise on the surface of a structure, as on a seed scale of a pine cone. Angiosperm (covered) seeds develop within a fruit, as the peas in a pod. *See* FLOWER; FRUIT.

Structure. One or two tissue envelopes, or integuments, form the seed coat which encloses the seed except for a tiny pore, the micropyle (**Fig. 1**). The micropyle is near the funiculus (seed stalk) in angiosperm seeds. The hilum is the scar left when the seed is detached from the funiculus. Some seeds have a raphe, a ridge near the hilum opposite the micropyle, and a bulbous strophiole. Others such as nutmeg possess arils, outgrowths of the funiculus, or a fleshy caruncle developed from the seed coat near the hilum, as in the castor bean. The fleshy, edible aril of the Philippine kamanchile completely encloses the seed to form a fruitlike structure. The embryo consists of an axis and attached cotyledons (seed leaves). The part of the axis above the cotyledons is the epicotyl (plumule); that below, the hypocotyl, the lower end of which bears a more or less developed primordium of the root (radicle). The epicotyl, essentially a terminal bud, possesses an apical meristem (growing point) and, sometimes, leaf primordia. The seedling stem develops from the epicotyl. An apical meristem of the radicle produces the primary root of the seedling, and transition between root and stem occurs in the hypocotyl. *See* APICAL MERISTEM; ROOT (BOTANY); STEM.

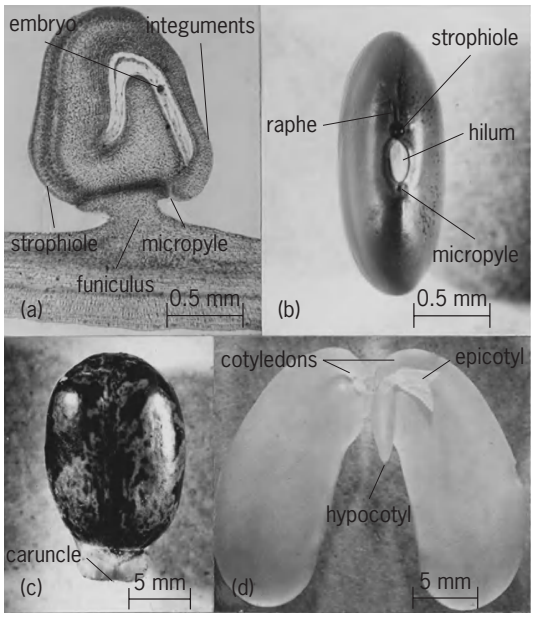


Fig. 1. Seed structures. (*a*) Median longitudinal section of pea ovule shortly after fertilization, showing attachment to pod tissues. (*b*) Mature kidney bean. (*c*) Mature castor bean. (*d*) Opened embryo of mature kidney bean.

Two to many cotyledons occur in different gymnosperms. The angiosperms are divided into two major groups according to number of cotyledons: the monocotyledons, including orchids, lilies, grasses, and sedges; and the dicotyledons, such as beans, roses, and sunflowers. Mature gymnosperm seeds contain an endosperm (albumen or nutritive tissue) which surrounds the embryo. In some mature dicotyledon seeds the endosperm persists, the cotyledons are flat and leaflike, and the epicotyl is simply an apical meristem (**Fig. 2**). In other seeds, such as the bean, the growing embryo absorbs the endosperm, and food reserve for germination is stored in fleshy cotyledons. The endosperm persists in common monocotyledons, for example, corn and wheat; and the cotyledon, known as the scutellum, functions as an absorbing organ during germination (**Fig. 3**). Grain embryos also possess a coleoptile and a coleorhiza sheathing the epicotyl and the radicle,

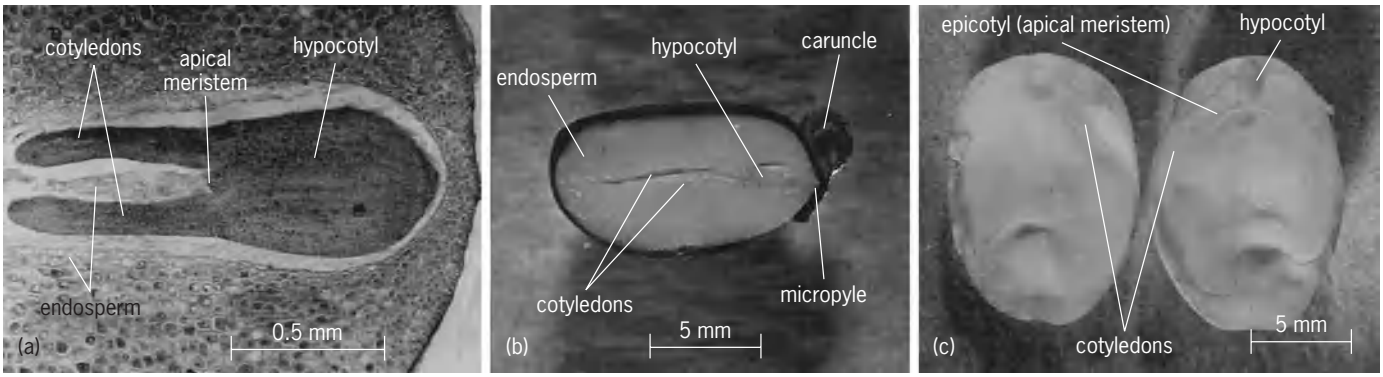


Fig. 2. Dicotyledon seeds. (*a*) Median longitudinal section of embryo of silk tassel bush (*Garrya elliptica*) embedded in endosperm which has been removed from mature seed. (*b*) Castor bean cut longitudinally. (*c*) Castor bean removed from seed coat and split longitudinally between cotyledons.

respectively. The apical meristems of lateral seed roots also may be differentiated in the embryonic axis near the scutellum of some grains.

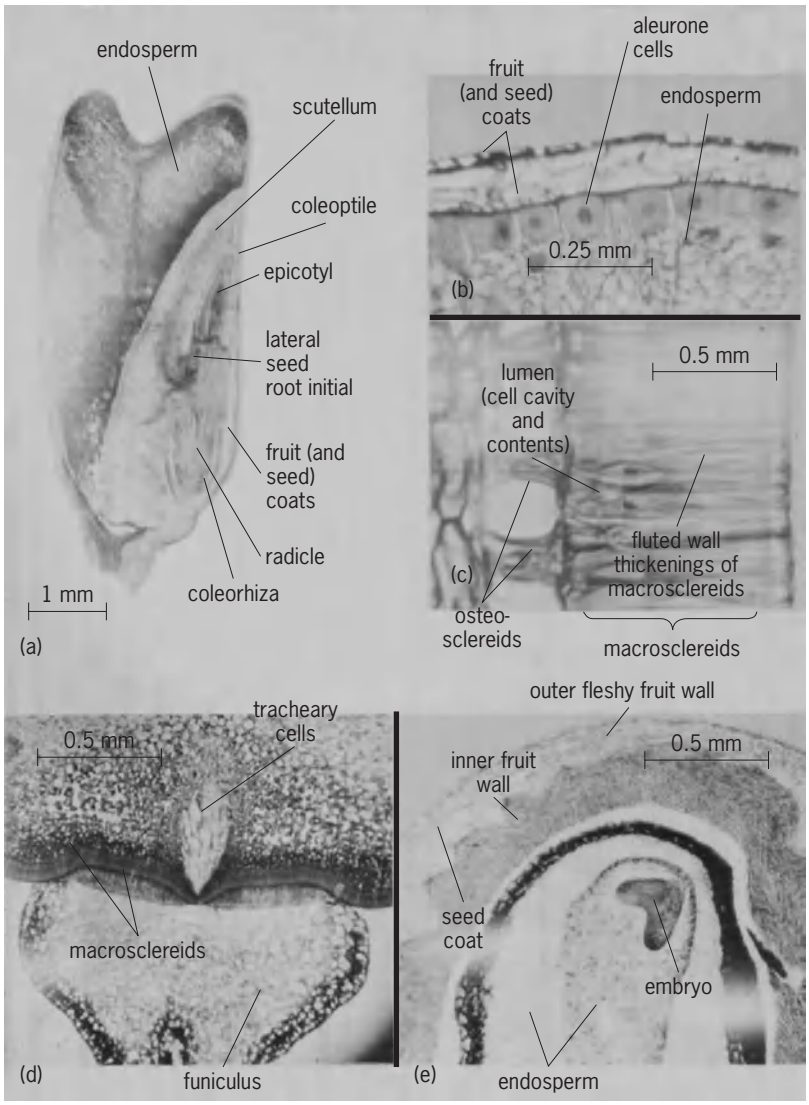


Fig. 3. Monocot seeds. (*a*) Section of corn kernel (*from M. J. Wolf et al., Cereal Chem., 29:321–382, 1952*). (*b*) Section of outer cell layers of wheat grain. (*c*) Section of outer cells of mature pea seed coat. (*d*) Section of hilum area of nearly mature pea. (*e*) Section through portion of young raspberry drupelet (segment of berry).

Monocotyledon and dicotyledon seeds also differ in seed coat structure. In grains, or caryopsis fruits, the mature fruit wall and seed coat may be fused and the outermost endosperm cells form an aleurone layer, rich in proteins. Flour brans consist of fragments of aleurone and fruit-seed coats. Most of the thiamine and riboflavin of grains occur in these tissues and near the scutellum.

Cellular structure of the seed coat varies in the dicotyledons. An outer, water-impervious layer of wax or cuticle is usually present. The outermost cells in bean and pea seed coats, known as macrosclereids (Malpighian cells), have fluted, cellulosic wall thickenings; underlying them is a layer of osteosclereids, short bone-shaped cells. Two layers of macrosclereids occur at the hilum in peas and beans. Underlying them is a group of tracheary cells with reticulate (netted) wall thickenings, and a surrounding spongy tissue of branched cells and large air spaces. This structure appears to be well suited for water absorption during germination.

Many so-called seeds consist of hardened parts of the fruit enclosing the true seed which has a thin, papery seed coat. Among these are the achenes, as in the sunflower, dandelion, and strawberry, and the pits of stone fruits such as the cherry, peach, and raspberry. Many common nuts also have this structure.

Dispersal. Mechanisms for seed dispersal include parts of both fruit and seed. Some dry fruits have membranous air sacs that aid in seed dispersal by water. In cockleburrs and beggarticks the seeds are enclosed in spiny, barbed fruits that are readily carried in the fur of animals. Seed dispersal by wind is aided by hair tufts of dandelion achenes, wings of elm and maple fruits and of conifer seeds, and by seed-coat hairs of willow, cottonwood, and milkweed seeds. Cotton fibers consist of greatly elongated epidermal hairs of the seed coat as do also the kapok fibers of the silk-cotton tree (*Ceiba pentandra*). *See* POPULATION DISPERSAL.

Economic importance. Propagation of plants by seed and technological use of seed and seed products are among the most important activities of modern society. Specializations of seed structure and composition provide rich sources for industrial

exploitation apart from direct use as food. Common products include starches and glutens from grains, hemicelluloses from guar and locust beans, and proteins and oils from soybeans and cotton seed. Drugs, enzymes, vitamins, spices, and condiments are obtained from embryos, endosperms, and entire seeds, often including the fruit coat. Most of the oils of palm, olive, and pine seeds are in the endosperm. Safflower seed oil is obtained mainly from the embryo, whereas both the seed coat and embryo of cotton seed are rich in oils. *See* FOOD; PLANT ANATOMY; REPRODUCTION (PLANT).

Roger M. Reeve

Physiology. Physical and biochemical processes of seed growth and germination are controlled by genetic and environmental factors. Conditions of light, temperature, moisture, and oxygen affect the timing and ability of a seed to mature and germinate. Seed development (embryogenesis) is concerned with the synthesis and storage of carbohydrate, protein, and oil to supply nutrients to the germinating seedling prior to soil emergence. Seed development occurs in several stages: rapid cell division, seed fill, and desiccation (**Fig. 4**). The timing of each stage is species-specific and environmentally influenced. In mature monocot seeds (for example, corn), starch in the endosperm is the main storage product, while small quantities of oil and protein are stored primarily in the embryo and aleurone (testa/pericarp), respectively. Major storage reserves in mature dicot seed cotyledons may be starch as in pea, or protein and oil as in soybean. In the case of soybean, seeds develop over a period of about 60 days. Protein and oil accumulation is confined to a period of carbon and nitrogen import from the vegetative plant, resulting in rapid dry-weight increases (seed fill). During late seed fill and desiccation, there is an accumulation of small amounts of the raffinose saccharides, raffinose and stachyose. Synthesis of raffinose and stachyose in soybean is correlated with increases in galactinol synthase activity, the enzyme that catalyzes the committing step to this biosynthetic pathway (**Fig. 5**). Total starch and soluble sugars account for only approximately 12% dry weight of a dry soybean seed. With desiccation, a viable seed will germinate if provided with suitable environmental conditions for germination and for overcoming any dormancy constraints.

Dormancy. Seed dormancy is the inability of a living seed to germinate under favorable conditions of temperature, moisture, and oxygen. Dormancy does not occur in all seeds, but typically occurs in plant species from temperate and colder habitats such as maple, pear, hazel, and pine. This process allows for a delay in seed germination until environmental conditions are adequate for seedling survival. At least three types of seed dormancy are recognized: primary, secondary (induced), and enforced. Primary dormancy occurs during seed maturation, and the seed does not germinate readily upon being shed. Internal factors such as chemical inhibitors, immaturity of the embryo, and impermeability of hard seed coat tissues to water and oxygen may be important in causing primary dormancy. Secondary and enforced dormancy occur after the seed is shed and may be caused by adverse environmental factors such as high or low temperature, absence of oxygen or light, low soil moisture, and presence of chemical inhibitors. Seeds with secondary dormancy will not germinate spontaneously when environmental conditions improve, and need additional environmental stimuli. Seeds with enforced dormancy germinate readily upon removal of the environmental limitation.

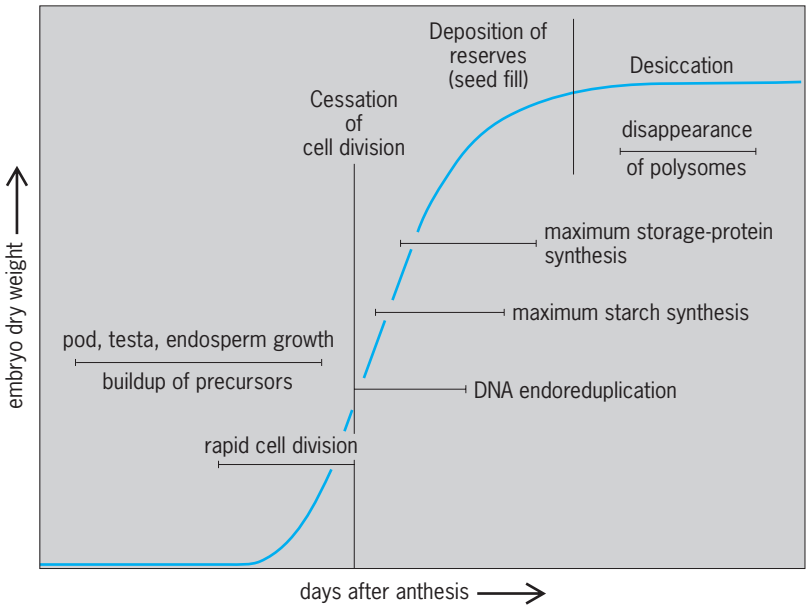


Fig. 4. Process of legume seed development (embryogenesis). (*After L. S. Dure, Seed formation, Annu. Rev. Plant Physiol., 26:259–278, 1975*)

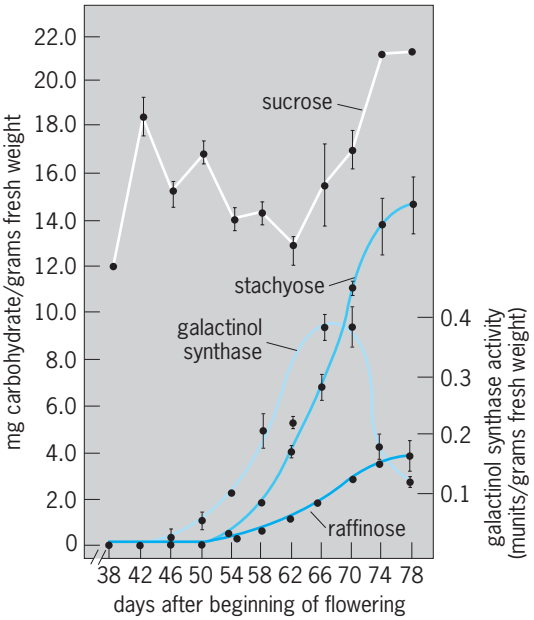


Fig. 5. Changes in levels of the sugars sucrose, raffinose, and stachyose, and concomitant increases in galactinol synthase activity in maturing soybean seeds. The term munits/grams fresh weight represents the amount of enzyme (galactinol synthase) necessary to form 1 nanomole of galactinol per minute at 30°C (86°F) per gram fresh weight. (*After L. W. Handley et al., Relationships between galactinol synthase activity and sugar composition of leaves and seeds of several crop species, J. Amer. Soc. Hort. Sci., 108:600–605, 1983*)

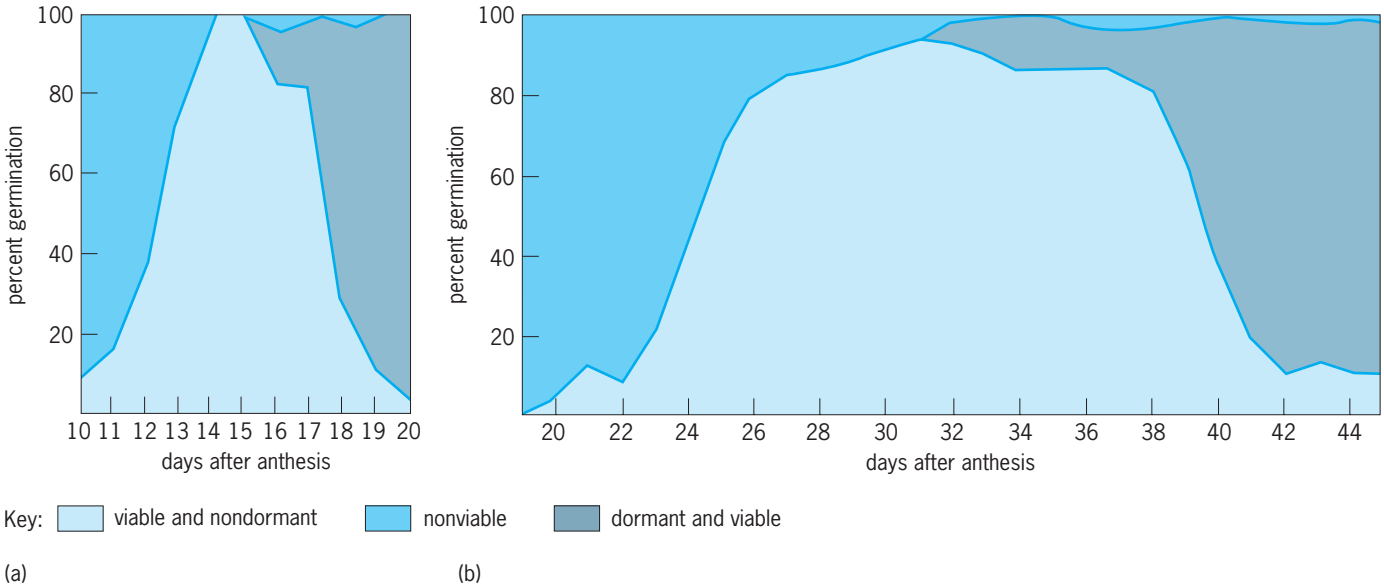


Fig. 6. Seed dormancy and viability during development under (*a*) field and (*b*) greenhouse conditions. Black medick (*Medicago lupulina*) seeds were detached from the plant and tested for germination. Three phases of germinability occurred depending on seed age; environmental conditions affect the length of each phase. (*After S. S. Sidhu and P. B. Cavers, Maturity-dormancy relationships in attached seeds of Medicago lupulina L. (black medick), Bot. Gaz., 138:174–182, 1977*)

Dormant seeds have all normal metabolic functions associated with nondormant seeds, except that the respiratory pentose phosphate pathway in dormant seeds may be inadequate to support germination. *See* DORMANCY; PLANT RESPIRATION.

Onset of dormancy. There is no one biochemical factor that has been determined to actually cause dormancy; however, factors affecting this process are genetic, environmental, and chemical (hormonal). Onset of seed desiccation during maturation appears to impart germability to seeds of barley and wheat, while low moisture content typical of a mature seed (about 15%) correlates with dormancy. Moisture loss specifically from seed coat tissues also is a factor in dormancy in common purslane (*Portulaca oleraceae*) and prickly sida (*Sida spinosa*). In seeds of black medick (*Medicago lupulina*), three phases of germinability/dormancy occur during development, the timing of which varies with the environment (**Fig. 6**). During early seed development, these seeds cannot germinate if detached from the plant. The seeds are subsequently viable and nondormant during seed fill, and can germinate on (precocious germination) or off the plant. These seeds finally become dormant during desiccation because of the development of an impermeable seed coat.

Both temperature and light may affect the onset of dormancy. Dormancy may be related to environmental temperatures during seed development in roses. The lower the daily average temperature the plants were grown at, the more pronounced the seed dormancy. In barley and wheat, embryos grown at 68–70°F (20–21°C) have greater incidence of dormancy than those grown at 82°F (28°C). In some cases such as seeds of *Syringa*, higher temperatures than normally encountered (64–75°F; 18–24°C) induce dormancy. Thermoperiod (different day/night temperatures) also may influence seed dormancy. In seeds of scarlet pimpernel (*Anagallis arvensis*), low dormancy occurs if seeds develop in a day/night temperature regime of 86°/77°F (30°/25°C), and extreme dormancy occurs in the seeds at 68°/59°F (20°/15°C) conditions.

Light-induced dormancy may be a photoperiodic (day/light time) or a phytochrome response. Long days appear to have greater effect than short days on inducing dormancy in developing seeds in species of *Chenopodium*, *Lactuca*, and *Portulaca*. The extent of this dormancy is related to the number of long days the developing seed receives. This time may or may not be correlated with seed coat thickness, depending on the species. Dormancy in a few species such as *Carrichtera anna*, however, is induced by short days. Nonphotoperiodic light effects on dormancy occur in mouse-ear cress (*Arabidopsis thaliana*) and cucumber, and they are controlled by the pigment phytochrome. Seeds from plants grown in far-red-enriched light (sunlight, incandescent) become dormant upon maturation, while those grown in red-enriched light (white fluorescent) are nondormant. *See* PHOTOPERIODISM; PHYTOCHROME.

Regulation of dormancy may be partly controlled by hormones. Maturing seeds, whether they undergo dormancy or not, accumulate abscisic acid. Induction of dormancy was correlated with abscisic acid in the embryo of mouse-ear cress. Abscisic acid also has been shown to inhibit translation by messenger ribonucleic acid (mRNA), an essential step in protein synthesis during early germination. *See* ABSCISIC ACID; RIBONUCLEIC ACID (RNA).

Release from dormancy. Dormancy is terminated in a large number of species when an imbibed seed is illuminated with white light. Biochemical control of this process is related to the functioning of a

single pigment, phytochrome, frequently located in the seed coat or embryonic axis. Phytochrome imparts to the seed the ability to interpret light quality, such as that under an existing vegetative canopy, and to distinguish light from dark with respect to its position in the soil. Phytochrome also is affected by temperature and is involved in the seasonal control of the ending of dormancy.

Light of three spectral regions interacts with phytochrome, blue (450 nanometers), red (650 nm), and far-red (750 nm); red light promotes, while far-red and blue light inhibit, seed germination. Phytochrome exists in two forms, P_r which absorbs red light and P_{fr} which absorbs far-red light. In a dormant seed in the dark, the P_r phytochrome form predominates. When light in the red range hits the seed, the phytochrome pigment converts from the P_r to P_{fr} form, and dormancy is terminated. The ratio of the two forms is critical to breaking dormancy. Germination is obtained when the P_{fr}/P_{total} ratio reaches a certain threshold value that is species-specific. Exposure to red light necessary to break dormancy may be a short burst, intermittent, or continual up to several days. The promotive effect of red light can be reversed by exposure to far-red light or long-term exposure to temperatures above 77°F (25°C). Blue light elicits a high-irradiance response in the seed. Short exposure to blue light reverses the effect of the red light and inhibits germination. The high-irradiance response, however, affects primarily elongation of the radicle after germination.

Temperature also can be a factor in ending seed dormancy, typically in imbibed seeds. Dormancy is broken via diurnal (daily) fluctuations or one-time shifts in temperature. In tobacco, diurnal alternations in temperature between 68 and 86°F (20 and 30°C) induce seed germination better than at either one temperature. Seed chilling or stratification at 34–59°F (1–15°C) for a few hours to days is a common requirement of temperate woody plants such as maple, apple, and hawthorn for germination. In many seeds with this requirement, stratification at 41°F (5°C) for several days prior to returning the seeds to ambient temperature is optimal for breaking dormancy in fully imbibed seeds.

Dormant seeds also may gradually overcome blocks to germination without imbibition (after-ripening). After-ripening is a time of physiological and anatomical maturation after the seed is shed from the plant that frequently occurs in cereals such as barley, rice, and wheat, and in grasses such as wild oat. After-ripening is stimulated by oxygen and high temperatures (86–104°F; 30–40°C).

Hormones that promote germination of dormant seeds include gibberellins, cytokinins, ethylene, and auxins. Exogenously applied gibberellins promote germination in dormant seeds, but internal concentrations of this hormone have not been shown to increase during dormancy. Cytokinin levels also have not been shown to change in dormant seeds, but cytokinins function in conjunction with gibberellins to induce germination when exogenously applied. Ethylene has stimulated germination of some species, and it has been linked to phytochrome and the cellular membrane state. Auxin does not promote germination when exogenously applied, but it does increase endogenously and is involved in loosening and expansion of cell walls, which is necessary for germination. Additional chemical compounds that stimulate seed germination when applied exogenously include plant or fungal chemical products (fusicoccin), respiratory inhibitors (cyanide, azide, hydroxylamine, hydrogen sulfide, carbon monoxide), oxidants, carbon dioxide nitrates or nitrites, sulfhydryl compounds, and certain organic compounds (ethanol, methanol). *See* AUXIN; CYTOKININS; ETHYLENE; GIBBERELLIN.

Germination. Germination is the process whereby a viable seed takes up water and the radicle (primary root) or hypocotyl emerges from the seed under species-specific conditions of moisture, oxygen, and temperature. Dormant seeds must undergo additional environmental stimuli to germinate. The germinating seed undergoes cell expansion, as well as increases in respiration, protein synthesis, and other metabolic activities prior to emergence of the growing seedling (**Fig. 7**).

As water, dissolved oxygen, and solutes from the soil enter the seed, there is a release of gas, unrelated to respiration, that was trapped in the seed following desiccation. Imbibition is accompanied by rapid leakage of substances of cytoplasmic origin such as sugars, organic acids, proteins, gibberellins, phenolics, phosphate, and potassium ions. Radicle expansion begins soon after the start of imbibition, and initial cell wall elongation proceeds slowly, lasting up to 46 h. This period is followed by rapid cell wall elongation, mobilization of metabolic reserves, and translocation of these materials to the growing radicle. Cell division may or may not be necessary for radicle protrusion. Cell expansion typically precedes cell division in seeds of corn, barley, broad bean, and pea.

Respiration of germinating seeds. Respiration in dry seeds is nearly nondetectable. In imbibed seeds, the three major respiratory pathways (glycolysis, pentose phosphate pathway, and Krebs cycle) are very active; they provide energy [adenosine triphosphate (ATP)], reducing power [β-nicotinamide adenine dinucleotide (NADH) and β-nicotinamide

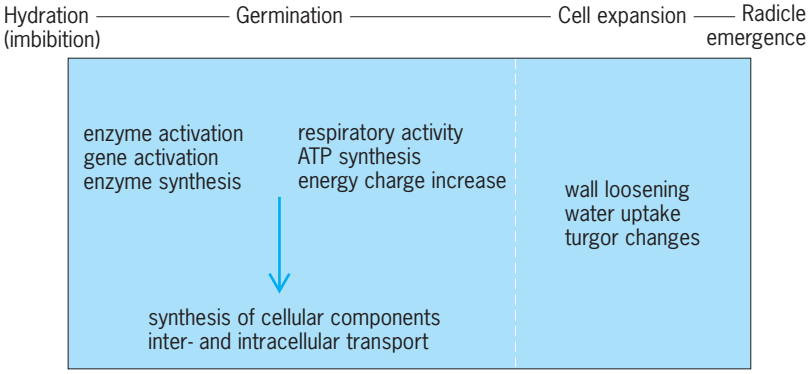


Fig. 7. Biochemical events during seed germination. (*After J. D. Bewley and M. Black, Physiology and Biochemistry of Seeds, vol. 2, Springer-Verlag, 1982*)

adenine dinucleotide phosphate (NADPH)], and key metabolic intermediates.

Respiration occurs in four phases as the seed imbibes; the length of each phase varies with species and environment. Aerobic respiration increases rapidly with imbibition and lasts approximately 10 h. This rapid respiration results from physical activation of preexisting cellular systems. A lag in respiration occurs 10–25 h after the start of imbibition. The seed is now completely hydrated, preexistent enzymes have been activated, and loosening and extension of the cell walls begin. Some of the respiration at this point may be anaerobic, and this lag may be due to a lack of oxygen. The seed may be held at this stage and even desiccated without loss of viability. Once rehydrated, the seed starts germination at the same point at which it was desiccated. The growing radicle subsequently breaks through seed coat tissues, and there is a second burst of rapid aerobic respiration. Oxygen is plentiful to the radicle with the breaking of the seed coat, and activity of newly synthesized mitochondria and respiratory enzymes in dividing cells increases dramatically. Finally, respiration decreases with the depletion of stored metabolic reserves. *See* ADENOSINE TRIPHOSPHATE (ATP); BIOLOGICAL OXIDATION; CITRIC ACID CYCLE.

Protein–nucleic acid synthesis. Protein synthesis is a prerequisite for radicle emergence, while synthesis of deoxyribonucleic acid (DNA) occurs only after germination and is associated with seedling growth. Protein synthesis begins when embryo cells are hydrated sufficiently to allow for the assembly of the necessary ribosomes and messenger RNA, within a few minutes to hours after the start of imbibition. In isolated embryos of rye, rice, and wheat, and dissected axes of lima bean and french bean, protein synthesis commences within 30–60 min after the start of inbibition. Protein synthesis results in formation of metabolic enzymes also found in the developing seed, as well as unique, so-called germination proteins such as certain proteolytic enzymes.

Mobilization of storage reserves. Carbon and nitrogen for seedling growth prior to soil emergence (heterotrophic growth) are derived from stored seed reserves in the form of carbohydrates, oils, and protein. Carbohydrate is the first storage compound to be used for energy by the growing embryo, beginning with imbibition. Starch stored in the endosperm of monocots such as barley and corn is degraded by enzymes synthesized in aleurone cell layers and secreted to the endosperm during imbibition. The resulting sugar, glucose, is absorbed by the scutellum, enzymatically converted into sucrose, and transferred to the growing embryonic axis. The scutellum also may synthesize and secrete hydrolytic enzymes. In dicots such as pea, where starch is also the major storage carbohydrate, this carbon source is mobilized directly in the cotyledons and transferred as sucrose to the growing axis. In the case where oligosaccharide sugars (sucrose, raffinose, stachyose) are the main storage carbohydrate as in cucumber and soybean (**Fig. 8**), there is a rapid enzymatic hydrolysis of raffinose and stachyose to sucrose with a concomitant formation of starch. This starch is degraded after germination to provide additional carbon energy for the seedling. Mobilization of raffinose saccharides is controlled by a balance of supply and demand between the cotyledons and growing radicle. Formation of starch may be a regulating mechanism for this process.

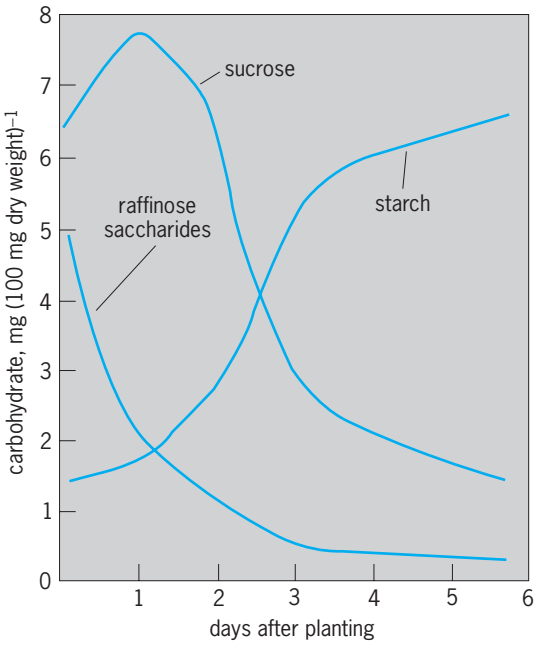


Fig. 8. Depletion of sucrose and raffinose saccharides (raffinose, stachyose) and accumulation of starch in etiolated soybean cotyledons when grown on moist germination paper. (*After C. S. Brown and S. C. Huber, Reserve mobilization and starch formation in soybean (Glycine max) cotyledons in relation to seedling growth, Physiol. Plant., 72:518–524, 1988*)

Genetic and hormonal control of carbohydrate mobilization in germinating seeds has been studied extensively in barley. Exogenously applied gibberellins and possibly those occurring endogenously appear to control the synthesis and release of hydrolytic enzymes for nutrient mobilization in the barley aleurone layer. For the starch-degrading enzyme α-amylase, gibberellins promote increases in an mRNA specific for this enzyme. Abscisic acid inhibits this gibberellin-induced synthesis of α-amylase and other hydrolytic enzymes.

Primarily as a postgerminative process, proteins and lipids are mobilized to supply nitrogen and carbohydrates, respectively, to the growing seedling. While protein is the main nitrogen source, RNA and amino acids also can contribute. Protein breakdown typically occurs in three stages. First, there is hydrolysis of proteins for the synthesis of additional hydrolytic enzymes for further protein degradation. Second is the hydrolysis of the main reserve proteins to provide amino acids for the growing seedling. Third, the seed storage tissue senesces and provides additional amino acids from cellular proteins and nuclear material for seedling growth. Oil-storing seeds typically are dicot species (for example, soybean, sunflower, and peanut), and mobilization of these reserves lags up to 2–3 days after imbibition. This

degradation then proceeds at a constant rate for 3–8 days. Mobilization of lipids is enzymatic, and cellular components (oil bodies, glyoxysomes, mitochondria, and cytosol) are involved in the conversion of lipids to carbohydrate for transport to the growing radicle. *See* CARBOHYDRATE; LIPID; PROTEIN.

Cadance A. Lowell

Storage. Seed storage is necessary for food production and conservation of genetic resources. Seed companies and farmers need to keep seeds alive for short periods to ensure production of a high-quality crop. Conservationists need to keep seeds alive for decades to ensure that valuable genes will not be lost even if a plant species or population becomes extinct in the wild. In an effort to slow the continued erosion of genetic diversity in cultivated and noncultivated plant species, national and privately owned gene banks are storing seeds as germplasm for the future.

Seeds possess a remarkable attribute which allows storage until the next growing season or the next century. They survive drying. Drying slows the molecular motions needed for chemical reactions. When seeds dry, their metabolism stops and they go into "suspended animation." Many species of seeds survive complete desiccation, so they can be stored under standard conditions and their life spans can be predicted using models. This type of seed is termed orthodox. All major food crops (corn, wheat, and soybeans) produce orthodox seeds. Recalcitrant seeds represent another category: those which do not survive complete drying. Recalcitrant seeds must remain metabolically active and so cannot be stored for more than a year. Temperate forest trees and plants from tropical rainforests or aquatic ecosystems often produce recalcitrant seeds. Many tropical fruits, such as cacao, mangos, and avocados, produce recalcitrant seeds. The ability to survive complete desiccation is acquired during the final days of maturation in orthodox seeds. When scientists discover the details of these processes, they may be able to store recalcitrant seeds more effectively. *See* CACAO; CORN; PLANT METABOLISM; SOYBEAN.

The best way to preserve biological material is to keep it dry and cold. In the early 1960s, J. F. Harrington developed "thumb rules" to describe the relative longevity of orthodox seeds under various storage conditions. He stated that the longevity of a seed doubles for every 1% reduction in water content or 5°C (9°F) reduction in temperature. Seed gene banks use these principles to keep seeds alive for decades. The International Plant Genetic Resources Institute in Rome recommends drying seeds to water contents of 3–7%, packaging them in airtight bags, and then placing them in a freezer at −18°C (−0.4°F). Oily seeds such as peanuts store best at 3% water, and starchy seeds such as wheat store best at 7%. Equilibrating seeds to 20% relative humidity will achieve the optimum moisture content no matter what the chemical composition.

The U.S. Department of Agriculture's National Seed Storage Laboratory (NSSL) in Fort Collins, CO, is the largest seed gene bank in the world. The NSSL stores over 415,000 accessions or bags of seeds, representing over 8000 species of plants. Each bag contains more than 2000 seeds, so there are literally millions of individual plants being stored. In 1977, NSSL scientists began to store some seeds in liquid nitrogen vapor (−150°C or −238°F) with the hope that these seeds would survive centuries, if not millennia.

Longevity. While storing seeds under dry, cold conditions can extend life spans, the innate longevity of a seed is determined by the seed quality. Seed quality is a poorly understood feature that is a product of both the genetic background and conditions of growth and harvest. Seed species have characteristic longevities. Species from Malvaceae, Onagraceae, Polygonaceae, and Leguminosae tend to produce long-lived seeds, while species from Compositae, Papaveraceae, and Liliaceae tend to produce short-lived seeds. Usually there is tremendous variability in longevity among cultivars of a species and among harvest years of the same cultivar. Environmental conditions such as drought, temperature, and nutrient deficiencies and premature or late harvests can adversely affect seed longevity. Extremes in longevity are noted in seeds collected from wild populations since growing conditions are variable from site to site and year to year.

Records of seed longevity are often anecdotal but indicate remarkable potential for long-term survival (see **table**). In some surveys, seeds survived decades in unrefrigerated conditions. W. J. Beal buried seeds from 23 different species in inverted bottles in the autumn of 1879 with the intention of measuring how long they would remain dormant in the

Results of some key studies in seed longevity

Longevity, years	Species	Storage conditions	Researchers and year of publication
1700–110	*Chenopodium album* and other species	Soil	S. Odum, 1965
1288–271	*Nelumbo nucifera*	Soil	J. Shen-Miller et al., 1995
295–75	*Carex bigelowii*	Soil	J. B. Mcgraw et al., 1991
158–25	500 species	Museum	P. Becquerel, 1907
100	23 species	Soil	A. Kivilaan and R. S. Bandurski, 1981
50	600 species	Museum	A. J. Ewart, 1980
60–45	15 vegetable species	5 to −18°C	E. E. Roos and D. A. Davidson, 1992
50	98 species	Vacuum tubes, 15–25°C	F. W. Went, 1969 (unpublished data)
39	107 species	Soil	E. H. Toole and E. Brown, 1946
25	17 crucifer species	Vacuum tubes, −5°C	R. H. Ellis et al., 1993
20	9 species	−15°C	C. M. Rincker, 1983

soil. After 100 years, some seeds were still able to germinate. The sacred lotus (*Nelumbo*) is most often and most reliably reported as the longest-living seed. Seeds from herbarium specimens collected 250 years ago still germinated, and germinable seeds from the soil were carbon-dated at over 1200 years (see table). Archeological dating of soil samples collected along with germinable seeds of *Chenopodium album* suggested this sample may have lived for 1700 years. There also are claims that *Lupinus arcticus* survived for 10,000 years in the permafrost of Alaska, but the evidence to support this claim is weak. There is no scientific evidence for the popular myth of seeds surviving for thousands of years in Egyptian tombs.

Aging. Even when stored under optimum conditions, seeds eventually succumb with time. As they age, seeds lose the ability to tolerate stress, then they germinate slowly or produce abnormal seedlings, and finally they die. The cells in aging seeds change: membranes become leaky, enzymes lose catalytic activity, and chromosomes accumulate mutations. It was once proposed that seeds consume food reserves during storage, but the huge amount of reserves contained in seeds and the low level of metabolism measured in dried seeds challenge this hypothesis. Only a few chemical and physical reactions can occur when seeds are very dry, and most scientists believe that peroxidative reactions, in which free radicals arise from activated oxygen and attack cells, cause the deterioration in dry seeds. Dry seeds cannot repair the lesions to their cell machinery because metabolism has been shut off. It is ironic that the ability to stop metabolism allows seeds to survive for a very long time, but also leads to their eventual death.

Darren H. Touchell; Christina Walters; Eric E. Roos

Bibliography. J. D. Bewley and M. Black, *Physiology and Biochemistry of Seeds in Relation to Germination, Viability, Dormancy and Environment Control*, 1982; J. D. Bewley and M. Black, *Seeds: Physiology of Development and Germination*, 2d ed., 1994; L. S. Dure, Seed formation, *Annu. Rev. Plant Physiol.*, 25:259–278, 1975; O. L. Justice and L. N. Bass, Principles and Practices of Seed Storage, *Agri. Handb.*, no. 506, 1978; J. Kigel and G. Galili (eds.), *Seed Development and Germination*, 1995; A. M. Mayer and A. Poljakoff-Mayber, *The Germination of Seeds*, 4th ed., 1989; D. R. Murray (ed.), *Seed Physiology*, vol. 2: *Germination and Reserve Mobilization*, 1984; D. A. Priestly, *Seed Aging*, 1986; T. F. Stuessy and S. H. Sohmer (eds.), *Sampling the Green World: Innovative Conecepts of Collection, Preservation, and Storage of Plant Diversity*, 1996; J. R. Wilcox (ed.), Soybeans: Improvement, Production, and Uses, *Amer. Soc. Agron. Monog.*, no. 16, 1987.

Seiche

A short-period oscillation in an enclosed or semienclosed body of water, analogous to the free oscillation of water in a dish. The initial displacement of water from a level surface can arise from a variety of causes, and the restoring force is gravity, which always tends to maintain a level surface. Once formed, the oscillations are characteristic only of the geometry of the basin itself and may persist for many cycles before decaying under the influence of friction. The term "seiche" appears to have been first used to describe the rhythmic oscillation of the water surface in Lake Geneva, which occasionally exposed large areas of the lake bed that are normally submerged. *See* WAVE MOTION IN LIQUIDS.

Behavior. The most straightforward example of a surface seiche is that occurring in a long, narrow lake of uniform depth; in this case the water surface elevations have a simple physical analogy with the vibrations of a plucked string. Since the wave length is long relative to the water depth, seiche behavior conforms to the theory of shallow water waves. Thus the velocity is given by $\sqrt{gh}$, where g is the acceleration of gravity and h is the water depth. A standing wave can be represented by the sum of two waves traveling in opposite directions, each being reflected at the ends of the basin. The natural period of oscillation of a uniformly deep, completely enclosed rectangular basin is given by Merian's formula (1), where

$$T = \frac{2L}{n\sqrt{gh}} \tag{1}$$

L is the length of the basin and n is the number of nodal lines present, one for the fundamental or uninodal seiche, two for the binodal seiche, and so forth (**Fig. 1**). However, the formula is only approximate, and calculation of the higher modes of oscillation starts to depart seriously from that given by Eq. (1) due to the importance of nonuniformity of the basin shape.

The **table** gives examples of observed periods of surface seiches in different lakes. Figure 1 shows profiles of a standing wave in a closed basin of the fundamental, binodal, and trinodal seiche. Below the nodes there is maximum horizontal and no vertical motion; below the antinodes there is maximum vertical and no horizontal motion. **Figure 2** shows simultaneous records of water height near each end of Lake Vättern during a uninodal seiche. The two measurements show opposite movement of the water at each antinode. When several different seiches occur in a basin, their effects are superposed. Thus it is quite possible, for example, for a vessel to be located at the node of the fundamental seiche but at the antinode of the binodal seiche. The amplitudes of the different harmonics are determined by the initial disturbance.

It is also possible for transverse seiches to exist together with the longitudinal seiche. The appropriate generalization of Merian's formula for a deep, rectangular basin of length L and breadth B is given by Eq. (2), where $\alpha = L/B$ and m and n define the harmonic nodes in the transverse and longitudinal directions, respectively. However, the earlier comments

$$T_{mn} = \frac{2L}{\sqrt{gh}}\,(\alpha^2 m^2 + n^2)^{-1/2} \tag{2}$$

regarding nonuniformity of basin shape still apply, and Eq. (2) cannot be generally applicable to natural basins.

Seiches are also a common feature of semienclosed basins such as bays, gulfs, and harbors, the natural period for an open-end rectangular basin of uniform depth being twice that of the same-size basin closed at both ends. The node and antinode occur at the open and closed ends, respectively. A problem of great practical significance is the oscillation of water in harbors, which can generate strong and unpredictable currents affecting the safety of ships entering or leaving the harbor; under extreme conditions these can also cause damage to moored vessels.

Generation and decay. Seiches can be generated when the water is subject to changes in wind or atmospheric pressure gradients or, in the case of semienclosed basins, by the oscillation of adjacent connected water bodies having a periodicity close to that of the seiche or of one of its harmonics. Other, less frequent causes of seiches include heavy precipitation over a portion of the lake, flood discharge from rivers, seismic disturbances, submarine mudslides or slumps, and tides. The most dramatic seiches have been observed after earthquakes.

The amplitude and persistence of the resulting seiche depend not only on the magnitude of the energy source but also on the energy losses within the water body. Such losses include dissipative effects resulting from friction on the sides or bottom of the basin. For semienclosed basins, energy can also be lost by the radiation of waves away from the mouth. If the seiche is generated by an impulsive event such as a sudden change in atmospheric pressure gradient, the seiche amplitude is seen to decay by a nearly constant fraction with each succeeding period. The table shows, for example, that in Lake Geneva the decay has been estimated to be about 3% with each seiche period, whereas in Lake Erie it is 32%. In general, the rate of decay is greater for basins that are shallow or have narrow constrictions and complex topography.

In bays and gulfs a dominant source of energy is the tide. Since there is quite a rich array of tidal frequencies, it is sometimes possible to determine the natural seiche frequency of a bay simply by observing the amplification of different tidal components. If the tidal period is close to that of the seiche period for the bay, resonance increases the tidal amplitude.

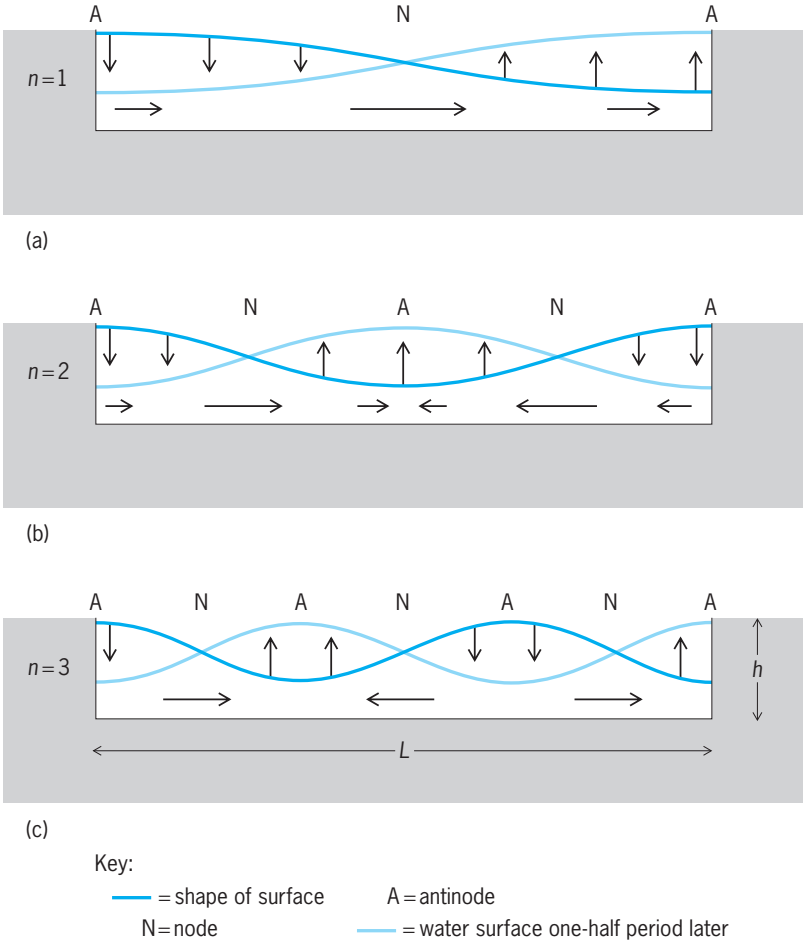


Fig. 1. Schematic diagram of a (*a*) uninodal, (*b*) binodal, and (c) trinodal seiche in a rectangular basin. Vertical and horizontal vectors show the direction of flow corresponding to the indicated change of water surface. Vertical movements are greatly exaggerated.

An example of this is the Bay of Fundy (Canada), where the tide can exceed 45 ft (15 m). The bay has a fundamental mode seiche period of about 13.3 h, close to the semidiurnal lunar tidal period of 12.4 h. A similar, near-resonance exists in the Gulf of Mexico, where the period of the fundamental seiche mode is close to 24 h, with resulting amplification of the diurnal tide. In smaller bays and harbors the generation of seiches has been attributed to surf beats. This effect is due to the coupling of radiation stress, in effect the thrust of the waves on the coast, with

Observed surface seiche periods in typical lakes and observed decay rates

Lake and location	Observed periods of oscillation, min						Fractional decrease in amplitude of free uninodal seiche over each successive period
	T_1	T_2	T_3	T_4	T_5	T_6	
Geneva (Switzerland–France)	74.0	35.5					.030
Garda (Italy)	42.9	28.6	21.8	15.0	12.1	9.9	.045
Loch Earn (Scotland)	14.5	8.1	6.0	4.0	3.5	2.9	
Erie (United States–Canada)	858.0	542.3	350.9	250.5			.322
Königssee (Germany)	11.6						.204
Vättern (Sweden)	179.0	97.5	80.7	57.9	48.1	42.6	.113
Yamanaka (Japan)	15.6	10.6	5.5				.099
Baikal (Russia)	278.2						

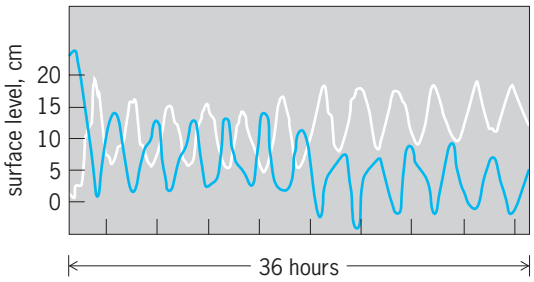


Fig. 2. Simultaneous recordings taken near each end of Lake Vättern during a uninodal seiche. The reading at one end is a mirror image of the observed displacement at the other end. 1 cm = 0.4 in. (*After A. Defant, Physical Oceanography, vol. 2, Pergamon, 1961*)

rhythmic changes in the height of the wind waves or swell. Since the swell can originate from distant storms, the existence of seiches is not necessarily closely correlated with local meteorological conditions. Coastal seiches can also be generated by traveling fronts or pressure disturbances, especially when the atmospheric disturbance travels at a speed close to that of the gravity wave ($\sqrt{gb}$). A less frequent but often important generation mechanism in harbors is the tsunami or seismically formed ocean wave. *See* TSUNAMI.

Effect of rotation. The response of large bodies of water is significantly influenced by the Coriolis effect produced by the rotation of the Earth. The effective period of the rotation decreases with latitude and is given by $^1/_2\ \tau/\sin\lambda$, where τ is the period of rotation of the Earth (24 h) and λ the latitude. The Coriolis force acts at right angles to the direction of motion of the water, to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. The Coriolis force introduces a transverse slope to each of the traveling waves that combine to form a standing wave pattern, or seiche. The transverse motion, added to the motion along the channel, results in the movement of water in an elliptical path rotating once with each seiche period. The transverse motion has the effect of shrinking the nodal lines of zero vertical motion to small areas known as amphidromic points.

In the plan views of **Fig. 3***a* and *b*, arrows show surface currents at one-fourth and three-fourth cycles, respectively. After high water at the channel head, Coriolis deflection, indicated by broken arrows, causes an increase of water level in the right half of the current in both Fig. 3*a* and *b*, with a corresponding decrease to the left. The broken line is one of mean level with respect to the deflection. Water profiles along nodal lines are shown beneath Fig. 3*a* and *b*. Point O, the amphidromic point, is thus the only position of no water level change. Line O-N (O-N′) connects points having simultaneous high (low) water. Other such cotidal lines are shown in Fig. 3*c* for different phases of one seiche oscillation. The superposition of the transverse motion on the primary longitudinal motion causes this line to rotate counterclockwise (dotted arrow) for one complete oscillation in the Northern Hemisphere. *See* CORIOLIS ACCELERATION; TIDE.

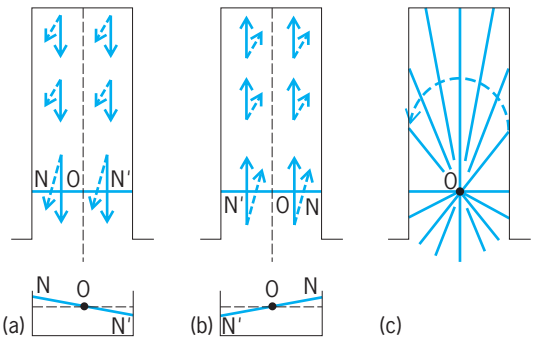


Fig. 3. Schematic representation of transverse oscillations in a bay in the Northern Hemisphere leading to the development of an amphidromic point. (*a*) Ebb current. (*b*) Flood current. (*c*) Cotidal lines.

Internal seiches. Internal seiches are standing waves that occur in water that has vertical variations in temperature or salinity resulting in a vertical density stratification. The waves produce vertical movements of water at depth, with very little vertical motion at the surface. They are typically generated by the piling up of water at one end of the lake by wind, although in semienclosed basins they may be forced by oscillations in the density structure at the mouth of the basin. **Figure 4** shows the measurements of two isotherm depths in Lake Alpnach (Switzerland) generated by a wind pulse as well as by diurnal wind forcing. The oscillations decay by one-third over about 1 day. As for surface seiches, the residence time of the energy varies, depending on the excitation (tens to hundreds of joules per square meter) and the depth of the basin, from less than 1 day up to several weeks.

Since the relative density differences within the water column are much smaller than the density difference at the air-water interface, the same amount of energy produces internal seiches of much greater amplitude than the corresponding surface seiche. However, the internal wave speed is much slower than the surface wave. The relative density difference in lakes during the summer is typically of the order of 0.002, so that wind effects, which might produce surface seiches of only a few centimeters, can produce internal seiches of meters to tens of meters (Fig. 4). In the Great Lakes of North America and the long and deep Rift valley lakes of Baikal and East Africa, amplitudes can reach almost 100 m with periods of many weeks to months.

In natural waters, the two-layer approximation, as depicted for surface seiches in Fig. 1, is a simplification, and in nature the density varies continuously with depth. Subsequently, the density structure allows several vertical modes, with characterizing vertical wavelengths, to exist. Such an example is given in Fig. 4 where, beside the fundamental seiche period of about 8 h, a second vertical mode with a 24-h periodicity is prominently displayed. The characteristic of the second vertical mode is the periodic widening and thinning of the middle layer (between 11 and 17°C in Fig. 4). The fluid of the lighter top layer (above 17°C) and the fluid in the heavier deep layer (below 11°C) move in phase, whereas the fluid

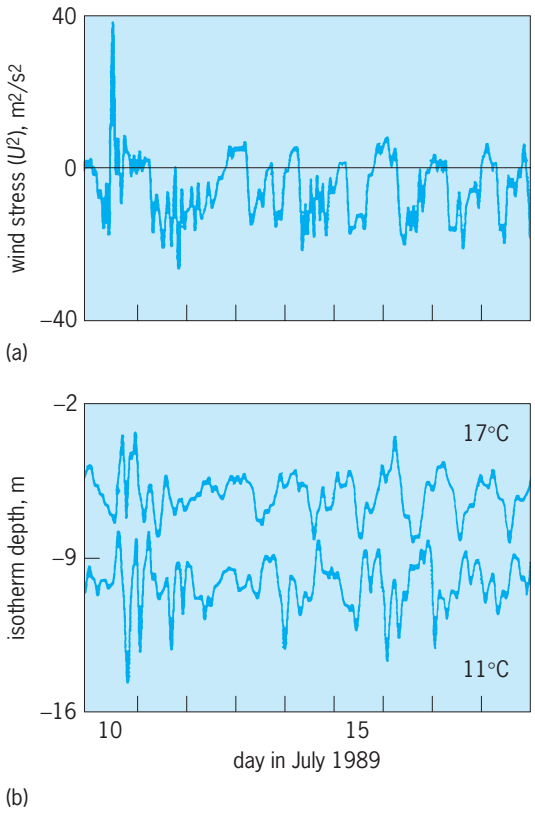


Fig. 4. Seiche measurements in Lake Alpnach, in Switzerland. (*a*) Time series of the lake-parallel component of the square of the wind speed (proportional to the wind stress) over the lake (maximum depth = 35 m). (*b*) Times series of the two selected isotherms of 11 and 17°C. The fluid layer between the two lines shows a strong second vertical mode characteristics of 24-h periodicity (V2H1). The fundamental mode (V1H1) of an 8 h period is also prominent. (*After M. Münnich, A. Wüest, and D. M. Imboden, Observations of the second vertical mode of the internal seiche in an alpine lake, Limnol. Oceanogr., 37:1705–1719, 1992*)

in the middle layer has the exactly opposite direction. Higher vertical resolution (by dividing the water column in more homogeneous layers) leads to even more vertical modes, generally expressed by *VpHq*, where *p* indicates the *p*-th vertical mode, and *q* the *q*-th horizontal mode. According to Eq. (2), the horizontal modes are combinations of the two horizontal coordinates (modes *m*, *n*). The Earth's rotation also alters the structure of internal seiches since the transverse slope of the internal interface induced by the Coriolis force is greatly increased for internal seiches.

Which modes are excited, and to what extent, in a particular water body depends on the forcing and the stratification, but is almost impossible to predict. As in the case of surface seiches, internal seiching is a mechanism that selectively extracts energy from the forcing (mostly wind) at a frequency close to that of the internal seiche. For example, as shown in Fig. 4, the second vertical mode in Lake Alpnach is resonantly forced at a 24-h period by the local diurnal wind.

The latter example demonstrates that the amplitude of internal seiches can reach a significant fraction of the basin depth. This enormous amplitude results in a distortion of the smooth sinusoidal perturbation of the interface predicted by linear theory, and the peak and trough of the wave travel at slightly different speeds. In addition, interaction with the topography and bottom friction leads to turbulence and the generation of short-period internal waves radiated into the water body. Especially in large basins and lakes, where resonant seiching is difficult to reach, most internal energy is contained in progressive short-period waves. Internal shear, bottom friction, and wave breaking leads to damping of the internal seiche energy. As a rule, the decay rate of the energy is about 1 day per 40 m of water depth (40 days in Lake Baikal). However, for very weak and almost linear seiching, damping may reduce the decay rate.

The enormous amplitude of internal seiches results in a distortion of the smooth sinusoidal perturbation of the interface predicted by linear theory. In effect, the peak and trough of the wave travel at slightly different speeds, resulting in steepening of a portion of the interface. Under the right conditions, the steepened interface can break down into a series of shorter waves, possibly accompanied by turbulence and mixing and usually referred to as an internal surge. Alfred Wüest; David M. Farmer

Bibliography. J. R. Apel, *Principles of Ocean Physics*, 1987; A. Defant, *Physical Oceanography*, vol. 2, 1961; J. Miles, Harbour seiching, *Annu. Rev. Fluid Mech.*, 6:17–36, 1974; M. Münnich, A. Wüest, and D. M. Imboden, Observations of the second vertical mode of the internal seiche in an alpine lake, *Limnol. Oceanogr.*, 37:1705–1719, 1992; B. W. Wilson, Seiches, *Advan. Hydrosci.*, 8:1–94, 1972.

Seismic exploration for oil and gas

Prospecting for oil and gas using exploration seismology, a geophysical method of determining geologic structure by means of prospector-induced elastic waves. By studying body waves such as compressional and shear waves propagating through the Earth's interior, the constituent and elastic properties of its solid and liquid core, its solid mantle, and its thin crust are defined. The major differences between earthquake seismology and petroleum exploration seismology are scales and knowledge of the location of seismic disturbances. Earthquake seismology studies naturally generated seismic waves, which have periods in minutes and resolution in kilometers. In exploration seismology, artificial sources are used that have periods of tenths of a second and tens of meters of resolution. Production seismology requires higher-frequency seismic waves and better resolution, often resolution in the order of a few meters. *See* EARTH; EARTH INTERIOR.

Seismic technology. Computer technology allows resolution of some of the theoretical complexities of elastic wave propagation so that deeper insight into the wave field phenomena can be obtained. The availability of a large number of channels in the recording instrument facilitates three-dimensional

and three-component acquisitions. Powerful supercomputers allow manipulation of larger and larger data sets, and they have facilitated display and interpretation of them as a single data unit through the use of advanced computer visualization techniques. *See* COMPUTER; SUPERCOMPUTER.

The availability of powerful workstations led to the wide use of interactive processing and interpretation. When such is coupled with technically advanced algorithms, the amount of information that the interpreter can obtain from the data increases significantly. Better quality control is provided, fine-tuning analysis is achieved more easily, and the data can be enhanced to meet specific objectives. As a result of technological advances, the working relationship among professionals of different disciplines has changed. Processing specialists, interpreters, geologists, and engineers use workstations concurrently to test new exploration concepts through computer simulations or modeling that can be validated through special processing of the seismic data. If there are discrepancies between the model and the real data, a hypothesis can be proposed based on information derived from the data. This process can be iterated until the Earth model derived is consistent with all available surface and subsurface geophysical, petrophysical, geological, and engineering data sets. *See* ALGORITHM; MODEL THEORY; SIMULATION.

Seismic method. The seismic method as applied to exploration of oil and gas involves field acquisition, data processing, and geologic interpretation.

Field acquisition. Seismic field acquisition requires placement of acoustic receivers (geophones) on the surface in the case of land exploration, or strings of hydrophones in the water in the case of marine exploration. The conventional geophone is a vertical-component velocity phone. A triaxial phone with two horizontal components may be used if shear-wave recording is desirable. A group of several receivers is generally used for each receiver point location. *See* HYDROPHONE.

After all instruments have been placed, an artificial earthquake is initiated by an impulsive or a vibratory source at the shot point. This generates elastic seismic waves that propagate through the subsurface. Commonly used sources are dynamite, vibrators, and air guns. These elastic seismic waves travel through different geologic strata at various speeds according to the acoustic properties of the composite rocks and minerals as well as the composite fluids in the reservoirs under various conditions. When these seismic waves arrive at the interfaces of acoustic impedance contrasts (impedance is a product of velocity and density), part of the energy is transmitted and part of the energy is reflected. The reflected waves eventually emerge back at the surface and are picked up by the receivers. The electric signal generated by the receivers is recorded digitally on magnetic tapes or cassettes in a multichannel seismic recorder, ready for preliminary processing in the field or further enhancement in processing centers.

Data processing. Seismic data processing is usually done in large computing centers with digital mainframe computers or a large number of processors in parallel configurations. Seismic data processing has several goals. One aspect of the processing is to compensate for wave propagation effects such as spherical spreading and attenuation, and to account for the time differences for the ray paths between different source-to-receiver distances. This includes seismic velocity estimation from the data. The processing also accounts for effects generated by different shots and the near-surface conditions such as the varying thickness and lithologic composition of the weathered layers. In addition, it ensures spatial and temporal consistency of the propagation wavelet. Seismic data processing can also improve the signal-to-noise ratio. Random noise is attenuated by the stacking of many redundant traces after normal moveout correction, which is the correction that has to be applied to seismic data to take into account the variation of reflection arrival time because of variation in the shotpoint-to-geophone distance. Organized noises such as surface waves, wind noise, repeated refraction, and multiples are filtered according to the principles of communication or information theory. Temporal and spatial resolution are achieved by using techniques known as deconvolution and migration, respectively. Deconvolution is an inverse filtering process that takes into account the characteristics of the shot signatures as well as some of the propagation effects of the wavelet as it propagates through the subsurface. Migration is a process that takes into account the true spatial position of reflectors. Migration ensures proper imaging of the reflective strata when it is performed correctly.

The end result of seismic data processing is the production of a subsurface profile similar to a geologic cross section. It is commonly plotted in a time scale, but it is also possible to plot it in depth. The accuracy of the depth plot, however, varies with the skill of the processor, the quality of data, and the knowledge of the prospecting area such as formation thickness and velocity information. These time or depth profiles are used for geologic interpretation.

Geologic interpretation. Geologic interpretaion of seismic data has two key components, structural and stratigraphic. Structural interpretation of seismic data involves mapping of the geologic relief of different subsurface strata by using seismic data as well as information from boreholes and outcrops. Stratigraphic interpretation looks at attributes within a common stratum and interprets changes to infer varying reservoir conditions such as lithology, porosity, and fluid content. These seismic attributes can be amplitude, velocity, travel time, absorption, frequency, phase, or amplitude-versus-offset, an interpretation technique that uses the difference in reflection strength as a function of offset (source-to-receiver separation) to infer reservoir properties. The interpretation of reservoir parameters from seismic data is an inversion process that is inherently nonunique. Information from other disciplines such as petrophysics, rock physics, and engineering must therefore be used as constraints for stratigraphic interpretations.

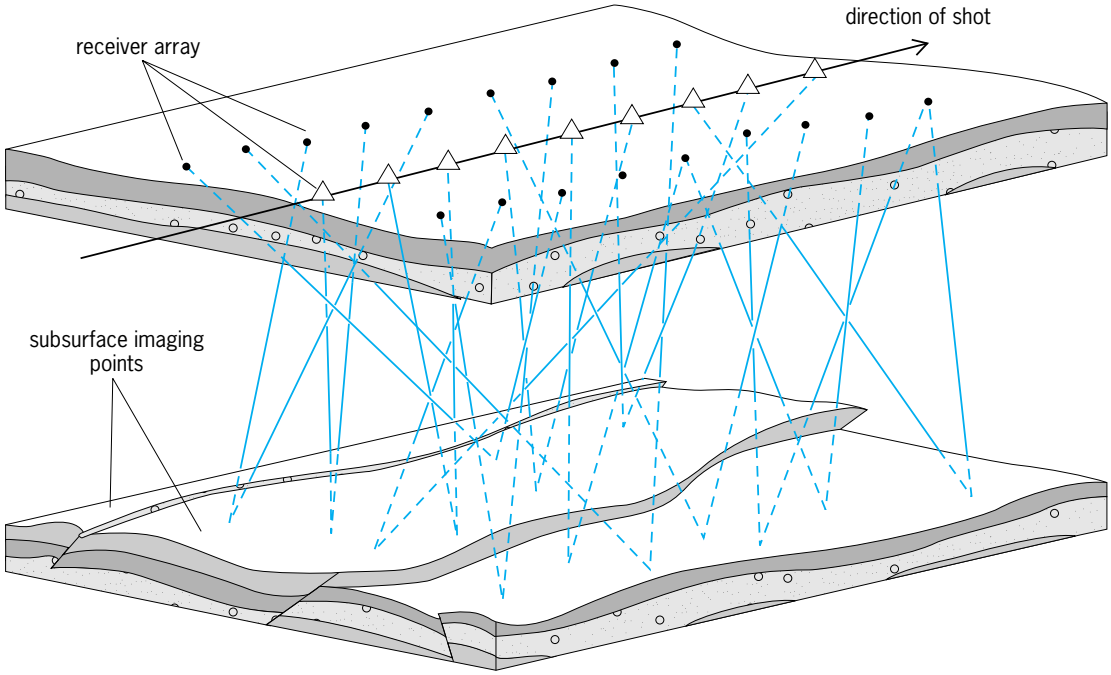


Fig. 1. Layout for a three-dimensional seismic survey.

Two- and three-dimensional seismic. Historically, surface seismic acquisition is done by placing sources and receivers along a straight line so that it can be assumed that all the reflection points fall in a two-dimensional plane formed between the line of traverse and the vertical. This is known as two-dimensional seismic.

Three-dimensional seismic is a method of acquiring surface seismic data by placing sources and receivers in an areal pattern (**Fig. 1**). One example of a simple three-dimensional layout is to place the receivers along a line and shoot into these receivers along a path perpendicular to this line. The common midpoints of a three-dimensional survey also form an areal surface. When all these traces or time series are displayed, they form a closely spaced data volume that represents an acoustic image of a three-dimensional subsurface. However, because of the large data set involved, display and processing of a three-dimensional seismic survey requires the use of workstations that include sophisticated computer graphics. *See* COMPUTER GRAPHICS.

A three-dimensional seismic survey provides a more accurate and detailed image of the subsurface. It offers significantly higher signal quality than the two-dimensional data commonly acquires. It also improves both spatial and temporal resolutions. The three-dimensional seismic technique is being applied to exploration and production of oil and gas, accounting for more than half of the seismic activity in the Gulf of Mexico and North Sea.

Three-component seismic. This refers to seismic surveys where either three-component receivers or three-component sources, or both, are being used.

In seismology, there are two modes of body-wave propagation in the subsurface. These waves are labeled primary (P), or compressional, and secondary (S), or shear. In the P wave, the volume of the rock changes, but not the shape, during the passage of the wave in response to alternating compressive and tensional stresses. The S wave, changes shape but not volume in response to the passage of shear stresses. The shear waves move perpendicular to the direction of propagation. The speed at which each body wave travels is a function of the elasticity of the rock.

Conventional seismic typically uses velocity phones in the vertical direction only. They are adequate for P-wave exploration. To record shear waves in three dimensions, however, three-component geophones are required. It is common practice to place the three-component phones orthogonally in the vertical direction and in the directions parallel and perpendicular to the direction of shooting.

If only P-wave sources are used in the survey, then only converted shear waves are recorded by using the three-component phones. Some surveys involve both P-wave and S-wave sources in three orthogonal directions, producing a total of nine seismic sections, one for each source-receiver direction. This is sometimes called elastic seismic or vector seismic. Such a survey contains azimuthal information related to additional constants of elasticity.

Production or reservoir seismology. Production seismology is the application of seismic techniques to problems related to the production and exploitation of petroleum reservoirs. The internal heterogeneity of a petroleum reservoir has a profound influence on its production performance. The internal structure of a reservoir is complex and generally poorly understood. Structural deformations, fractures, lithologic variations, and diagenetic alterations contribute to the creation or destruction of conduits and barriers to fluid flow. In addition, it may be important to

monitor changes in fluid flow or composition during the producing life of the field and in an enhanced oil recovery process. Since production geophysics is the only effective method available that can image the reservoirs under in-place conditions, it has become an active field of applied research aimed at improving descriptions and understanding of reservoirs and their fluid flow behaviors. *See* PETROLEUM ENHANCED RECOVERY; PETROLEUM RESERVOIR ENGINEERING.

The use of seismic waves to image detailed rock and reservoir properties is hampered by limited resolution. As the elastic seismic waves propagate through the Earth, they attenuate because of spherical divergence. Their higher-frequency components, however, are even more attenuated because of scattering and absorption, resulting in a drastic decrease in resolution. For this reason the majority of successful cases in production seismology have been limited to close-spacing imaging such as is found in shallow reservoirs or reservoirs very close to the wellbore.

By placing acoustic sources at appropriate distances from the receivers in one or more boreholes, higher-frequency seismic signals can be propagated further in the subsurface, resulting in a greater improvement in resolution. Borehole geophysics makes possible the mapping of reservoir heterogeneity, monitoring of secondary or tertiary recovery, and their use as proximity indicators to possible oil and gas accumulations.

Vertical seismic profiling. This is a measurement procedure in which a seismic signal generated at the surface of the Earth is recorded by geophones secured at various depths to the wall of a drilled well. When the source is directly at the top of the well, it is called zero-offset vertical seismic profiling; otherwise it is known as offset vertical seismic profiling (**Fig. 2***a*).

A primary result of the vertical seismic profiling survey is the time-to-depth relationship from which interval velocities as a function of depth can be determined. A zero-offset vertical seismic profiling survey provides a link between sonic log and surface seismic information in two important ways. First, it calibrates the integrated sonic log times to the vertical seismic profiling times, which are directly related to the seismic times. Synthetic seismograms generated by using calibrated sonic logs match the surface seismic data better. Second, the zero-offset vertical seismic profiling links seismic reflection character in time to the well information in depth. It helps to correlate the seismic response to the lithologic interfaces. Vertical seismic profiling also helps interpreters to understand and resolve some of the ambiguities and uncertainties associated with processing and interpretation of surface seismic data. *See* WELL LOGGING.

Offset vertical seismic profiling can also be used as a predictive tool for exploration. It can image the reservoirs ahead of a drill bit, investigate reservoir conditions several hundred meters around the wellbore, or monitor the propagation of a steam front in a thermal recovery process.

Reverse vertical seismic profiling. This is a reverse configuration where the source is placed at a given depth in a borehole and the receivers are deployed in different patterns at the surface (Fig. 2*b*). Although it is a more time-efficient and cost-effective method of downhole imaging. It requires a powerful source to be effective.

Cross-well tomography. Cross-well seismology is a method of seismic imaging where both the sources and receivers are placed in separate boreholes (Fig. 2*c*). By placing both sources and receivers at depths below the highly attenuating near-surface materials such as the weathered layers, seismic waves in frequencies up to several kilohertz can be propagated, resulting in a great improvement of resolving power. The limiting factor is the amount of acoustic energy that can be generated from downhole sources without risking the safety of the wellbore. This in turn limits the distance between the wells at which surveys can be safely conducted.

In a typical cross-well seismic survey between two wells, travel time and amplitude measurements for each source-receiver location across the wells are made. As each ray of transmissive seismic wave

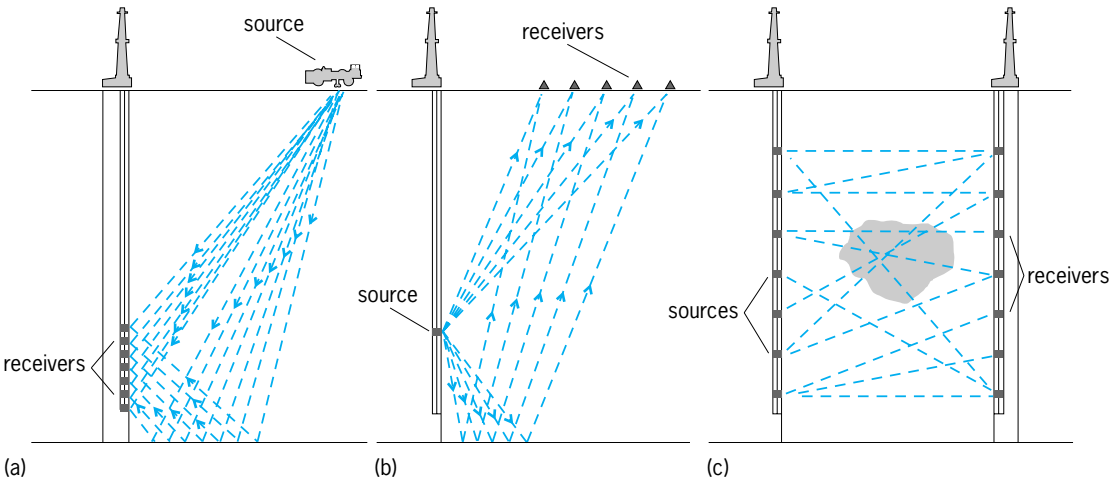


Fig. 2. Borehole seismic geometry. (*a*) Offset vertical seismic profiling. (*b*) Reverse vertical seismic profiling. (*c*) Cross-well tomography.

passes through the reservoir under investigation, it is modified according to the elastic properties of the reservoir. Tomography is an inverse processing technique that is used to reconstruct a two-dimensional image of these reservoir properties. Tomographic reconstruction techniques have been developed and applied in medical science for more than two decades. Conventional tomography assumes straight ray paths. When the effect of diffraction and scattering are considered in the mathematical model, it is called diffraction tomography. When reflections from different strata are considered, it is called reflection tomography. Cross-well seismic tomography has been successfully used to estimate velocities between wells, mostly for the purposes of monitoring enhanced oil recovery. *See* COMPUTERIZED TOMOGRAPHY.

Single-well seismic imaging. This is a technique where both seismic source and receiver are deployed in a single wellbore. The power of the source and the separation between source and receiver determine the depth of investigation.

Real-time seismic monitoring. This seismic technique is used to monitor the change in the conditions of a reservoir as it is producing through time by repeated surveys. It has been demonstrated to work most effectively in heavy-oil thermal recovery projects where steam front progressions have been successfully monitored by using repeated three-dimensional or cross-well seismic surveys. *See* GEOPHYSICAL EXPLORATION; PETROLEUM GEOLOGY; SEISMOLOGY.

Tai. P. Ng

Bibliography. J. K. Costain and C. Coruh, *Basic Theory in Reflection Seismology (Handbook of Geophysical Exploration Seismic Exploration)*, 2004; R. E. Sheriff (ed.), *Reservoir Geophysics*, 1992; M. A. Slawinski, *Seismic Waves and Rays in Elastic Media (Handbook of Geophysical Exploration Seismic Exploration)*, 2003; O. Yilmaz, *Seismic Data Processing*, 1988.

Seismic risk

The probability that social or economic consequences of earthquakes will equal or exceed specified values at a site, at several sites, or in an area, during a specified exposure time. Historically the term seismic risk has been used to describe an assortment of earthquake effects that range from ground shaking, surface faulting, and earthquake-induced landsliding to economic loss and casualties. As more quantitative methods for estimating the effects of earthquakes have been developed, terminology has become more precise. Although the term seismic risk is still sometimes used in a general sense to mean the potential for both the occurrence of natural phenomena and the economic and life loss associated with earthquakes, it is useful to differentiate between the concepts of seismic hazard and seismic risk. Seismic hazard may be defined as any physical phenomena (for example, ground shaking or ground failure) that are associated with an earthquake and that may produce adverse effects on human activities.

The exposure time is the time period of interest for seismic hazard or risk calculations. In practical applications, the exposure time may be considered to be the design lifetime of a building or the length of time over which it is of interest to estimate numbers of casualties. For example, the risk that a certain number of casualties will occur is the probability associated with all combinations of seismic hazards that may result from the earthquakes and all possible numbers of casualties resulting from those hazards. A general expression for the total risk is given by the equation below, in which $P(R_i)$ is the probability that the state

$$P(R_i) = \sum_{\text{all } j} P(R_i|S_j)P(S_j)$$

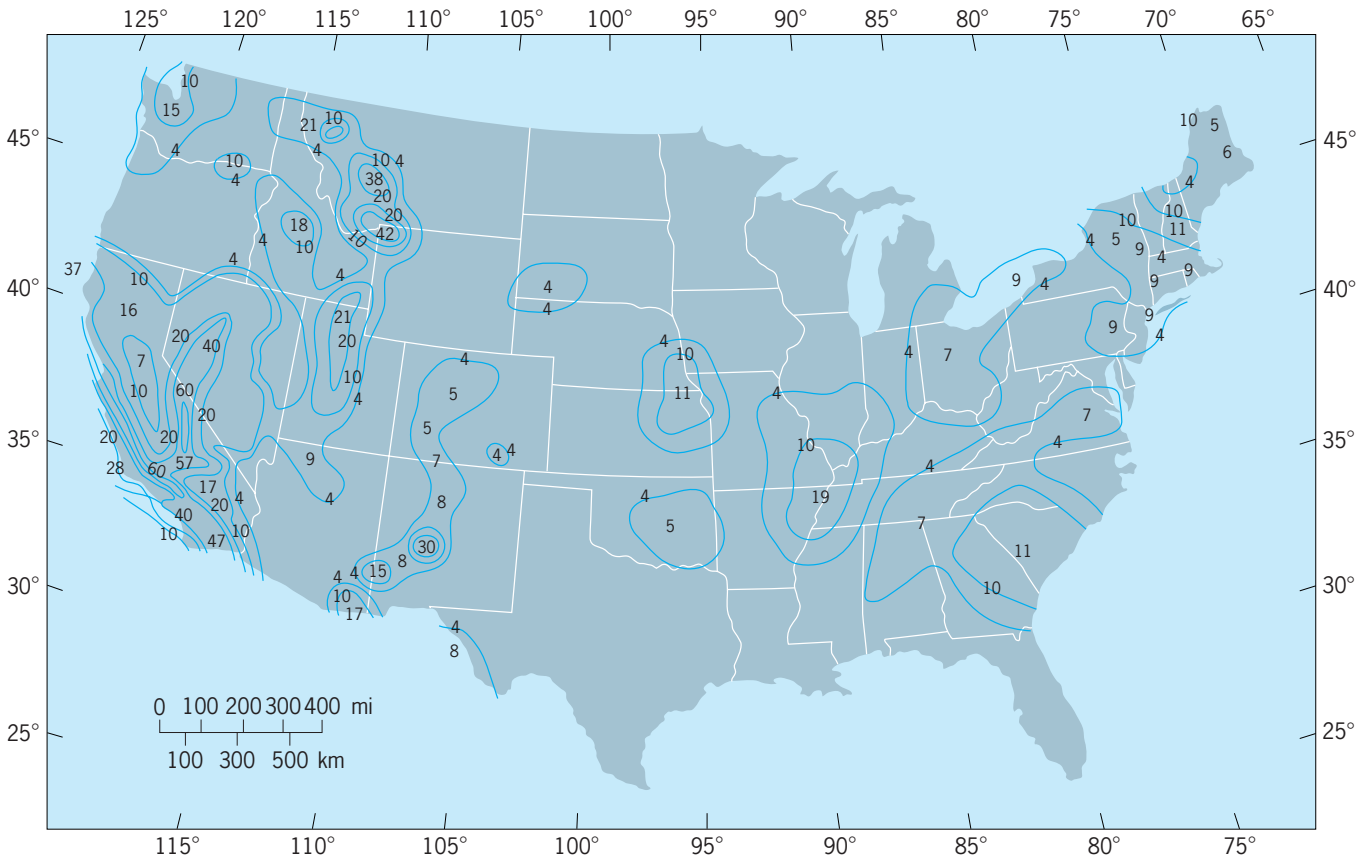
of the system is i, S_j means that the seismic hazard is level j, $P(S_j)$ is the probability that seismic hazard is level j, and $P(R_i|S_j)$ is the probability that the behavior state of the system will be R_i, given that seismic input, S_j, takes place. Thus, evaluation of seismic risk requires that the seismic hazard be specified.

Seismic hazard analysis. A seismic hazard can be any one of various physical phenomena that result either from surface faulting during shallow earthquakes or from the ground shaking resulting from an earthquake.

Physical phenomena. Surface faulting or rupturing of the Earth's surface usually occurs only in shallow earthquakes (12 mi or 20 km depth or less), but is not observed in all shallow earthquakes. The relative displacement of the two sides of a fault and the fault length both increase approximately exponentially with earthquake magnitude. The relative displacement between the two sides of a fault during great earthquakes may be as much as 40 to 90 ft (20 to 30 m), and the fault length in a single event may be 180 to 240 mi (300 to 400 km).

Landslides may be triggered by earthquake ground shaking, and under certain conditions can be catastrophic. The magnitude 7.4 earthquake that occurred off the coast of Peru in 1970 triggered a landslide above the village of Yungay in the Peruvian Andes, which resulted in the death of more than 15,000 people. Ground shaking also may cause a complex type of ground failure known as liquefaction, that is, the transformation of a granular material from a solid state into a liquefied state as a consequence of increased pore-water pressure. Liquefaction is known to cause widespread damage when concomitant to large earthquakes, as has been exemplified at Puerto Montt, Chile (1960), Niigata, Japan (1964), Anchorage, Alaska (1964), San Fernando, California (1971), and Caucete, Argentina (1977). Even moderate levels of ground shaking may result in liquefaction and subsequent ground failure. *See* LANDSLIDE.

Earthquakes may also result in tsunamis, which are water waves with periods of about 5 to 60 min that are believed to be caused by underwater tectonic displacements. Tsunamis may also be caused by landslides and subaqueous slides that may or may



Acceleration hazard map of the United States. The numbers represent the horizontal acceleration in rock with a 90% probability of not being exceeded in 50 years. The acceleration is expressed as a percentage of the acceleration of gravity (≈10 m/s^2 or 32 ft/s^2). Therefore the numbers represent units of 10 cm/s^2 (4 in./s^2). Other numbers represent the highest acceleration attained within the contours. (*After S. T. Algermissen and D. M. Perkins, A Probabilistic Estimate of Maximum Acceleration in Rock in the Contiguous United States, U. S. Geol. Surv. Open-File Rep. 76–416, 1976*)

not have been triggered by an earthquake. Tsunamis have caused extensive property damage and life loss, particularly on Pacific islands and in coastal areas on the periphery of the Pacific Ocean, for example, those associated with the May 23, 1960, Chilean earthquake and the March 27, 1964, Alaskan earthquakes. *See* TSUNAMI.

Maps. Until about 1965, seismic hazards were estimated almost exclusively by use of deterministic methods. A number of seismic zoning maps have been prepared for the United States which provide estimates of earthquake effects. These have been compiled on the basis of a consideration of the regional geologic structure and its relationship to the size and spatial distribution of historic earthquake activity and of the resulting earthquake effects, described in degrees of intensity.

An intensity scale is a descriptive scale of earthquake effects arranged in degrees of increasing severity. The intensity scale most commonly used in the United States is the Modified Mercalli Intensity Scale (1931). This scale has 12 degrees, with I signifying earthquakes not felt or felt only rarely, under especially favorable circumstances, and XII signifying total destruction. Degrees I–V of the scale are based mostly on human reaction and perception of ground shaking; degrees VI–IX are based principally on the severity of damage to structures; and degrees X–XII are based mainly on geologic effects.

Deterministic hazard maps, which outline areas of potential fault rupture, landsliding, and liquefaction, have also been prepared.

Probabilistic models. Since 1965, interest in the probabilistic treatment of seismic hazard has increased. The most widely used (and perhaps the simplest) model assumes that earthquakes are exponentially distributed with regard to magnitude and that they follow a Poisson arrival process in time. The Poisson assumption is not in accord with concepts of the nature of earthquake occurrence, but it is suitable for the engineering evaluation of earthquake hazards. The occurrence of moderate to large earthquakes can closely approximate a Poisson process, and these earthquakes are the most important events in hazard and risk analyses. In addition, a suitable time-dependent model for large-earthquake occurrence is difficult to specify because of insufficient data. Bayesian statistics have been used with some success to model the earthquake process, but specification of prior probabilities will present difficulties until the earthquake mechanism is more clearly understood. A time history of ground motion resulting from an earthquake at any particular location of interest completely describes the ground-shaking hazard, but this

information is rarely available. Seismic intensity has been used frequently as a mapping parameter for seismic hazard, but because of its descriptive nature, it is somewhat unsuitable for formulating engineering specifications for the design of earthquake-resistant structures. The seismic ground-shaking hazard has been specified also by using peak ground acceleration, velocity, or displacement as a mapping parameter. Single-parameter descriptions of ground motion are not entirely satisfactory for the earthquake-resistant design of buildings, because the properties of ground motion that cause building damage cannot be described entirely by a single parameter. A probabilistic acceleration map of the United States is shown in the **illustration**. It has been used by the Applied Technology Council in the construction of two maps of the United States: one for effective peak acceleration and another for effective peak velocity. These two maps are used in constructing generalized spectra for ground motion on different kinds of superficial materials. These spectral shapes are considered to be the descriptions of ground motion for the determination of suitable earthquake-resistant designs for structures. The ground motions specified have an approximately 10% probability of being exceeded at least once in 50 years.

Risk studies. Seismic-risk analysis has been defined here as it involves probability, but risk studies can be deterministic or probabilistic, depending upon the requirements of the particular study. Practical techniques for probabilistic risk analysis in the earthquake-resistant design of structures have become highly developed. Seismic-risk estimates conducted on a regional or national basis for disaster preparedness and for economic studies have been largely deterministic in nature. Risk studies provide the framework for the evaluation of acceptable risks, those judged appropriate for determination of structure design or development of social or economic policy. Such analyses on a regional and national scale have yielded significant results. Studies have predicted that about 70% of the earthquake damage in the United States will occur in the western regions, even though this area contains only about 50% of the construction in the country.

While large earthquakes cause serious losses, the potential for cumulative losses from moderate earthquakes as least as great as those from individual large earthquakes exists because of their much more frequent occurrence. The Richter magnitudes for moderate earthquakes range from 5.5 to 6.5.

Earthquake losses in the United States are expected to become progressively larger with time because of the increase in population and the dispersion of the population. The difficulties in estimating seismic risk on a regional and national level are great because of the problem in estimating the construction characteristics, spatial distribution and value of structures, and their probable resistance to earthquake ground shaking. Seismic risk also is involved with the nature and importance of secondary economic losses resulting from an earthquake, such as those associated with suspension of business activities during building repair or restoration of utilities, increases in repair costs because of material shortages, and lack of skilled workers. The difficulty of evaluating total earthquake loss has prevented accurate and complete estimates of earthquake tolls. *See* EARTHQUAKE; RISK ASSESSMENT AND MANAGEMENT.

S. T. Algermissen

Bibliography. S. T. Algermissen and D. M. Perkins, *A Probabilistic Estimate of Maximum Acceleration in Rock in the Contiguous United States*, U.S. Geol. Surv. Open-File Rep. 76–416, 1976; S. Balassanian, A. Cisternas, and M. Melkumyan (eds.), *Earthquake Hazard and Seismic Risk Reduction: Advances in Natural and Technological Hazards Research*, 2000; C. A. Kircher and A. K. Chopra (eds.), *Seismic Engineering: Research and Practice*, 1989; F. Mulargia and R. J. Geller (eds.), *Earthquake Science and Seismic Risk Reduction,* NATO Science Series: IV: *Earth and Environmental Science*, 2004.

Seismic stratigraphy

Determination of the nature of sedimentary rocks and the fluids they contain from analysis of seismic data. Prior to the 1970s, seismic data were used only to map the structure of the Earth. Since then, improvements in seismic data quality have revealed reflection patterns that indicate the environment in which the sediments were deposited, and as a result the nature of the rocks. *See* SEQUENCE STRATIGRAPHY.

Further improvements in computers, data quality, data acquisition, and seismic theory have expanded the stratigraphic information that can be obtained from the patterns seen in seismic data. The better understanding of seismic stratigraphy has become important in both hydrocarbon exploration and production. *See* SEISMIC EXPLORATION FOR OIL AND GAS.

Sedimentation and seismic reflections. Seismic energy is partially reflected where changes occur in acoustic impedance, the product of density and velocity. At perpendicular incidence, reflectivity, R, is given by Eq. (1), where p and V are density and velocity, respectively, and the subscripts 1 and 2 indicate that these properties are measured above and below an interface where the values change. The fraction of the energy reflected is R^2. One expects density and velocity to change where rock type changes, and reflections to be associated with massive changes in rock type. However, stratification in sedimentary rocks is usually much smaller than a seismic wavelength, so that most reflections are the interference composites of many component reflections.

$$R = \frac{p_2V_2 - p_1V_1}{p_2V_2 + p_1V_1} \tag{1}$$

Deposition is usually episodic rather than smoothly continuous. On a macroscale, periods of land subsidence are interrupted by uplift events, or rivers shift course and begin depositing sediments in a new area, or similar events shift the locale of deposition. On a microscale, periods of quiet

water are interrupted by violent storms that rearrange sediments, or by floods which suddenly bring in large amounts of new material, or by similar events which rearrange sediments over a wide area. The microevents are associated with stratal surfaces; the macroevents often result in a geographic shift in the region of deposition, and are associated with unconformities. *See* ACOUSTIC IMPEDANCE.

Fairly large changes in density and velocity occur within the same rock type. Such changes are associated with changes in porosity or composition, such as changes in the sorting within a sand or changes in the type and degree of cementation. At least locally, such changes generally follow stratal surfaces which are relict depositional surfaces (isochronal surfaces); that is, they were the surface of the solid earth at some point in time. Stratal surfaces generally parallel each other; thus the component reflections do not change very much in short distances, do interfere in the same way on nearby seismic traces, and do produce coherent reflections. Consequently the reflections seen on seismic data generally show the attitude of isochronal surfaces.

The shifts in the geography of deposition associated with macroevents often result in the new depositional unit lying unconformably on older rocks. Usually the base of the unit is an unconformity associated with a hiatus in deposition, and also often is an erosional surface. The next macroevent which marks the end of the depositional unit may leave the top of the unit exposed to erosion and the next unit may be deposited elsewhere, so that the top is also apt to be an unconformity. The sediments above and below the unconformity are apt to differ, and erosion on the unconformities is apt to have changed the nature of the rock. Thus the acoustic impedance is apt to change significantly, and the unconformity will probably be a distinctive reflector. Often unconformities are responsible for the strongest reflections.

Thus seismic reflections can be associated with isochronal surfaces and unconformities (and fluid contact surfaces). The reflections associated with unconformities permit defining seismic-sequence units which are associated with depositional units, and those associated with isochronal surfaces permit determining the setting within the depositional unit, which may be indicative of the depositional environment. *See* SEDIMENTOLOGY; UNCONFORMITY.

Seismic-sequence analysis. The first step is to separate seismic-sequence units, also called seismic-facies units. This is usually done by mapping unconformities (**Fig. 1**) where they are shown by angularity. Angularity below an unconformity may be produced by erosion at an angle across the former bedding surfaces or by toplap (offlap), and angularity above an unconformity may be produced by onlap or downlap, the latter distinction being based on geometry. The unconformities are then followed along reflections from the points where they cannot be so identified, advantage being taken of the fact that the unconformity reflection is often relatively strong. The procedure often followed is to mark angularities in reflections by small arrows before drawing in the boundaries.

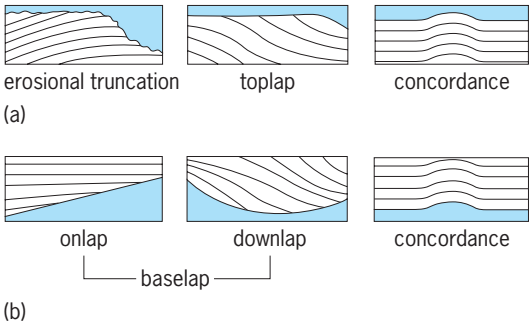


Fig. 1. Relation of strata to boundaries of depositional sequences. (*a*) At the upper boundary of a sequence, strata may terminate mainly as a result of erosion, toplap, or offlap (initially inclined strata may terminate mainly as a result of nondeposition) or top concordance (strata may be parallel to the boundary). (*b*) At the lower boundary, strata may terminate progressively against an initially inclined surface by onlap or downlap; strata may also parallel the lower boundary (base concordance). (*After C. E. Payton, ed., Seismic Stratigraphy: Applications to Hydrocarbon Exploration, Amer. Ass. Petrol. Geol. Mem. 26, 1977*)

Seismic-facies units are three-dimensional, and many of the conclusions from them are based on their three-dimensional shape. The appearance on seismic lines in the dip and strike directions is often very different. For example, a fan-shaped unit might show a progradational pattern in the dip direction and discontinuous, overlapping arcuate reflections in the strike direction (**Fig. 2**). *See* FACIES (GEOLOGY).

Reflection-character analysis. Often information from boreholes suggests that a particular interval may change nearby in a manner which increases its likelihood to contain hydrocarbon accumulations. Lateral changes in the wave shape of individual reflection events may suggest where the stratigraphic changes or hydrocarbon accumulations may be located. *See* DRILLING AND BORING, GEOTECHNICAL.

Where sufficient information is available to develop a reliable model, expected changes are postulated and their effects are calculated and compared to

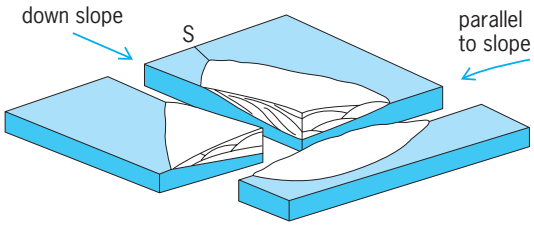


Fig. 2. Diagram of a fan-shaped depositional unit. The unit was fed from an entrenched source (S) and deposited on the continental slope or basin floor. The pattern in the downdip direction is a progradational one of successive units being deposited seaward on earlier units, whereas the pattern parallel to the slope shows the deposition jumping from one lobe to another to produce an overlapping cusp pattern.

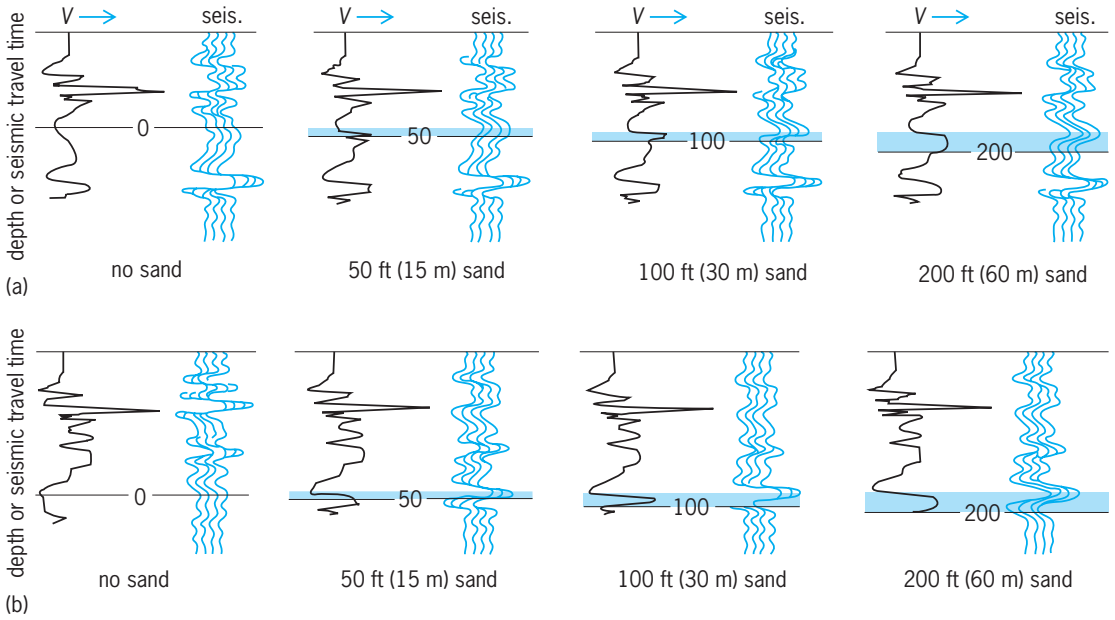


Fig. 3. Portions of sonic logs (V) and seismic traces (seis.) for four thicknesses of sand units; the sands are shaded. (*a*) Variable lower Wade. (*b*) Variable Medrano. The appearance of the sands produces a significant change in seismic wave shape, but the seismic wave shape does not change significantly as the sand thickens from 50 to 200 ft (15 to 60 m). (*After C. E. Payton, ed., Seismic Stratigraphy: Applications to Hydrocarbon Exploration, Amer. Ass. Petrol. Geol. Mem. 26, 1977*)

observed seismic data. The procedure is called synthetic seismogram manufacture; it usually involves calculating seismic data based on sonic and density logs from boreholes, sometimes based on a model derived in some other way. The sonic and density data are then changed in the manner expected for a postulated stratigraphic change, and if the synthetic seismogram matches the actual seismic data sufficiently well, it implies that the changes in earth layering are similar to those in the model. While the solution is nonunique, it gives an interpreter a feel for the relation between stratigraphic changes and their evidences in the seismic data, and is often very useful in actual practice. For example, interposing a possible reservoir sand into the middle of a shale (**Fig. 3**) may result in a significant reflection even where the sand thickness is only one-twentieth of a wavelength, but the reflection character may be insensitive to the thickness of the sand. As another example, the amplitude of a reflection may be markedly different where the pore spaces in the sand are filled with gas or liquid hydrocarbons rather than salt water, but only a minor amount of gas in salt water may produce equivalent effects.

Measurements made on seismic data such as amplitude, frequency, apparent polarity, and velocity are useful in suggesting where stratigraphic or structural features or hydrocarbon accumulations may be located. The general technique is to display attribute measurements, often as a color-coded overlay, along with the seismic section, and then examine the behavior of the attributes laterally along the reflection events. Constancy of attributes implies that there is no significant change in the rocks, a gradual change is generally interpreted as having a stratigraphic

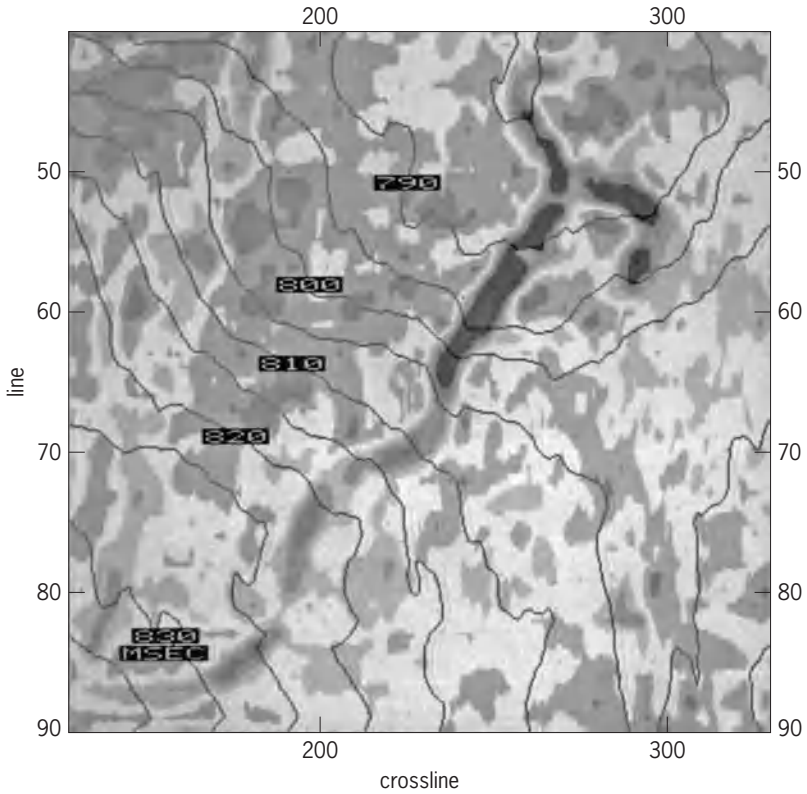
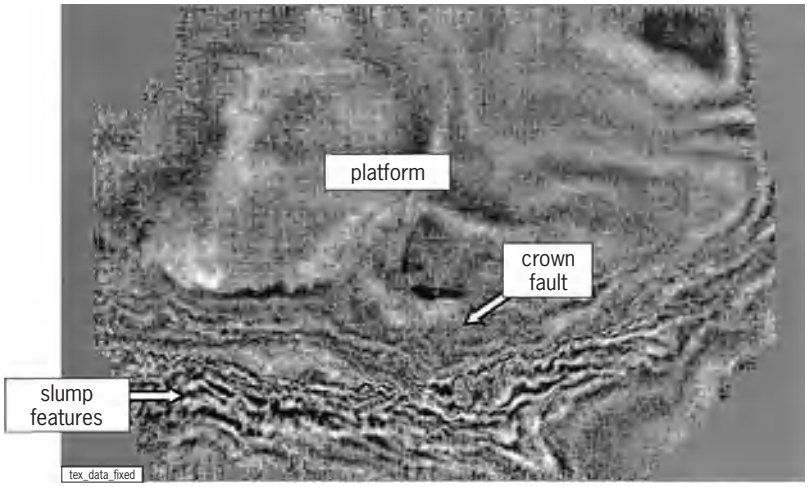
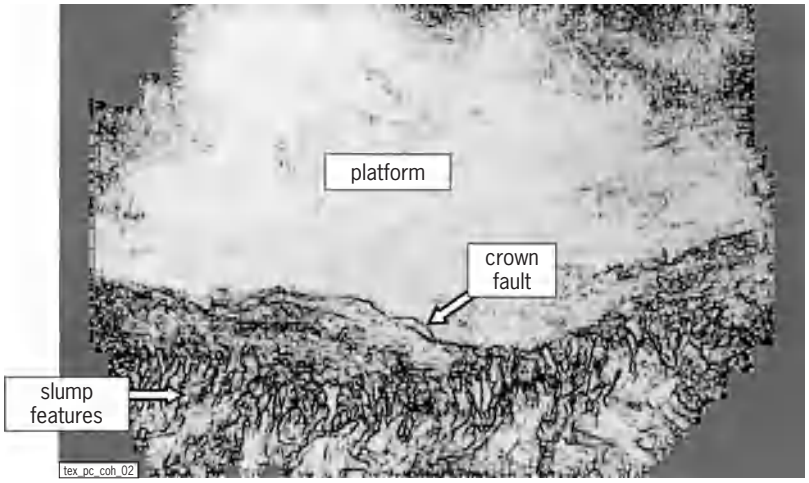


Fig. 4. Section through a 3D seismic volume in the Gulf of Mexico, showing a channel. The thin lines are arrival-time contours along a reflection event. They show the present-day relief on what was once the surface of the Earth. The different color intensities indicate different fluids filling the pore spaces. (*From A. Brown, 2004, courtesy of Chevron U.S.A. Inc.*)



(a)



(b)

Fig. 5. Horizontal time slice at 900 ms through a 3D seismic volume in the Delaware Basin of New Mexico. The northern portion cuts through a relatively featureless carbonate platform. The edge of the platform has begun to break and slump as indicated by the crown fault. The southern part of the survey shows a suite of slumps from higher above the platform. Eolean sands in the slumping formation make excellent but compartmentalized reservoirs. (*a*) Seismic data. (*b*) Coherence volume. Coherence is an attribute that emphasizes incoherencies in the seismic data.

cause such as change in bed thickness or facies, and a sudden change may indicate a fault or a hydrocarbon accumulation. Attribute measurements and display may aid in defining the limits of hydrocarbon accumulations so that field development may be planned in a more economical and efficient manner.

The reflection coefficient equation (2) can be

$$p_2V_2 = P_1V_1\frac{1+R}{1-R} \qquad (2)$$

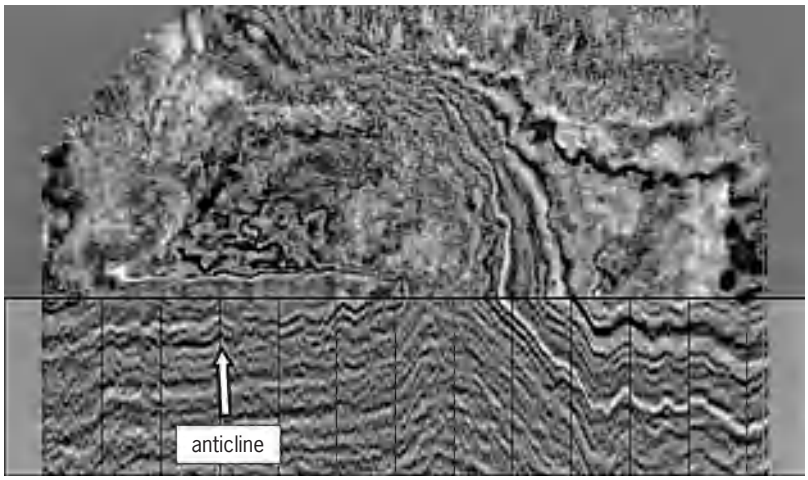
solved for the acoustic impedance. The acoustic impedance below an interface can thus be determined from that above the interface if noise can be removed from the seismic data sufficiently that the data represent the Earth's reflectivity. The result is a synthetic acoustic-impedance log (saile) or seismic log. The calculation process requires information not present in the seismic data, and additional velocity data usually provide the necessary supplemental information. For some purposes a seismic log display may be easier to interpret than the seismic data from which it is derived. The major problems with seismic logs usually involve the adequacy of isolating earth reflectivity effects from noise and equivalent wavelet shape, and there is controversy over their use. *See* GEOPHYSICAL EXPLORATION; SEISMOGRAPHIC INSTRUMENTATION; SEISMOLOGY; STRATIGRAPHY; WELL LOGGING.

Three-dimensional seismic methods. Whereas seismic stratigraphy was developed mainly by interpreting the vertical patterns that improved the vertical resolution revealed more clearly, the biggest single improvement in data quality has been in three-dimensional (3D) seismic methods that reveal horizontal (areal) patterns. The horizontal shape of features often reveals distinctive characteristics, such as the fan shape of delta lobes and the sinuosity that characterizes meandering streams (**Fig. 4**). Displays of quantities calculated from data measurements (attributes) often help us to see aspects that are otherwise difficult to see (**Figs. 5** and **6**). A multitude of different attributes are used to emphasize different properties. Having 3D data permits displaying attributes over slices through the data volume not only vertically and horizontally, but also along warped surfaces that follow seismic reflections. Some attributes require prior interpretation (such as picking of reflections, faults, and so on).

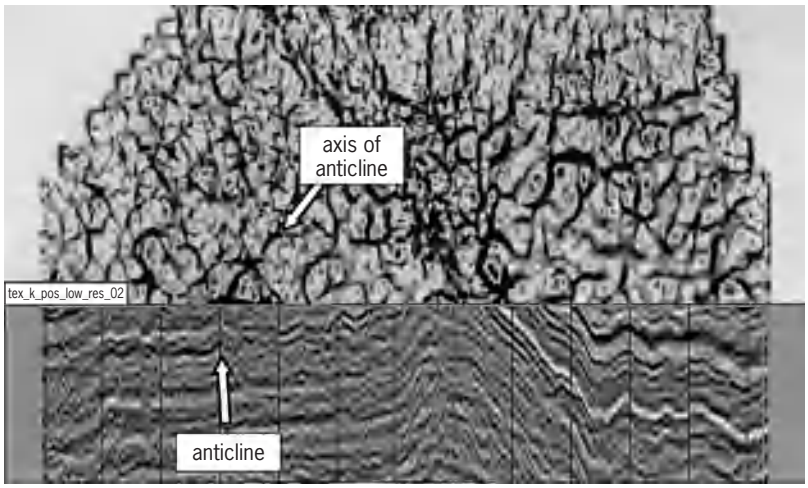
The nature of the fluid in a rock's pore spaces alters rock properties, especially velocity and density, which change the reflectivity at interfaces. Changes in reflectivity with the angle of incidence because of the conversion of pressure waves (P waves) into shear waves (S waves) are obscured by stacking. Prestack amplitude versus offset (AVO) analysis consists of examining the variation in amplitude with source-geophone distance (offset) and with incidence angle. Because the pore fluid affects P waves and S waves differently, AVO analysis often permits identifying the pore fluid, which has great economic value. Attribute displays that show where pore fluid changes occur help to locate where the hydrocarbons are and where stratigraphic features and faults affect fluid flow paths. Changes as hydrocarbons are produced (time-lapse or 4D seismic) show the movement of the fluids and have considerable economic value in managing production.

Multicomponent recording can provide S-wave data, in addition to P-wave data. Because P and S waves depend on different elastic properties, having both types of data permits a more definitive identification of the nature of the rocks and the fluids they contain. In areas where reservoir gas has permeated the overlying rocks so much that P-wave data is of very poor quality, the conversion of P waves to S waves upon reflection allows the mapping of the S waves, since they are relatively unaffected by the intruded gas. Multicomponent recording also helps determine the rock's anisotropic character, which helps reveal fractures that affect fluid flow. *See* PETROLEUM GEOLOGY; PETROLEUM RESERVOIR ENGINEERING.

Robert E. Sheriff



(a)



(b)

Fig. 6. Folded slice in the Delaware Basin, New Mexico. The northern part is a time slice at 1700 ms and the lower part is a vertical section at he southern edge of the time slice. (*a*) Seismic data. (*b*) Curvature attribute over the time-slice portion. Curvature is an attribute that shows complex folding. It is calculated for very small surfaces to fit the immediate surroundings of each point. Here it shows the orientation and extent of a small anticline. Curvature calculations do not require prior interpretation of the data.

Bibliography. A. R. Brown, *Interpretation of Three-Dimensional Seismic Data*, 6th ed., 2004; D. Emery and K. Myers (eds.), *Sequence Stratigraphy*, 1996; R. G. Loucks and J. F. Sarg, *Carbonate Sequence Stratigraphy*, 1993; R. E. Sheriff (ed.), *Reservoir Geophysics*, 1992; R. E. Sheriff and L. P. Geldart, *Exploration Seismology*, 2d ed., 1995; J. C. van Wagoner and G. T. Bertran (eds.), *Sequence Stratigraphy of Foreland Basin Deposits*, 1995; P. Weimer and H. W. Posamentier (eds.), *Siliciclastic Sequence Stratigraphy: Recent Developments and Applications*, 1994; P. Weimer and R. M. Slatt, *Petroleum Systems Of Deepwater Settings*, 2004.

Seismographic instrumentation

Various devices or systems of devices for measuring movement in the Earth. Ground motion is generally the result of passing seismic waves, gravitational tides, atmospheric processes, and tectonic processes. Seismographic instrumentation typically consists of a sensing element (seismometer), a signal-conditioning element or elements (galvanometer, mechanical or electronic amplifier, filters, analog-to-digital conversion circuitry, telemetry, and so on), and a recording element (analog visible or direct, frequency modulation, or digital magnetic tape or disk). Seismographs are used for earthquake studies, investigations of the Earth's gravity field, nuclear explosion monitoring, petroleum exploration, and industrial vibration measurement.

Design requirement. Seismographic instruments may be required to measure ground motions accurately over a range approaching 12 orders of magnitude, from as small as 10^{-11} m (the Earth's background noise level in the 2 mHz to 100 Hz band at very quiet sites) to as large as several meters (strong ground motion in the near-field of a very large earthquake). The instruments may be required to measure frequencies as low as $\sim 10^{-5}$ Hz (the semidiurnal gravitational tides), and even lower frequencies that are involved in tectonic strain monitoring, to as high as $\sim 10^{4}$ Hz (as observed from acoustic

emissions from rock failures in mines at distances of a few meters). Seismic waves from earthquakes are observed in the bandwidth of ~3 × 10^{-4} Hz (the gravest free oscillations of the Earth) to ~200 Hz (a local earthquake recorded by a seismometer installed in a borehole 100+ meters deep). In exploration seismology the frequency range of interest is typically 10–1000 Hz. No single instrument can operate over such a large dynamic range and frequency bandwidth. Thus a variety of instrumental designs are seen in seismometry.

Seismometers. The seismometer is the basic sensing element in seismographic instruments, and there are two fundamentally different types: inertial and strain. The inertial seismometer generates an output signal that is proportional to the relative motion between its frame (usually attached to the ground or a point of interest) and an internal inertial reference mass. The strain seismometer (or linear extensometer) generates an output that is proportional to the distance between two points.

In the inertial seismometer (**Fig. 1**), the inertial mass is an element of a damped mechanical oscillator, in either a spring-mass or a pendulum configuration. The restoring force may be either an elastic force (including the inertial elasticity of a piezoelectric device), an electrical spring, or gravity. Damping may be achieved by the use of viscous fluids or by electrical feedback to provide a force that resists the relative velocity of the suspended mass.

Regardless of the design details, the behavior of inertial seismometers is described by the basic second-order differential equation for the harmonic oscillator, where x is the relative motion between the

$$\frac{d^2x}{dt^2} + 2\zeta\omega_n\frac{dx}{dt} + \omega_n^2 x = \frac{d^2z}{dt^2}$$

mass and the frame, ζ is the damping factor ($\zeta = 1$ is critical damping), ω_n is the natural frequency (in radians per second), and z is the frame (ground) motion. The relative displacement, x, or the relative velocity, dx/dt, is sensed by an appropriate transducer. Displacement transducers are typically optical, variable-capacitance, or linear vector differential transformer (LVDT) devices. Velocity transducers invariably used coil-magnet systems until recently.

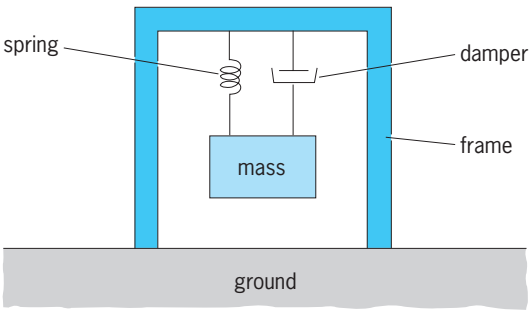


Fig. 1. Principle of the vertical-component inertial seismometer. Motion is measured between the mass and the frame.

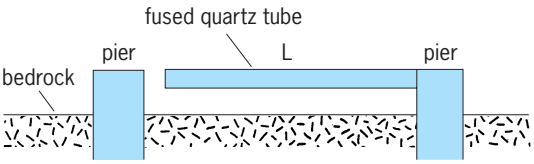


Fig. 2. Principle of the linear extensometer in a strain seismograph. Motion is measured between the left pier and the free end of the quartz tube.

However, a new type of velocity transducer has been developed which uses a molecular electronic transfer cell to detect the motion of an electrolytic fluid. In the case of piezoelectric (lead zirconate titanate, or PZT) seismometers, the voltage generated across the piezoelectric element is proportional to acceleration at frequencies below the resonant frequency (normally very high) of the piezoelectric device. *See* HARMONIC OSCILLATOR; PIEZOELECTRICITY; TRANSDUCER.

The output of a seismometer is a signal, usually electrical, that is proportional to some function of the ground motion, z. It is possible for this signal, over some frequency range, to be directly proportional to the ground displacement (z), the ground velocity (dz/dt), or the ground acceleration (d^2z/dt^2). Within such frequency ranges, seismographs are commonly referred to as displacement seismographs, velocity seismographs, or accelerographs, respectively. That is, the recorded amplitude represents the appropriate ground-motion quantity multiplied by a frequency-independent constant (the magnification, the velocity sensitivity, or the acceleration sensitivity of the seismograph) for motion in the particular frequency range.

A strain seismometer (or linear extensometer) [**Fig. 2**] is capable of providing a continuous precise measurement of the distance between two points, spaced typically 1 m to 1 km (3.3 ft to 0.6 mi) apart. A wide variety of designs have been implemented, using as reference lengths (L) from one point to the vicinity of the other point such standards as fused silica rods or light beams. Rod-based extensometers normally use a capacitance displacement transducer to sense variations in the position of the reference rod end with respect to the other reference point. Light-beam extensometers invariably incorporate some type of interferometer to detect changes in the length of the optical path between the two reference points. Longer paths require evacuated tubes and laser sources to achieve adequate sensitivity and stability in the length change measurement. The frequency range of interest is from zero to no more than ~1 Hz in most systems. *See* INTERFEROMETRY.

The quantity measured in these instruments is the change ΔL in the reference length L. The ratio $\Delta L/L$ gives the horizontal component (typically) of the strain between the two points. With a displacement resolution of 10^{-9} m (3.3×10^{-9} ft), a 100-m (330-ft) extensometer can measure changes in strain of 1 part in 10^{-11}. However, practical limits in environmental controls (temperature, atmospheric pressure, and

mechanical stability) yield long-term stabilities no better than 10^{-7} per year for such instruments.

Secular (long-term) strains are also measured using very long baseline interferometry (VLBI) and geodetic Global Positioning System (GPS) techniques. VLBI uses phase differences in signals from extraterrestrial radio sources (quasars) observed simultaneously at points separated by continental dimensions. Accuracies of a few centimeters in a few thousand kilometers are observed, for a resolution of 10^{-9} in relative distance change between the two remote points. The GPS system is used to observe geodetic strain by absolute position measurement on the Earth's surface. The geodetic deformation is monitored with both continuous (permanent) stations and campaign measurements (temporary deployments at existing benchmarks) over a period of a few years. Accuracies of a few millimeters are attained. *See* GEODESY; SATELLITE NAVIGATION SYSTEMS.

Seismoscopes. A seismoscope is a device that indicates only the occurrence of relatively strong ground shaking and not its time of occurrence or duration. A typical seismoscope inscribes a hodograph of horizontal strong ground motion on a smoked watch glass.

Dilatometers. A dilatometer continuously and precisely measures volumetric strain. The quantity measured is the change ΔV in the reference volume V, and the ratio $\Delta V/V$ gives the volumetric strain. Dilatometers are typically installed in boreholes in competent rock (preferably granite) at a depth of 100–300 m (330–1000 ft).

Tiltmeters. A tiltmeter monitors the relative change in the elevation between two points, usually with respect to a liquid-level surface. The horizontal distance between the reference points may be as little as a few millimeters or as large as several hundred meters. A displacement transducer is typically employed to sense the vertical separation between the two liquid-level surfaces. Environmental effects limit the long-term stability to about 10^{-6} radian.

Tilt at a point can be measured by an inertial seismometer. Tilt of the horizontal surface is indistinguishable inertially from a horizontal acceleration of magnitude $g \sin \theta$, where θ is the angle and g is the acceleration due to gravity. Any seismometer with DC (direct current or zero frequency) response to acceleration (for example, a horizontal pendulum equipped with a displacement transducer) is therefore a tiltmeter with constant tilt sensitivity throughout the frequency range where its output is proportional to acceleration. *See* ACCELEROMETER.

Gravimeters. The gravity meter is just a vertical-component accelerometer, that is, a pendulum sensing ground motion and equipped with a displacement transducer, analogous to the inertial tiltmeter. In its most widely used form, the pendulum is small for portability and ease of thermal stabilization, with a natural frequency of ~0.1 Hz. Transducer technology and environmental control methods allow the best gravimeters to have a repeatable accuracy of $\sim 10^{-8}$ g. Gravimeters are widely used in geophysical exploration, in the study of earth tides, and in the recording of very low frequency (0.0003–0.01 Hz) seismic waves from earthquakes. *See* EARTH TIDES; EARTHQUAKE; GEOPHYSICAL EXPLORATION; SEISMIC EXPLORATION FOR OIL AND GAS; SEISMOLOGY.

Recording systems. The complete seismograph produces a record of the properly conditioned signal from the seismometer, along with appropriate timing information. The recording system may be a simple as a mechanical stylus scratching a line on a smoke-covered drum in a portable microearthquake seismograph, or as complex as a multichannel computer-controlled system handling 25,000 24-bit digital words per second in a modern seismic reflection survey for petroleum exploration. The range between these extremes includes many special-purpose seismographs, all designed to record ground motion in a particular application.

Deployment. Many methods are used for deploying seismographs, and they depend upon whether the site is temporary or permanent and whether the site is land-based or on the ocean bottom. Seismic station installations are susceptible to several types of noise that contaminate or even mask the desired signals, effectively reducing the operating sensitivity of the instrument. These noise sources are of several types: atmospheric, cultural, ocean current, microseismic, and ambient temperature variation. The atmosphere affects the seismometer through pressure fluctuations directly on the mechanical system and indirectly on the surrounding ground. Cultural and industrial vibrations from human activities can be measured near any population center. In the case of seismographs deployed on the sea floor, ocean currents are a large noise source. Microseisms, the natural background vibrations of the Earth, are generally larger near continental margins than inland—a result of wave action on the coastline. For feedback broadband seismometers, ambient temperature variation is also a significant source of noise.

In the conventional land-based permanent seismographic station, the seismometer is placed upon a stable (usually concrete) pier in an environment as noise-free as possible. Two deployment methods have been devised to mitigate against noise sources. One technique uses seismometers in closely spaced arrays, relying on the coherency of the signal of interest over the array dimensions, and the lack of correlation of the noise components in the seismic wave field. This method is applied successfully in seismic reflection surveying, and in the arrays deployed to monitor from large distance the seismic waves from detonation of underground nuclear explosions. A second noise-reduction method uses borehole or deep mine installations to take advantage of the natural attenuation with depth of much of the noise field due to shallow propagating surface waves, to attenuate the effects of the surface ground motion due to atmospheric pressure fluctuation, and to provide temperature stability.

When deploying a temporary array of portable seismographic stations, as required for aftershock studies or research investigations, the siting criteria also include obtaining permission from the land

owners and concerns about vandalism. The portable seismographs are typically housed in small enclosures that are placed on the surface or buried in the ground a few feet. The stations are generally battery- and solar-powered and have sufficient storage capacity to run unattended for weeks to months. Siting the seismometer on hard rock and away from cultural noise sources is preferable but not always possible. Instrument pools of portable seismographs exist for a wide variety of seismic studies. The Program for Array Seismic Studies of the Continental Lithosphere (PASSCAL) instrumentation center, for example, maintains a large stockpile of portable seismographs that are used by researchers.

In the deployment of an ocean bottom seismograph (OBS), the logistics and installation problems are formidable. However, there is considerable incentive to place seismographs on the sea floor since it represents the largest uninstrumented area of the Earth's surface. A number of specialized methods have been devised for deploying, servicing, and retrieving OBS instrumentation. Borehole installations in the sea floor produce the lowest noise levels. Most OBSs use retrievable instrument capsules with onboard recording and sufficient battery capacity for unattended operation from a few weeks to a year or more. Some permanent OBS stations have been installed along cables that link them to shore and provide telemetry and power. Transoceanic telephone cables, decommissioned as a result of increasing reliance on satellite communications, are being considered for the installation of permanent OBS stations on the sea floor. OBS instrument pools exist for a wide variety of seismic studies. For example, both the Ocean Bottom Seismograph Facility of the Scripps Institution of Oceanography and the Woods Hole Oceanographic Institution Marine Seismology Group maintain a pool of OBS instrumentation for use by researchers. Thomas V. McEvilly; Robert A. Uhrhammer

Bibliography. M. E. Ander, T. Summers, and M. E. Gruchalla, LaCoste & Romberg gravity meter: System analysis and instrumental errors, *Geophysics*, 64:1708–1719, Society of Exploration Geophysicists, 1999; G. Anderson, H. Staudigel, and F. K. Wyatt, A seafloor long-baseline tiltmeter, *J. Geophys. Res. B*, 102:20269–20285, 1997; C. R. Bradley, *Very Low Frequency Seismo-acoustic Noise below the Sea Floor*, Ph.D. thesis, MIT/WHOI Joint Program in Oceanography, 1994; C. Braitenberg, The hydrologic induced strain-tilt signal, a review: Analysis of environmental data for the interpretation of gravity measurements, *Marees Terrestres: Bulletin d'Informations*, 131:10171–10181, Observatoire Royal de Belgique, 1999; W. K. Cloud, Instruments for earthquake investigation: Earthquake investigations in the western United States, 1931–64, *U.S. Coast Geod. Surv. Publ.*, 41-2:5–20, 1965; J. Dewey and P. Byerly, The early history of seismometry (to 1900), *Bull. Seismol. Soc. Amer.*, 59:183–227, 1969; J. Fowler, PASSCAL: A facility for portable seismological instrumentation, *EOS: Trans. Amer. Geophys. Union*, 75:66, 1994; B. S. Melton, Earthquake seismograph development: A modern history, parts 1 and 2, *EOS: Trans. Amer. Geophys. Union*, 62:505–510, 545–548, 1981; G. D. Myren and M. J. S. Johnston, Borehole Dilatometer Installation, Operation and Maintenance at Sites along the San Andreas Fault, California, *USGS Open-File Rep.*, no. 89-0349, 1989; R. A. Stephen et al., The Seafloor Borehole Array Seismic Systems (SEABASS) and VLF ambient noise, *Mar. Geophys. Res.*, 16:243–286, 1996; R. A. Uhrhammer, B. Romanowicz, and W. Karavas, Broadband seismic station installation guidelines, *Seismol. Res. Lett.*, 69:15–26, 1998; E. Wielandt, The leaf-spring seismometer: Design and performance, *Bull. Seismol. Soc. Amer.*, 72:2349–2367, 1982.

Seismology

The study of the shaking of the Earth's interior caused by natural or artificial sources. The revolution in understanding the Earth brought about by the paradigm of plate tectonics in the early 1960s provided a rationale that has since guided an intense period of investigation into the structure of the Earth's crust, mantle, and core. Plate tectonics provided Earth scientists with a kinematic theory for global tectonics that gave context to many of their studies. Ancient regions of crustal deformation could, for instance, be seen as the result of past plate boundary interactions, and the global distribution of earthquakes could be understood as the result of present plate interactions. The extraordinary insights into the nature of the solid Earth provided by kinematic plate tectonics also indicated new avenues in which to pursue basic questions concerning the dynamic processes shaping the Earth. Throughout the period in which plate tectonics was advanced and its basic tenets tested and confirmed, and into the latest phase of inquiry into basic processes, seismology (and particularly seismic imaging) has provided critical observational evidence upon which discoveries have been made and theory has been advanced. The use of the term image to describe the result of seismic investigations is closely linked mathematically to inversion and wavefield extrapolation and applies both to structures within the Earth and to the nature of seismic sources. *See* PLATE TECTONICS.

Theoretical seismology. A seismic source is an energy conversion process that over a short time (generally less than a minute and usually less than 1–10 s) transforms stored potential energy into elastic kinetic energy. This energy then propagates in the form of seismic waves through the Earth until it is converted into heat by internal (molecular) friction. Most earthquakes, for example, release stored elastic energy, nuclear and chemical explosions release the energy stored in nuclear and molecular bonds, and airguns release the pressure differential between a container of compressed air and surrounding water. Large sources, that is, sources that release large amounts of potential energy, can be detected worldwide. Earthquakes above Richter magnitude 5 and explosions above 50 kilotons or so are large enough to be observed globally before the seismic

waves dissipate below modern levels of detection. The largest earthquakes, such as the 1960 event in Chile with an estimated magnitude of 9.6, cause the Earth to reverberate for months afterward until the energy falls below the observation threshold. Small charges of dynamite or small earthquakes are detectable at a distance of a few tens to a few hundreds of kilometers, depending on the type of rock between the explosion and the detector. The smallest of these may be detectable for only a few seconds. *See* EARTHQUAKE.

Seismic vibrations are recorded by instruments known as seismometers that sense the change in the position of the ground (or water pressure) as seismic waves pass underneath. The record of ground motion as a function of time is a seismogram, which may be in either analog or digital form. Advances in computer technology have made analog recording virtually obsolete: most seismograms are recorded digitally, which makes quantitative analysis much more feasible.

Equation of motion. The dissipation of seismic energy in the Earth generally is small enough that the response of the Earth to a seismic disturbance can be approximated by the equation of motion for a disturbance in a perfectly elastic body. This equation holds regardless of the type of source, and is closely related to the acoustic-wave equation governing the propagation of sound in a fluid. Deviations of the elastic body away from equilibrium or rest are resisted by the internal strength of the body. An elastic material can be thought of as a collection of masses connected by springs, with the spring constants governing the response of the material to an externally applied force or stress. More formally, the strain or distortion of an elastic body when subjected to an applied stress is described with a constitutive relation, which is just a generalization of Hooke's law. Instead of just one spring constant, however, a generalized elastic solid needs 21 constants to completely specify its constitutive relation. Fortunately, the Earth is very nearly isotropic; that is, its response to an applied stress is independent of the direction of that stress. This reduces the complete specification of the constitutive relation from 21 to just 2 constants, the Lamé parameters λ and μ. A well-known result is that the equation of motion for an isotropic perfectly elastic solid separates into two equations describing the propagation of purely dilatational (volume changing, curl-free) and purely rotational (no volume changing, divergence-free) disturbances. These propagate with wave speeds α and β [Eqs. (1) and (2)], respectively,

$$\alpha = \sqrt{\lambda + \frac{2\mu}{\rho}} \tag{1}$$

$$\beta = \sqrt{\frac{\mu}{\rho}} \tag{2}$$

where ρ is the density and the Lamé parameters have units of pressure (force/area).

Elastic-wave propagation. These velocities are also known as the compressional or primary (P) and shear or secondary (S) velocities, and the corresponding waves are called P and S waves. The compressional velocity is always faster than the shear velocity; in fact, it can be shown that mechanical stability requires that $\alpha^2 \geq 4/3\beta^2$. In a fluid, $\mu = 0$ and there are no shear waves. The wave equation then reduces to the common acoustic case, and the compressional wave is just the ordinary sound or pressure wave. In the Earth, α can range from a few hundred meters per second in unconsolidated sediments to more than 13.7 km/s (8.2 mi/s) just above the core–mantle boundary. (The fastest spot within the Earth is not its center, but just above the core–mantle boundary in the silicate mantle. The core, being predominantly iron, is in fact relatively slow.) Wave speed β ranges from zero in fluids (ocean, fluid outer core) to about 7.3 km/s (4.4 mi/s) at the core–mantle boundary. *See* HOOKE'S LAW; SPECIAL FUNCTIONS.

A P wave has no curl and thus only causes the material to undergo a volume change with no other distortion. An S wave has no divergence, thus causing no volume change, but right angles embedded in the material are distorted. Explosions are relatively efficient generators of compressional disturbances, but earthquakes generate both compressional and shear waves. Compressional waves, by virtue of the mechanical stability condition, always arrive before shear waves.

Compressional and shear waves can exist in an elastic body irrespective of its boundaries. For this reason, seismic waves traveling with speed α or β are known as body waves. A third type of wave motion is produced if the elastic material is bounded by a free surface. The free-surface boundary conditions help trap energy near the surface, resulting in a boundary or surface wave. This in turn can be of two types. A Rayleigh wave combines both compressional and shear motion and requires only the presence of a boundary to exist. A Love wave is a pure-shear disturbance that can propagate only in the presence of a change in the elastic properties with depth from the free surface. Both are slower than body waves.

Solutions of the elastic-wave equation in which a wave function of a particular shape propagates with a particular speed are known as traveling waves. An important property of traveling waves is their causality; that is, the wave function has no amplitude before the first predicted arrival of energy. The complete seismic wavefield can be constructed by summing up every possible traveling wave, usually a countably infinite set. A traveling-wave formalism is useful especially when it is desirable to concentrate on particular arrivals of energy. An alternative set of solutions to the wave equation can be constructed by considering standing-wave functions. Individual standing waves by definition do not propagate, but rather oscillate throughout the body with a fixed spatial pattern. However, the sum of all possible standing waves must be equivalent to the sum of all possible traveling waves, because they must represent the same total wavefield. Standing waves, also known as free oscillations or normal modes, are useful representations of the wavefield in the frequency domain.

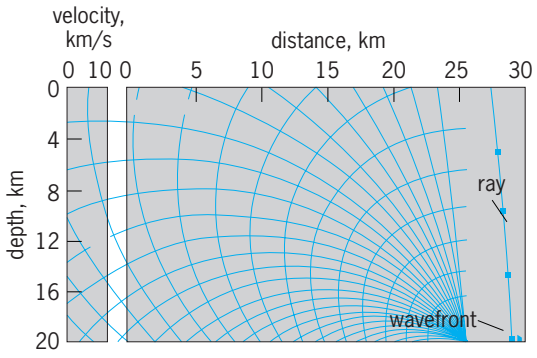


Fig. 1. Relationship between wavefronts and geometrical rays for a constant velocity gradient. 1 km = 0.6 mi.

Ray theory. Traveling-wave or full-wave theory provides the basis for a very useful theoretical abstraction of elastic-wave propagation in terms of the more common notions of wavefronts and their outwardly directed normals, called rays (**Fig. 1**). The validity of the ray treatment can be deduced from full-wave theory by taking infinite frequency approximations that treat the seismic disturbances as pure impulses having essentially no time duration and unit amplitude. Ray theory makes the prediction of certain kinematic quantities such as ray path, travel time, and distance by a simple geometric exercise involving what is essentially a generalization of Snell's law. Ray theory can be developed in the context of an Earth comprising flat-lying layers of uniform velocities; this is a very useful approximation for most problems in crustal seismology and can be extended to spherical geometry for global studies. Expressions will be developed for travel time and distance in both geometries.

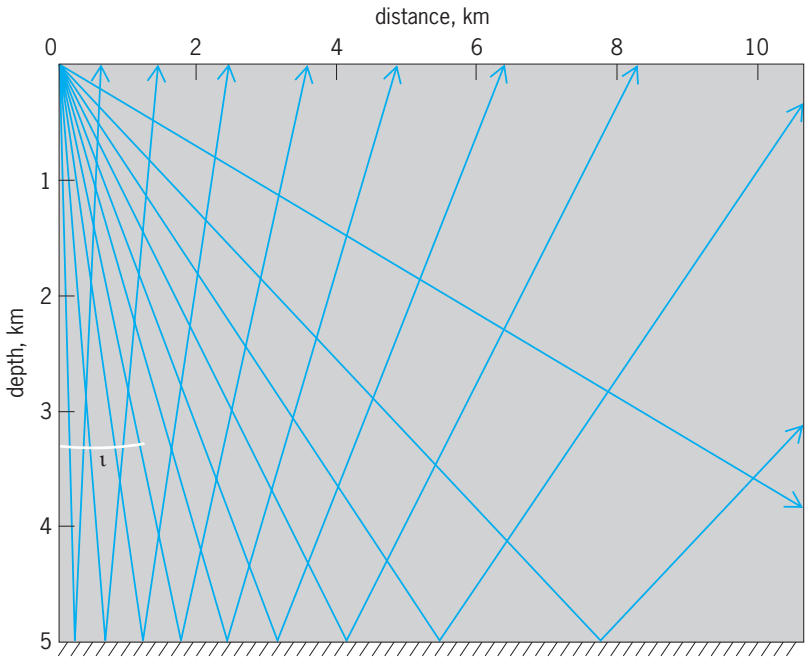


Fig. 2. Diagram showing seismic rays associated with a source at the Earth's surface bouncing off the base of a layer and being returned to the surface. The take-off angle ι for the ray that returns to the surface at a range of about 3.5 km (2.1 mi) is shown. 1 km = 0.6 mi.

Figure 2 is a diagram showing the rays of a seismic wavefront in a single layer expanding away from a source, a single layer. It is apparent from this simple ray diagram that the rays will intersect the base of the layer and be reflected to the surface. Each ray leaves the source at an angle ι, known as the take-off angle or angle of incidence. This angle can be used to characterize each of the rays by Eq. (3), where p is

$$p = \frac{\sin \iota}{\alpha} \tag{3}$$

the ray parameter. This term characterizes uniquely each ray in a system of rays. Its physical meaning is very important; if the term slowness is used to decribe the inverse of velocity, then p is the horizontal component of a ray's slowness. A vertical component q can also be defined, because unlike components of velocity the quantities p and q behave as vectors. This result makes it possible to cast the calculation of travel time and distance as geometric problems. Equation (4) gives the time $T(x)$ required to reach a given distance x from the source, and Eq. (5) gives

$$T^2(x) = T^2(0) + \left(\frac{x}{\alpha}\right)^2 \tag{4}$$

$$X(p) = \frac{2hp}{\sqrt{u^2 - p^2}} \tag{5}$$

the distance reached by a ray of given ray parameter, $X(p)$.

The first term on the right-hand side of Eq. (4) is the travel time of the vertical ray having a take-off angle of 0°, or $p = 0$. Equation (5) employs notation u to describe the slowness of the medium, rather than its velocity. Both Eqs. (4) and (5) are parametrized by layer thickness h and layer velocity α (or slowness u); and these are the unknowns needed to describe the Earth (α depends on the Lamé parameters and the density). Equation (4) is hyperbolic, which is sometimes observed in practice. This indicates that a simple representation of the Earth as a stack of uniform velocity layers might at times be reasonably close to reality.

Kinematic equations have been developed to describe what happens to rays as they impinge on the boundaries between layers. **Figure 3***a* is a diagram of a single ray propagating in the stack of horizontal layers that define the model Earth. At each interface, part of the ray's energy is reflected, but a portion also passes through into the layer below. The transmitted portion of the ray is refracted; that is, it changes the angle at which it is propagating. The relationship between the incident angle and the refracted angle is exactly the same as that describing the refraction of light between two media of differing refractive index, as shown in Eq. (6). This is the seismic analog

$$\frac{\sin \iota_{\text{incident}}}{\alpha_{\text{upper layer}}} = \frac{\sin \iota_{\text{refracted}}}{\alpha_{\text{lower layer}}} \tag{6}$$

of Snell's law. The left-hand side of Eq. (6) is the ray parameter for the ray in the upper layer, and the right-hand side is the ray parameter in the lower layer. Thus Snell's law for seismic energy is equivalent to stating

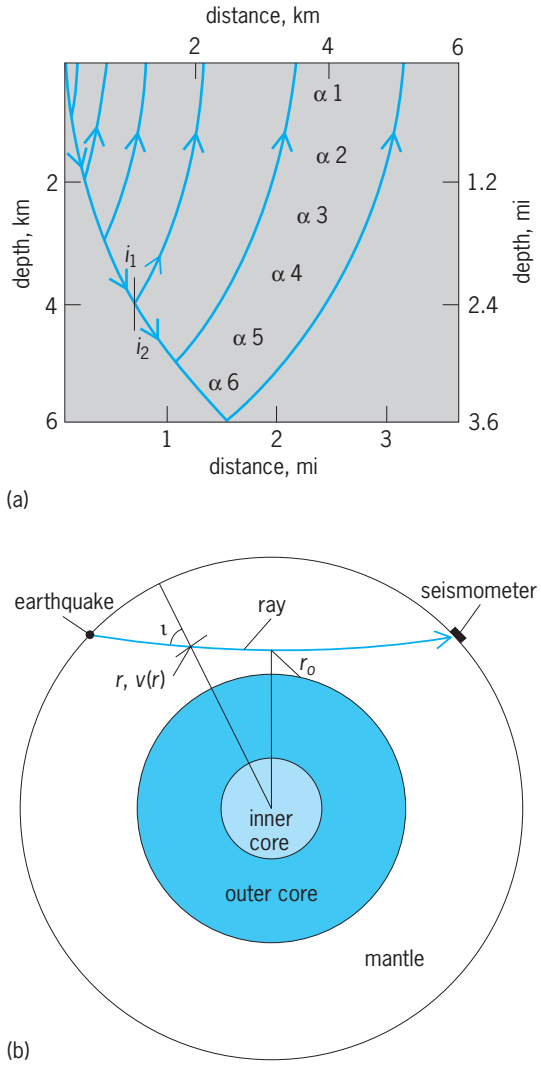


Fig. 3. Seismic ray paths. (*a*) A single ray passing through a multilayered Earth comprising a stack of uniform velocity layers will be reflected from each layer and also be refracted as it passes from one layer into the layer below in a manner that obeys Snell's law. Each ray therefore is considered to give rise to a new system of rays. (*b*) Ray diagram for a cross section of the spherical Earth. At the point labeled $v(r)$, r = radial distance and v = velocity. r_0 = radial distance to the turning point. ι is the angle of incidence.

that the ray parameter or the horizontal slowness is conserved on refraction. *See* REFRACTION OF WAVES.

Equation (7) for multilayers describes the distance

$$X_n(p) = 2\sum_{k=1}^{n} \frac{ph_k}{\sqrt{u_k^2 - p^2}} \tag{7}$$

reached by a ray of given p after passing through n layers, each of which is associated with a slowness u_k and thickness h_k before being reflected to the surface. It is the sum of the individual contributions to $X(p)$ from each of the layers. The equation for $T(x)$ cannot be written as a simple sum of hyperbolic contributions, although doing so in some special situations can be a reasonable approximation to the real travel-time behavior. Because of this, complex seismograms are generally analyzed by using a "parametric" form such as $T(x) = T[X(p)]$.

The generalization of these kinematic equations to the case where the medium slowness varies continuously with depth is straightforward. More care must be taken, however, in defining what happens to the ray at its deepest point of penetration into the medium. In Eq. (5), the ray is reflected from the base of the deepest layer. More generally, information is sought for the ray traveling downward, then reversing direction and traveling upward. In a medium with continuously varying velocities, this happens when the ray is instantaneously horizontal, that is, $\iota = 90^\circ$ and the ray parameter or horizontal slowness p is equal to the medium slowness $1/\alpha(z_{\max})$; the term $z_{\max}$ is defined to be the turning point of the ray and is the deepest point in the medium sampled by the ray. The integral form of the distance equation is given by Eq. (8), where $u(z) = 1/\alpha(z)$ represents any

$$X(p) = 2\int_0^{z_{\max}} \frac{p\,dz}{\sqrt{u^2(z) - p^2}} \tag{8}$$

relationship of slowness versus depth.

The case of a spherical Earth is treated by replacing the flat-Earth ray parameter by Eq. (9), where r

$$p = \frac{r\sin\iota}{v(r)} \tag{9}$$

is the distance from the center of the Earth, $v(r)$ is the velocity at that radius, and ι is the angle of intersection between the ray and the radius vector (Fig. 3*b*). The velocity function $v(r)$ denotes either $\alpha(r)$ or $\beta(r)$, depending on the circumstance. The radius r_0 at which $\iota = 90^\circ$ is known as the turning radius of the ray and denotes the point at which the ray begins its journey back toward the surface. If $v(r)$ is a constant independent of radius, the ray path is straight but the turning point is still well defined. In a spherical Earth, the ray travels in the plane containing the source, the receiver, and the center of the Earth. Like its flat-Earth counterpart, the spherical-ray parameter must be conserved along the ray.

Seismogram synthesis. These simple geometric arguments can be extended to the computation of amplitudes provided that there are no sharp discontinuities in the velocity as a function of depth. More exact representations of the amplitudes and wave shapes that solve the full-wave equation to varying extents can be constructed with the aid of powerful computers; these methods are collectively known as seismogram synthesis, and the seismograms thus computed are known as synthetics. Synthetics can be computed for elastic or dissipative media that vary in one, two, or three dimensions.

The fundamental rationale for computing synthetics is rooted in the notion of the seismogram as an observable entity, and the distributions of Lamé parameters and density within the Earth as unknowns. In other words, it is the province of seismological imaging to estimate the three-dimensional variation

of elasticity and density from observations of seismic-wave propagation. Beyond elucidation of basic structure, the variation of elasticity and density is a proxy for the variation of the chemical and thermal properties within the Earth. These in turn provide constraints on the long-term evolution of the Earth as a physical system.

Seismogram synthesis is an example of a forward problem; given a mathematical representation of the Earth and a model of the seismic source, attempts are made to compute synthetic seismograms (or some observables like travel time). For every forward problem, there is a concomitant inverse problem. The general statement of the inverse problem is: Given a set of observations of some measurable seismic disturbance at the Earth's surface, what can be said about that part of the Earth through which the disturbance passed and the characteristics of the source of the disturbance? The source and structural inverse problems are dichotomous in that something must be known about the source before the seismogram can be inverted for structure, and vice versa. This is not a problem in crustal or exploration imaging, where the properties of the artificial source are for the most part known and controlled. It is a problem when natural sources like earthquakes are used and specific steps must be taken either to isolate source and structure inversions or to do them jointly.

The development of inversion methods is a major area of research in seismology. The kinematic predictions of ray theory provide a useful starting point for a discussion of seismogram interpretation and inversion.

Crustal imaging. In a typical experiment for crustal imaging, a source of seismic energy is discharged on the surface, and instruments record the disturbance at numerous locations. Many different types of sources have been devised, from simple explosives to mechanical vibrators and devices known as airguns that discharge a "shot" of compressed air. The details of the source–receiver geometry vary with the type of experiment and its objective, but the work always involves collecting a large number of recordings at increasing distance from the source. **Figure 4***a* shows a seismogram known as a $T(x)$ resulting from one such experiment conducted at sea where an airgun source was fired and recordings were made by devices towed several meters below the sea surface from a research vessel. The strength of an arriving disturbance can be judged approximately by how dark the record appears.

This seismogram is complex, exhibiting a number of distinct arrivals with a variety of shapes and having amplitudes that change with distance, though they do not simply lose amplitude with increasing distance. Although this seismogram clearly does not resemble the structure of the Earth in any sensible way and is therefore not what would normally be thought of as an image, it can be analyzed to recover estimates of those physical properties of the Earth that govern seismic-wave propagation. Some simple deductions can be made by inspection. For instance, in a flat Earth the travel time equation is hyperbolic in distance. In the distance ranges up to about 15 km (9 mi), the observed $T(x)$ seismogram in Fig. 4*a* indeed exhibits a number of segments which have a roughly hyperbolic shape. This can be taken immediately to indicate that the Earth here probably contains at least a few simple layers of uniform velocity. Some segments, particularly those developed at ranges greater than about 20 km (12 mi), do not show a hyperbolic form. In fact, their curvature is in the opposite sense to that expected for a hyperbolic reflection form. This indicates that the arrivals have propagated through a structure in which there is a continuous increase in velocity with depth, and are therefore turning or diving rays as discussed above.

The formal procedure of travel-time inversion involves deriving the turning-point depth for each ray parameter by using a formula for $z_{\max}$ that is obtained from Eq. (5) by an integral transform known as the Abel transform. This approach has been used extensively by seismologists, often very successfully. Methods have been developed that allow observational uncertainties to be included in the procedure so that the inversion calculates all of the possible $v(z)$ functions allowed by the data and their associated uncertainties, thereby giving an envelope within which the real Earth must lie. The approach has several difficulties principally associated with the nonlinear form of the $X(p)$ equation. Although the desired quantity, $v(z)$, is single-valued (the Earth has only one value of velocity at any given depth), $X(p)$ can be multivalued. $X(p)$ also has singularities that are caused when the quantity under the square root goes to zero. Thus, care must be taken not to use rays that do not turn but are reflected at interfaces. In practice, separating out only those arrivals needed for the inversion can be very difficult. Furthermore, $X(p)$ data are not actually observed but must be derived from observations. Methodologies that can overcome these problems attempt to linearize the equation with respect to small variations in $X(p)$ and obtain an inversion that includes solution bounds derived from data uncertainties by linear programming.

Much of the difficulties have been further overcome by a coordinate transformation as shown in Eq. (10), where $\tau(p) = T - pX$.

$$\tau(p) = 2\int_0^{z_{\max}} dz\sqrt{u(z)^2 - p^2} \qquad (10)$$

The term $\tau(p)$ is known as the intercept time, and it can be thought of as the time that a vertically traveling ray takes to propagate upward from the turning point of a ray with ray parameter p. Although Eq. (10) still represents a nonlinear relationship between data $\tau(p)$ and desired Earth model $u(z)$, the square root no longer presents a problem, and this leads to considerable stabilization of the inversion. The observed seismogram can be easily transformed to $\tau(p)$ by a process known as slant stacking, in which a computer search is made for all arrivals having a particular horizontal slowness; those data are

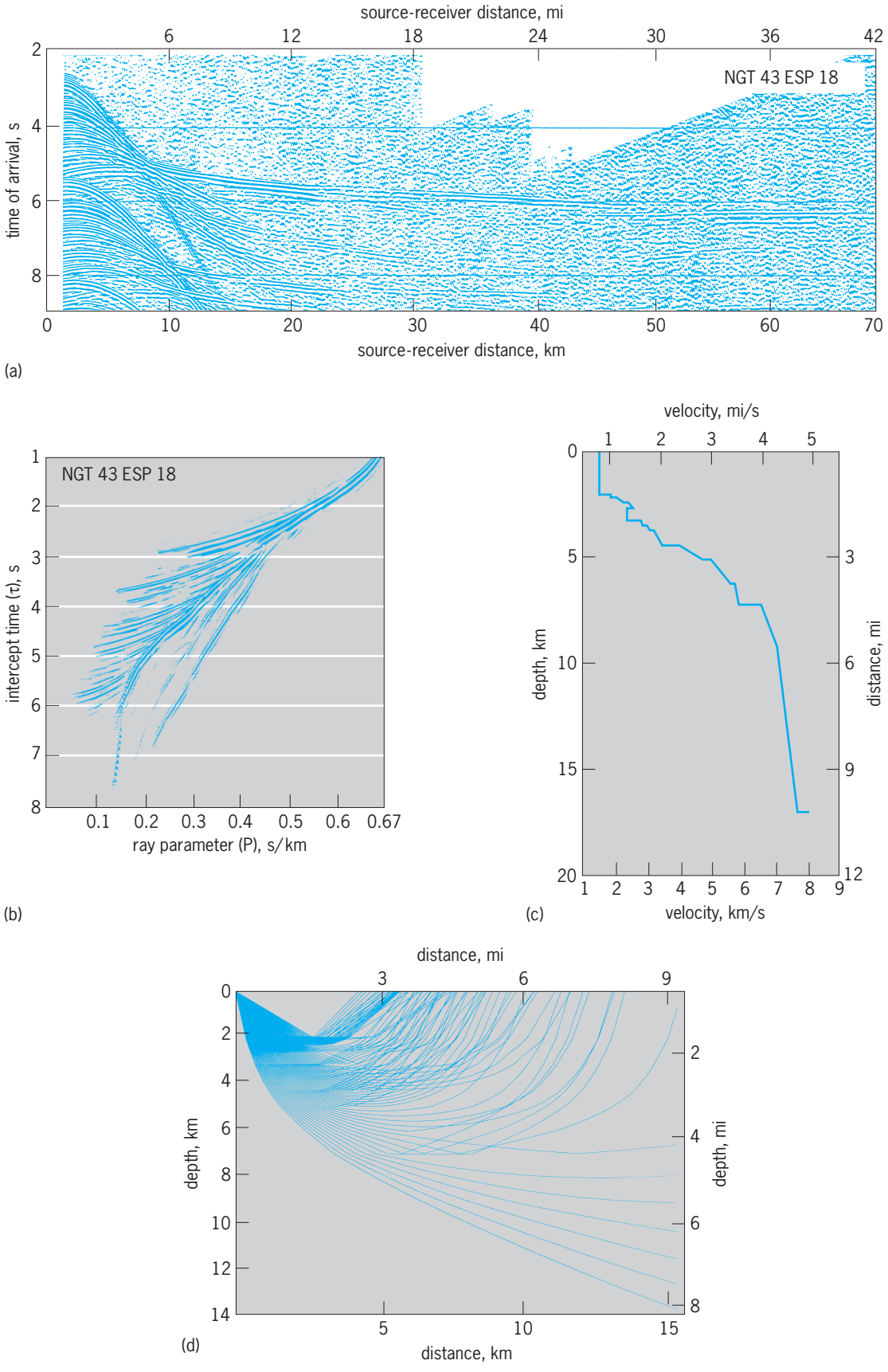


Fig. 4. Crustal imaging. (*a*) A *T*(*x*) seismogram recorded off the continental margin of East Greenland as part of a two-ship experiment designed to study the structure of the crust. (*b*) The *T*(*x*) seismogram data are transformed to $\tau(p)$ for analysis and for beginning the travel-time inversion. The part of the $\tau(p)$ data corresponding to the turning rays is organized into the strong band of energy in the seismogram on its lower right side. (*c*) By inversion and iterative forward modeling, an estimate of the change in compressional-wave velocity with depth is derived. (*d*) Rays that propagate through the structure derived from analysis of the seismogram take complex paths.

summed; and using the appropriate X and T data, $\tau(p)$ is computed (Fig. 4*b*). Most importantly, like the desired $\alpha(z)$ function these $\tau(p)$ data are single-valued, so that deriving the $\upsilon(z)$ function amounts to a point-by-point mapping from $\tau(p)$ to $\alpha(z)$.

Much of this theory has been known for quite some time, but the $\tau(p)$ approach to inversion and analysis of seismic data did not see extensive application until relatively recently. One reason is that the quality and sparseness of many early seismic experimental data did not allow them to be treated with $\tau(p)$ methods. Data often comprised just a dozen or so seismograms obtained by using explosive charges and a few recording stations, the records from which were made in analog form usually on paper records. The example seismogram (Fig. 4*a*) is typical of modern marine seismic data in comprising several hundred equally spaced traces. The seismic source consisted of an array of airguns. Seismic arrivals are recorded on another vessel towing a hydrophone array more than 2 km (1.2 mi) long containing several thousand individual hydrophones. Recording on this ship is synchronized to the shooting on the other vessel. Ship-to-ship ranges are measured electronically and written onto the same system that records the seismic data together with other information such as shot times. This experimental method leads to the dense sampling of the seismic wavefield that is

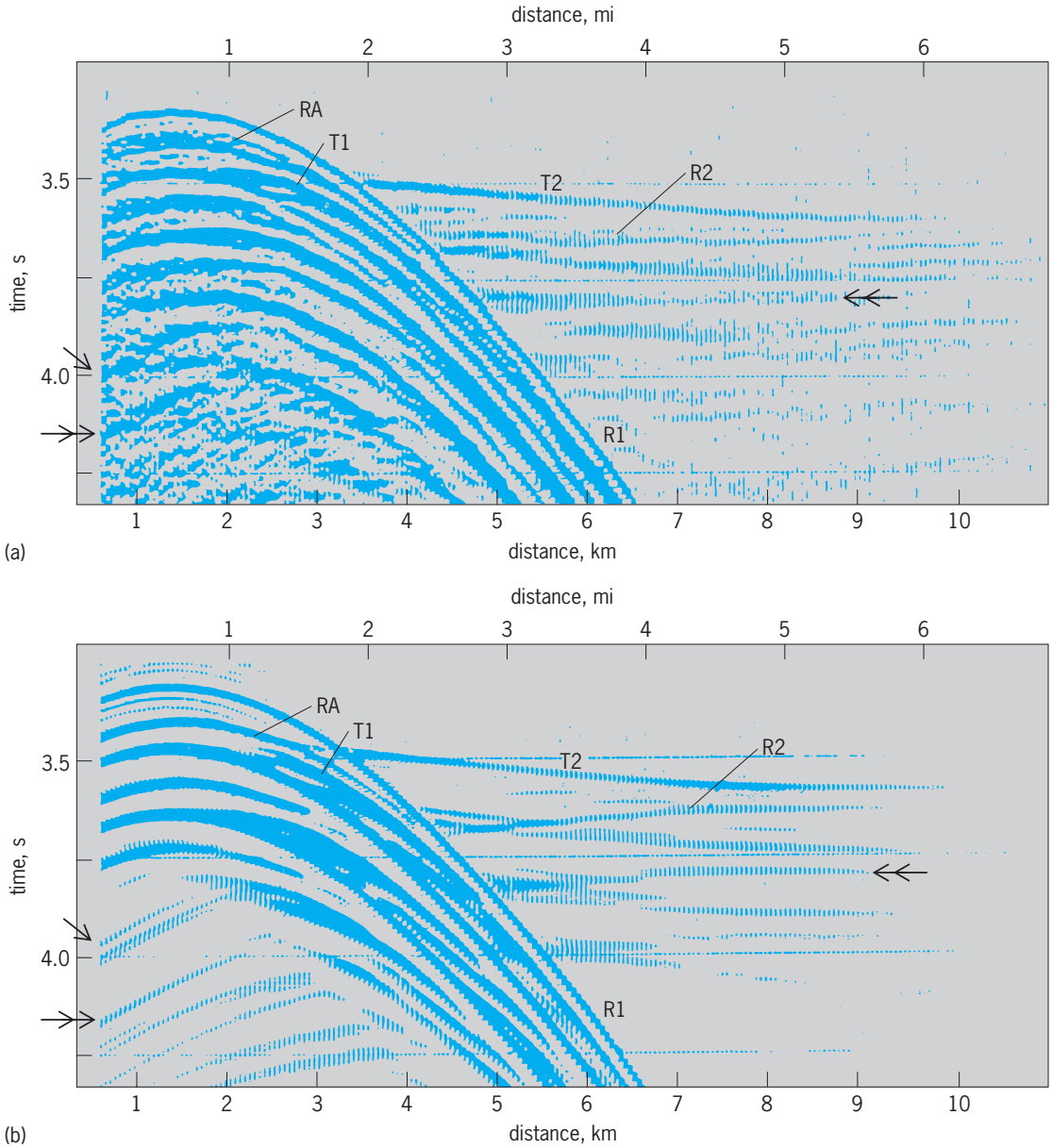


Fig. 5. Forward modeling; single and double arrows indicate the two types of reflected arrivals from the top of the axial magma chamber beneath the East Pacific Rise. (*a*) A *T*(*x*) seismogram obtained on the spreading center of the East Pacific Rise, used to obtain a velocity depth function by inversion and travel-time modeling. (*b*) A full synthetic seismogram, calculated and compared to the observed data as a way of resolving more detail in the structure than could be obtained from the travel times. In both parts, particular seismic phases [turning rays (T) and reflected rays (R)] are indicated; the vertical axis is a reduced time scale. (*From E. Vera et al., The structure of 0.0 to 0.2 m.y. old oceanic crust at 9°N on the East Pacific Rise from expanded spread profiles, J. Geophys. Res., 95:15529–15556, 1990*)

required for the correct computer transformation of the observed data into $\tau(p)$. The process is extremely demanding of computer time; it has come into common use only since relatively inexpensive, fast computers have made high computational power available to most researchers.

Figure 4*c* shows the result of analysis of the seismogram of Fig. 4*a*. It shows a profile of compressional-wave velocity against depth in the Earth at a location on the continental margin off East Greenland. The upper part of the crust where sedimentary layers are present is particularly complex, showing both uniform velocity layers, ones that show a gradational increase in velocity with depth, and one (at a depth of 3 km or 1.9 mi) in which the velocity is less than that of the layer above. This $\alpha(z)$ profile was constructed by the combination of several methodologies, including travel-time inversion from the $\tau(p)$ data in Fig. 4*b* and forward modeling of travel times and seismic amplitudes. Figure 4*d* depicts the rays that would propagate through such a structure; it is possible to recognize reflected arrivals that bounce off the interfaces between layers, and diving rays that smoothly turn in layers in which the velocity increases smoothly with depth.

Information from seismic waveforms. Real seismic disturbances have a finite time duration and a well-defined shape (see insert in Fig. 1). In passing through the Earth, any seismic disturbance changes shape in a variety of ways. The amount of energy reflected and transmitted at an interface depends on the ray incident angle (or equivalently on its ray parameter) and the ratio of physical properties across the boundary. It is sensitive both to the medium's compressional-wave and shear-wave velocities and to the ratio of densities. This is because, as well as splitting the arrival into reflected and refracted waves, boundaries act to convert waves from one mode of propagation to another. Thus, some of the compressional-wave energy incident on a boundary is always partitioned into shear energy.

Formulas for reflection and transmission coefficients are substantially more complex than those for the angles, and they are derived from considerations of traction across the boundary. At some incident angles the reflected energy becomes particularly strong. One point at which this happens is when the refracted wave travels horizontally below the interface, giving rise to a propagation mode known as a head wave, and essentially all the incident energy is reflected. This phenomenon is analogous to total internal reflection in optics. On the seismogram shown in Fig. 4*a*, there are several places where amplitudes become strong, and some can be related to critical-angle reflection, as this phenomenon is known in seismology.

Other strong regions correspond to places where energy is focused by the effect of changing subsurface gradients, which cause many diving rays to be returned to the surface near the same regions. Figure 4*d* was drawn by tracing a series of rays through the structure in Fig. 4*c*, with each ray incrementing in ray parameter by a fixed 0.005 s/km (0.008 s/mi). There are several regions where the density of rays returned to the surface is relatively sparse, while in other regions the density is very high (around 6 km or 3.7 mi, for instance). This pattern of amplitude variations can be used to refine models obtained by travel-time inversion and modeling to provide estimates of properties such as density, shear velocity, and attenuation (the amount of energy dissipated into heat by internal friction); however, it is often difficult to separate velocity and density uniquely, and estimates of acoustic impedance, the product of velocity and density, are more commonly derived.

Waveform information can be employed either by forward modeling or by using the $T(x)$ seismogram (Fig. 4*a*) directly in the inversion procedure. In the

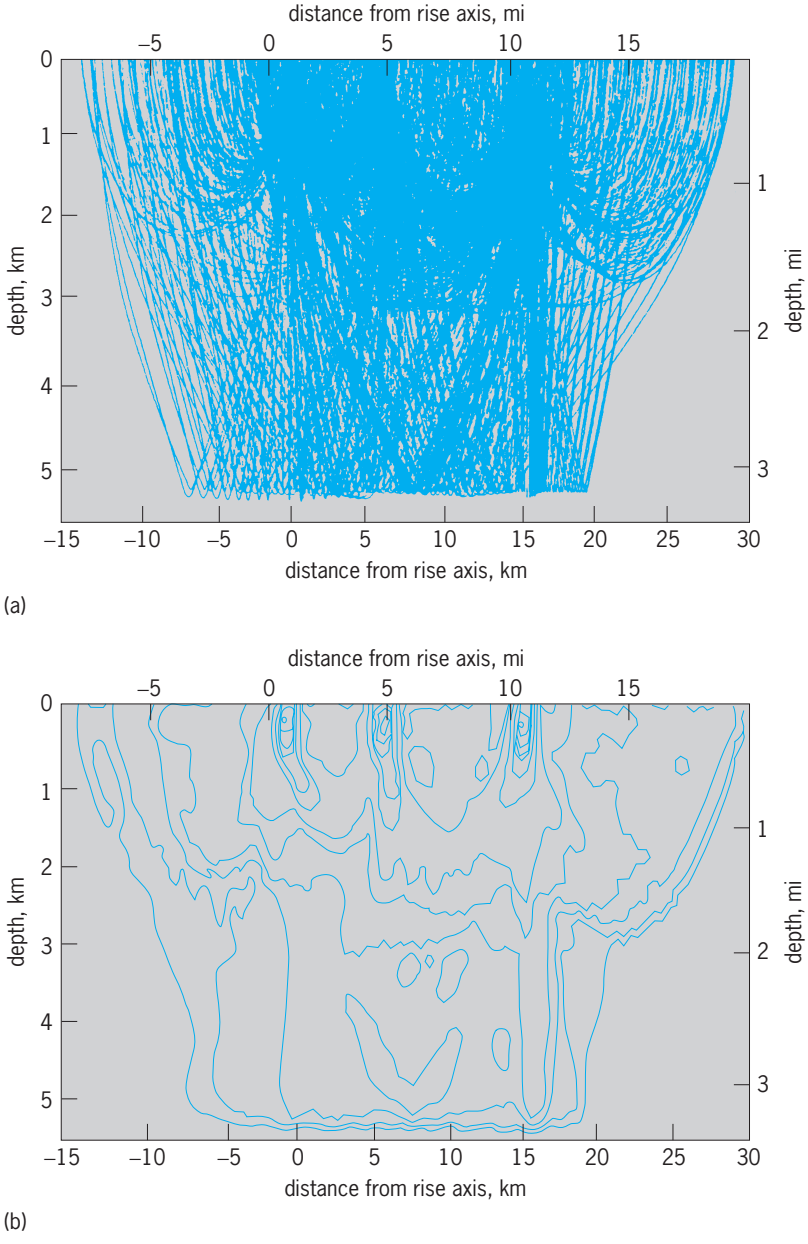


Fig. 6. Imaging by seismic tomography. (*a*) The type of ray pattern that can be used to conduct a tomographic inversion of travel-time data. (*b*) Travel-time partial derivatives that are closely related to the ray density and hence to the ability of the method to resolve structure.

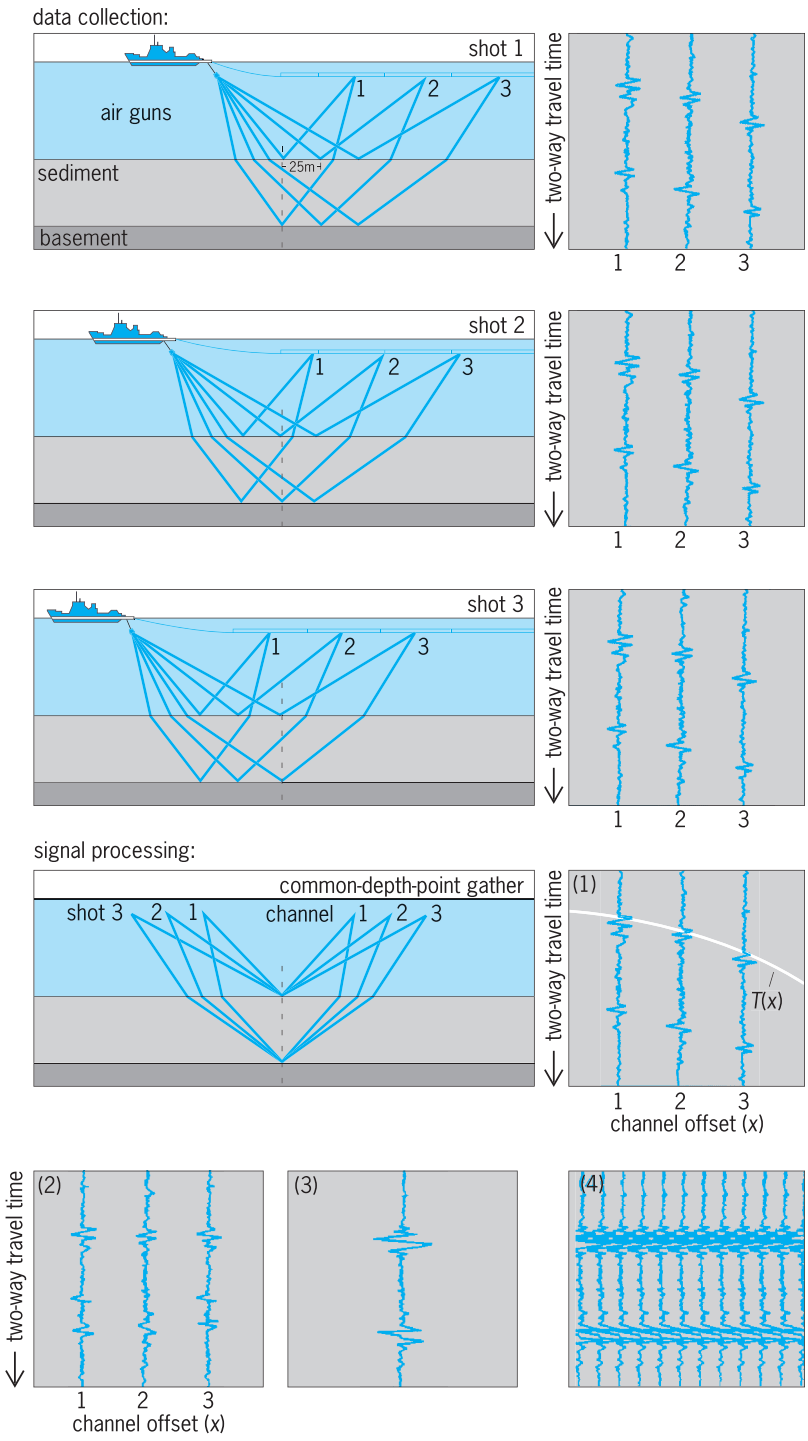


Fig. 7. Schematic representation of a standard multichannel seismic profiling experiment and the postcruise data processing. The first three rows show the ship towing an array of airguns and a recording streamer. The airguns fire at fixed intervals, sending acoustic energy (the heavy lines) into the water and the oceanic crust. The first (top) pulse represents the reflection of acoustic energy from the interface between the water and the sediment, and the second (bottom) pulse represents the reflection from the sediment–basement interface. The panels to the right show the *T*(*x*) seismograms recorded by each of the three hydrophones [channel offset (*x*) 1, 2, and 3] for shots 1, 2, and 3, with time increasing downward. The steps in the postcruise signal processing are denoted by (1)–(4) in the corner of the *T*(*x*) plots. Since the geometry of the shots and the recording streamer are maintained as the ship steams along, the *T*(*x*) seismograms of the three shots can be summed or "gathered" to reduce the effect of noise [plot (1)]. In (2), the common-depth-point gather is corrected for the effects of velocity moveout by fitting short sections of hyperbolas to the pulses. The results for each channel can then be gathered again to reduce noise further, as in (3). This is done continuously as the ship steams along, eventually producing an image of the interfaces, as in (4). (*After J. Mutter, Seismic images of plate boundaries, Sci. Amer., pp. 66–69, February 1986*)

forward methodology a synthetic seismogram is calculated on a computer; a model structure, usually derived from travel-time inversion, is used as a starting point, and the effect of propagation of a real seismic source through the structure is computed. The synthetic is then compared to the observed data, misfits noted, and adjustments made to the model until a satisfactory fit is obtained (**Fig. 5**). Computation of synthetic seismograms is enormously consuming of computer time, and the various methods in use all apply some computational shortcuts that make the problem tractable. Like inversion of travel-time data from $\tau(p)$, the advent of modern computers has made the computation of synthetics more realistic, though only supercomputers allow this to be done routinely.

These very fast machines may also allow waveform data to be incorporated more directly into the inversion of seismic data. Clearly the desired aim is to make use of all the information that the entire seismogram contains to yield estimates of all the physical properties that contribute to the observed waveform of the seismogram. The problem is a highly nonlinear one, and can be regarded as an optimization problem often addressed with a nonlinear least-squares procedure in which computed and observed data are systematically compared and adjusted until a fit is obtained that satisfies some optimization criterion. The computational demands of such a procedure are such that the use of supercomputers is essential. The automatic inversion of the complete seismogram to recover all the physical property information that affects the seismic waveform is one of the most challenging areas of research in crustal seismology.

Two- and three-dimensional imaging. A volume of the crust can be directly imaged by seismic tomography. In crustal tomography, active sources are used (explosives on land, airguns at sea) so that the source location and shape are already known. Experiments can be constructed in which sources and receivers are distributed in such a way that many rays pass through a particular volume and the tomographic inversion can produce relatively high-resolution images of velocity pertubations in the crust (**Fig. 6**). Crustal tomography uses transmitted rays like those that pass from a surface source through the crust to receivers that are also on the same surface.

Because most applications of tomography make use of rays that are refracted in their passage through a structure, they provide representations of the crust expressed in terms of smoothly varying contours of velocity, or velocity anomaly with respect to some reference. Many interfaces within the crust are associated with relatively small perturbations in velocity and occur on relatively small spatial scales, making their imaging by tomographic techniques essentially impossible. The finely layered strata of a sedimentary basin, for instance, cannot be imaged by such an approach.

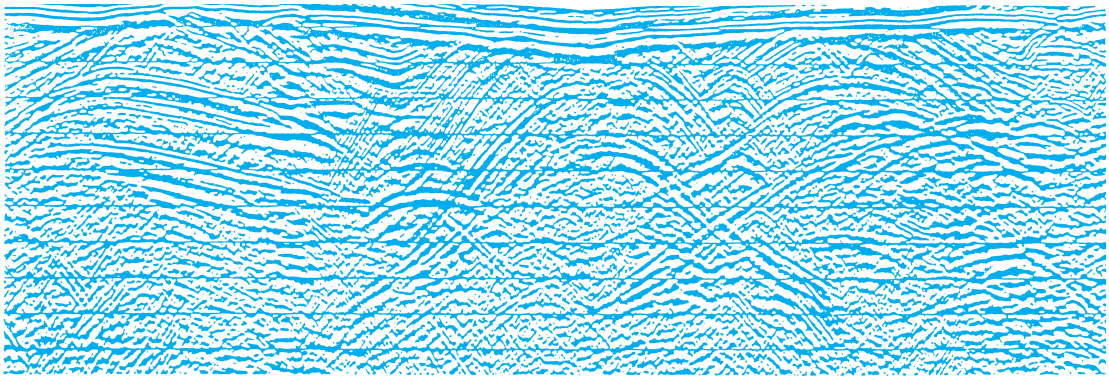
The most successful approach that has been devised to address the detailed imaging of crustal structure involves the use of reflected arrivals and is obtained by a profiling technique. In the reflection

profile technique, energy sources and hydrophone (or geophone) arrays are the same as those used to create seismograms used for travel-time inversion; but in the profiling experiment, both are towed from the same vessel or moved along the ground together.

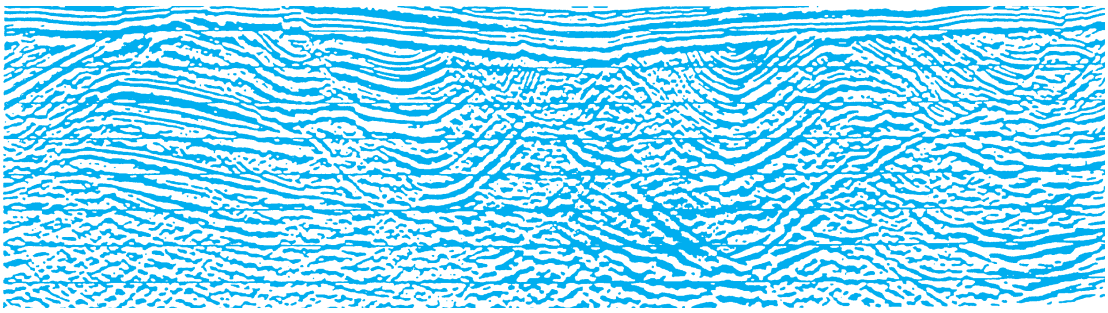
To produce an image from field recordings, they are first regrouped (gathered), corrected for the effect of varying source–receiver distance, and summed (stacked). The regrouping is made so that arrivals that are summed have come from the same reflection points and the time corrections are made assuming that all the arrivals obey hyperbolic travel times (**Fig. 7**) and Eq. (4). Although the latter assumption is not completely correct, it is an acceptable approximation for the relatively small source–receiver offsets (the distance between the shot position and the receiver) used in this type of experiment. After summation, the resultant seismogram has very high signal-to-noise characteristics and appears as if it were obtained in an experiment in which source and receiver were fixed at zero separation and moved along the profile line. Such a record is shown in **Fig. 8**.

Mathematically, if a recording of the wavefield at the surface has been obtained, and since it is known that propagation obeys the wave equation, then the equation can be used to move the wavefield back to its point of origin; that is, it can be extrapolated back down into the Earth to the place where it began to propagate upward. Having done this, and applying some condition that allows the strength of the reflection to be determined, it becomes possible to recover an essentially undistorted image of the structure. A conceptual similarity to travel-time inversion can be recognized in this methodology. In travel-time inversion, formulas are used for $X(p)$ or $\tau(p)$ to determine the turning-point depth of a ray. This could be restated by saying that the calculation extrapolates back down the ray from the surface to its turning point. Reflection imaging uses the wave equation to extrapolate the entire reflected wavefield back to its points of reflection, and can be considered the inversion methodology appropriate for the reflected wavefield.

Several wavefield extrapolation methods have been developed, all of which involve several stages of manipulation of the data, and all are very demanding of computer time. Figure 8 shows one example in which data obtained in shallow water in a sedimentary basin showing complex deformation structures have been imaged to reveal a variety of geologically interpretable structures. The velocity field in the structure must be known very well (and may not be) for the imaging to be successful. Most extrapolation methods involve some form of approximation to the full-wave equation to make the computations tractable. Most, for instance, allow only mild lateral variations in velocity. To do the job in a complete sense requires use of the full-wave equation and operating on prestack data.



(a)



(b)

Fig. 8. Seismic profiling. (*a*) Stacked reflection profile obtained by using the common depth point (CDP) profiling method described in Fig. 7. The record represents the acoustic response of the Earth to near-vertical-incidence seismic energy along a traverse. It is analogous to the seismogram in Fig. 4*a*, which represents the response of the Earth at one location to seismic energy at a wide range of angles. (*b*) Reflection seismic image produced by migrating the data from *a*; migration of reflection seismograms can be treated as the equivalent of inversion of refraction seismograms.

Advances, typically led by industry requirements, have seen surveys conducted to provide three-dimensional images, usually for petroleum reservoir evaluation. Survey lines are conducted on orthogonal grids in which the line spacing is as little as 50 m (160 ft) in both directions. Large exploration vessels carry two long hydrophone arrays set apart by 50 m (160 ft) on large booms or deployed on paravanes. Correct imaging of structure from such recordings requires a fully three-dimensional wavefield extrapolation procedure, and this has been developed successfully by using the power of supercomputers. Once the migrated image has been made, it can be manipulated in a smaller computer to observe structure in ways that cannot be achieved by conventional means. Horizontal slices can be made to show buried surfaces, diagonal cuts can be made to examine fault structure, and the whole image can be rotated to view the volume from any chosen direction.

The correct three-dimensional imaging of the Earth, either the whole Earth by tomographic means or crustal structure by tomography and reflection seismic imaging, represents a frontier research area in seismology. *See* GEOPHYSICAL EXPLORATION; GROUP VELOCITY; PHASE VELOCITY; SEISMIC EXPLORATION FOR OIL AND GAS; WAVE EQUATION.

Discoveries. Although the theory of plate tectonics allowed Earth scientists to recognize that the global mid-ocean ridge system represents the locus of formation of oceanic lithosphere, it gave little direct insight into the processes operating at these spreading centers. At a few scattered locations throughout the world, small slivers of oceanic crust known as ophiolites have been thrust into exposure onto continents as a result of collisional tectonics. Their study by geologists led to the proposition that large, shallow, relatively steady-state magma chambers are responsible for producing the crust at all but the slowest-spreading ridges. Efforts directed toward seismic imaging of spreading centers since 1980 produced some unexpected results. *See* MID-OCEANIC RIDGE; OPHIOLITE.

Reflection imaging of the fast-spreading center at the East Pacific Rise combined with two ship refraction seismic measurements in 1985, and later tomographic imaging, showed that a magma body indeed exists beneath the axis but that it is small and discontinuous (**Fig. 9**). The magma body is so small, in fact, that the term chamber seems barely applicable. It is typically not more than 2–3 km (1.2–1.8 mi) wide, and the inferred region of melt (judged from the thickness of layers determined from the results of inversion and modeling of refraction seismic

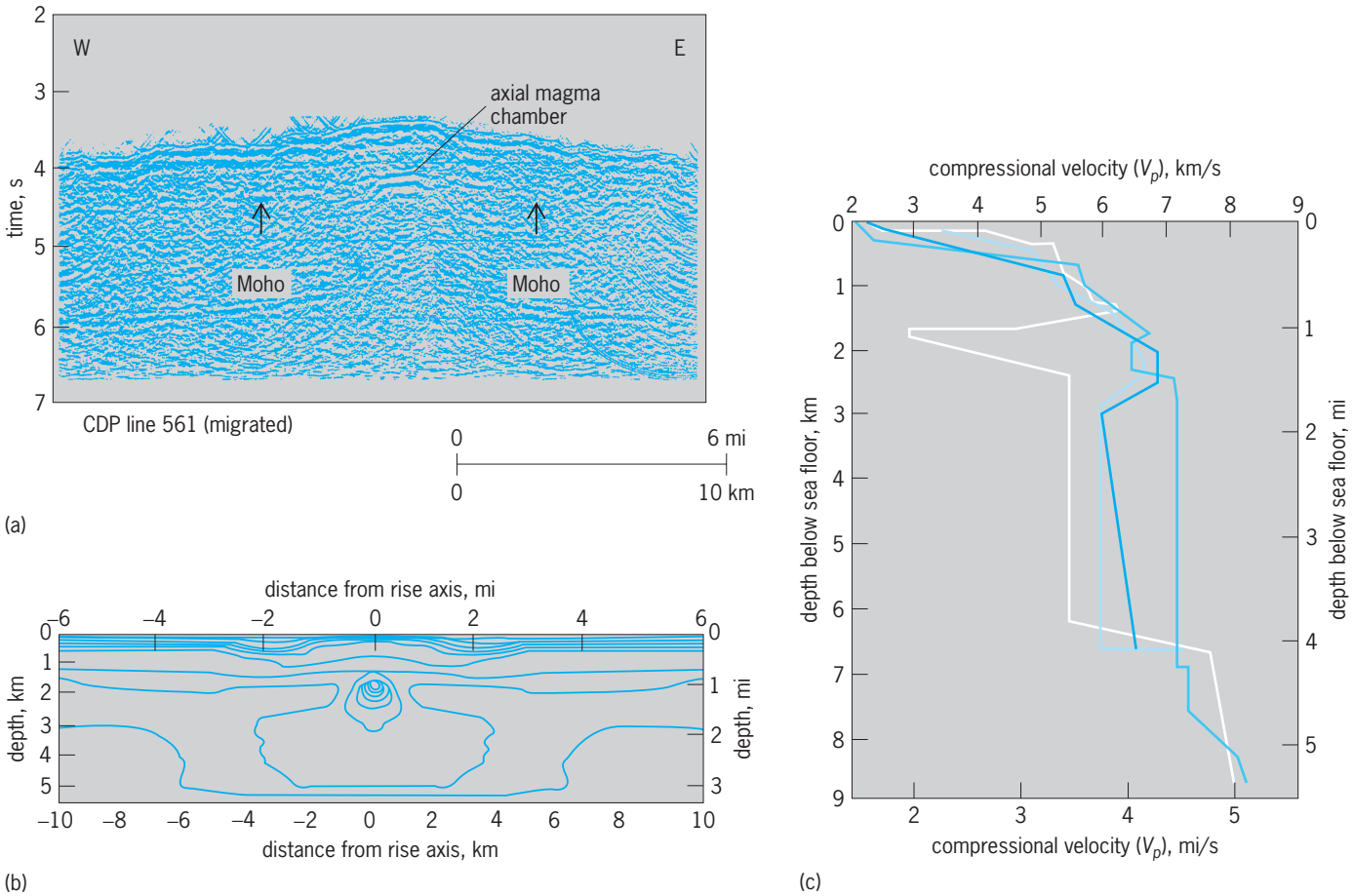


Fig. 9. Results of several different seismic investigations of the spreading center of the East Pacific Rise. (*a*) Reflection seismic image. (*b*) Velocity structure obtained by tomographic inversion. (*c*) Compilation of several velocity depth profiles obtained from inversion, travel-time, and synthetic seismogram modeling of refraction seismic data.

data) with greatly reduced velocities may be only a few hundred meters. This region is embedded in a broader zone of reduced velocities that may represent a region containing traces of melt but is more likely to be solid rock in which the temperature has been raised by proximity to the region of melt. The base of the crust, the Mohorovičić discontinuity (Moho), is a strong reflector formed very near the rise axis. *See* MOHO (MOHOROVIČIĆ DISCONTINUITY).

Images of continental crust and its underlying mantle have proven to be equally provocative (**Fig. 10**). Reflection profiling conducted on land in many regions of the world and on the shallow continental seas around Britain have shown that the deep crust is often characterized by an unreflective upper region and a band of strongly laminated reflections forming the lower third or so of the crust above a Moho that is quite variable in nature. The best available images show distinct events from within the mantle section itself. Mantle reflections may result from deep shear zones with reflectivity enhanced by deeply penetrating fluids. As exploration of the upper mantle has continued, it has become clear that structure in the mantle can be imaged by reflection methods to very great depths—perhaps even to the base of the lithosphere. It is possible that reflection methods will eventually be used to investigate the structure of the lithosphere as a whole, thereby complementing studies based on a different class of seismic methods that already have been developed. *See* LITHOSPHERE.

Global and regional seismograms. The basic unit of observation in global and regional seismology is a seismogram, but unlike their counterparts in crustal reflection and refraction, most seismometers used for larger-scale structural studies are geographically isolated from their neighbors. Thus the observational techniques and most methods of analysis used in global seismology have evolved quite differently from crustal seismology. Generally speaking, much more of each seismogram must be retained for analysis. This so-called spatial resolution gap is being closed slowly by the development of new portable instrumentation suitable for recording natural sources. Once these instruments are available in quantity, seismologists will be able to record seismic energy at closely spaced sites that will illuminate in much finer detail structures deep within the Earth.

Seismograms are classified according to the distance of the seismometer from the source epicenter. Those recorded within about 50–100 km (30–60 mi) of a large source are generally complex not only because of intervening structure but because the "dimensions" of the source are close to the propagation distance and different areas of stress release on the fault write essentially different seismograms. The term near-field is given to these seismograms to signify that the propagation distance for the seismic energy is less than a few source dimensions. Understandably, these seismograms are most useful for examining the details of the earthquake rupture. Beyond the near field to distances just past 1000 km (620 mi), the seismograms are dominated by energy propagating in the crust and uppermost mantle. These so-called regional seismograms are complicated, because the crust is an efficient propagator of high-frequency energy that is easily scattered; but there are still discernible arrivals. These seismograms are used to examine the velocity structure and other characteristics of relatively large blocks of crust. The domain beyond 1000 km (620 mi) is called teleseismic. Seismograms written at teleseismic distances are characterized by discrete and easily recognized body phases and surface-wave arrivals. These are relatively uncontaminated by crustal structure, and instead they are more sensitive to structure in the mantle and core.

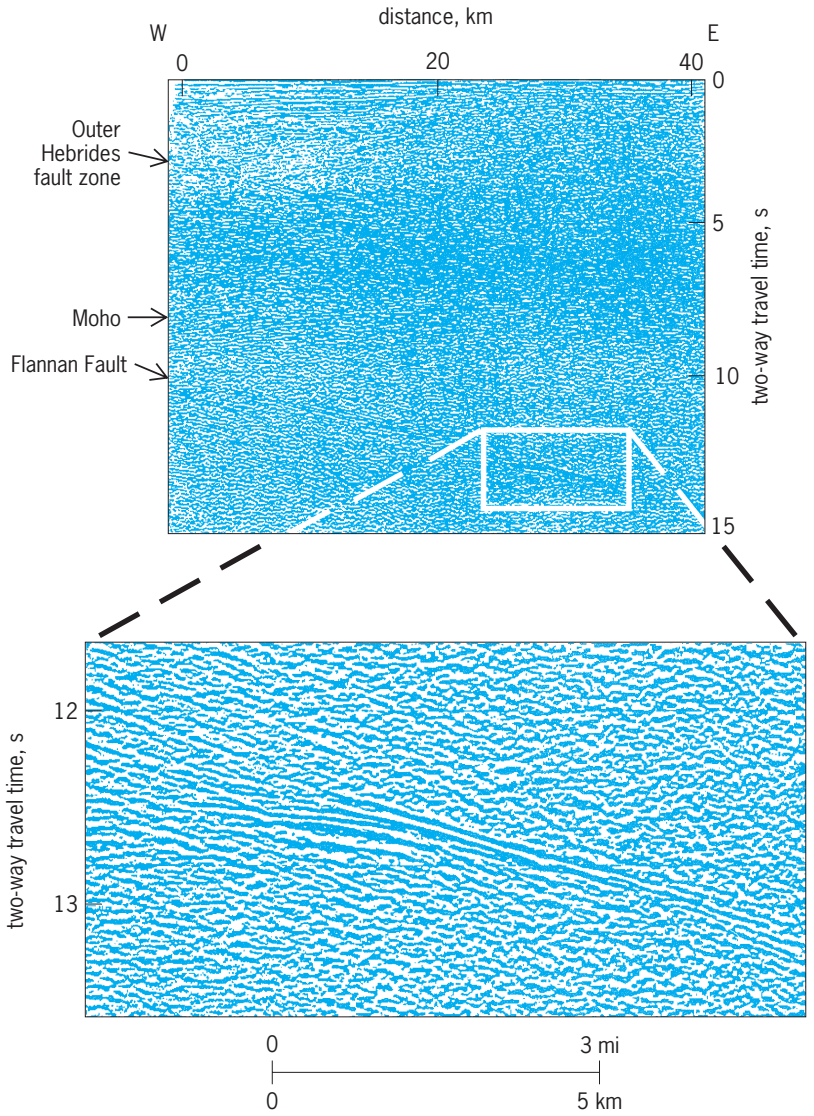


Fig. 10. Reflection seismic images of deep continental crust offshore Britain showing a variety of reflecting interfaces in and beneath the crust. (*From D. H. Mathews and M. J. Cheadle, Deep reflections from the Caledonides and Variscides west of Britain and comparison with the Himalayas, in M. Banazang and L. Brown, eds., Reflection Seismology: A Global Perspective, Geodynamics Ser., vol. 13, 1986*)

Figure 11 shows both regional and teleseismic seismograms; since seismic-wave motion is vector motion, three components (up-down, north-south, and east-west) are needed to completely record the

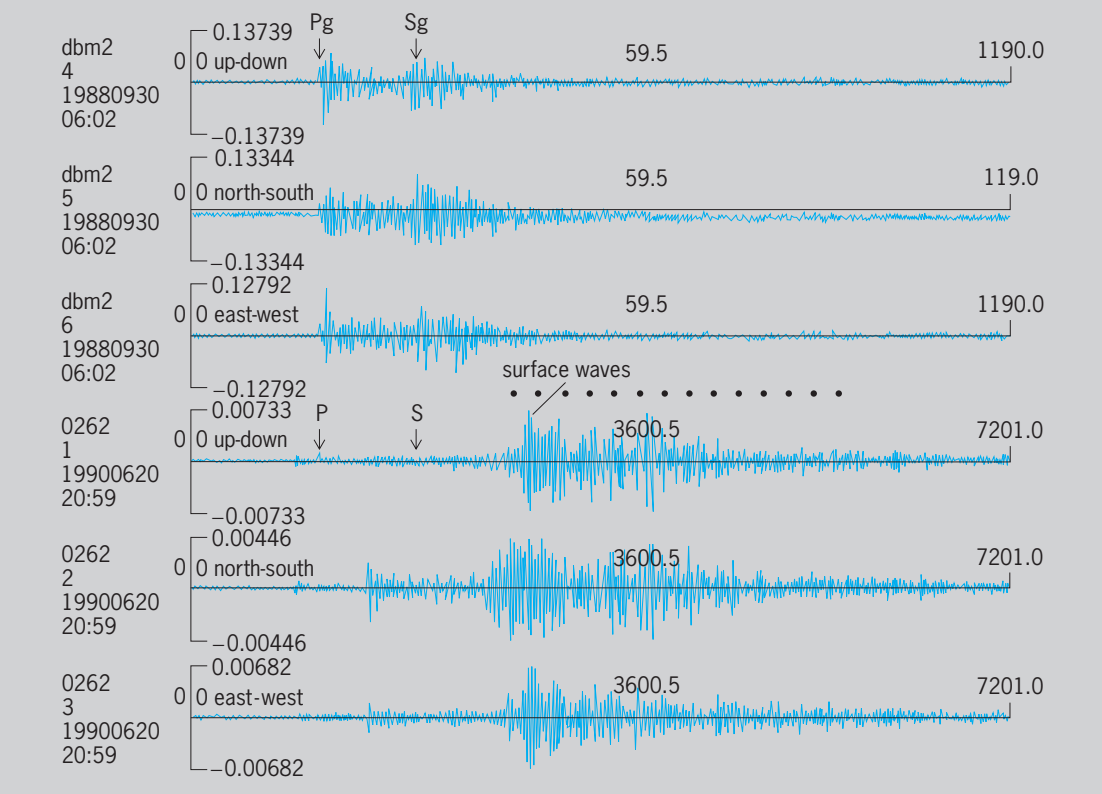


Fig. 11. Regional and teleseismic seismograms. The top three traces (seismometer installation dbm2) are the up–down, north–south, and east–west components of the regional seismogram from a 1-ton chemical explosion in the Adirondack Mountains of New York State (recorded at a distance of 107 km or 66 mi for 2 min). The regional phases Pg and Sg are indicated. These waves are guided by the crust over long distances and thus are very complicated. The bottom three traces (seismometer installation 0262) are the three components of the teleseismic seismogram produced by an earthquake in Iran (June 20, 1990; recorded at Palisades, New York, at a distance of 84°, or more than 9300 km or 5780 mi from the epicenter for 2 h). In addition to the P and S arrivals that have propagated through the Earth's mantle, the large-amplitude low-frequency arrivals later in the record are surface waves that have propagated in the Earth's outer layers. Detailed information is given in the text.

incoming wavefield. For the regional seismogram, the first arrival (Pg) is the direct P wave, which dives through the crust; the second arrival (Sg), beginning some 15 s after the Pg, is the crustal S phase. Even though explosions are inefficient generators of shear waves, it is common on regional seismograms to see substantial shear energy arising from near-source conversions of compressional to shear motion. The arrival time of the Pg phase can be picked easily, but the arrival time of Sg is somewhat more obscure. This is because some of the shear energy in the direct S is converted into compressional energy by scatterers in the crust near the receiver. This converted energy, traveling at compressional-wave speeds, must arrive before the direct S. The extended, very noisy wavetrains following the direct P and S are called codas; they represent energy scattered by small heterogeneities elsewhere in the crust. It is difficult to analyze these codas deterministically, and statistical procedures are often used. One result from the analysis of these and other seismograms is that the continental crust is strongly and three-dimensionally heterogeneous and scatters seismic energy very efficiently.

For the teleseismic seismogram, the teleseismic record is 60 times longer than the regional seismogram, and has been truncated only for plotting purposes. Plotted at this compressed time scale, it looks quite similar to the regional seismogram, but in fact there are important differences. The direct P and direct S waves are clearly evident; the S waves are larger on the horizontal components and larger than the P waves, a common feature for natural sources such as earthquakes. The very large-amplitude, long-period arrival starting some 20 min after the S are the surface waves. These waves are generally the most destructive in a major earthquake because of their large amplitude and extended duration. Between the direct P and the direct S are compressional body waves which have shallower take-off angles, and have bounced once (PP) or twice (PPP) from the surface before arriving at Palisades (the analogy here, though inverted, is like skipping a rock over water). The high level or energy between the S wave and the first surface wave is due to other multiply surface-reflected and -converted body waves, which are very sensitive to the structure of the upper mantle.

For purposes of analysis, the travel times of major phases have been inverted tomographically for large-scale structure. The International Seismological Centre (ISC) in England has been collecting arrival-time picks from station operators around the world since the turn of the century. Nearly one thousand

stations report arrival-time picks consistently enough so that their accuracy can be judged without viewing the original seismograms. More than 2 million of these P-wave arrival times have been inverted tomographically for the structure of the lower mantle, the core, and the core–mantle boundary. These inversions show that the core–mantle boundary and the mantle just above it are very heterogeneous. Tomographic inversions of special P phases that transit the boundary show that the boundary may have topography as well. Estimates of the amplitude of this topography range from a few hundred meters to several kilometers, but further work must be done.

Direct P phases dive deeply into the mantle and are not very sensitive to upper mantle structure unless the earthquakes are within 25° or so of the seismometer. Multiple bounce body phases are more sensitive to upper-mantle structure, but these must be picked carefully from the teleseismic records. This has been done for several thousand seismograms situated in the United States and Europe, and the resulting models give very finely detailed views of the structure of the mid to upper mantle. These methods are sensitive enough to resolve the remnants of old subducted slabs in the lower mantle and the large keels of high-velocity material which seem to underlie most old continents.

Going beyond tomographic methods requires analyzing more of the seismogram, for it is in the details of the wave shapes that information about major structures and the boundaries between them can be found. For example, surface waves disperse (different periods travel at different velocities) with wave speeds that depend on the fine details of the crust and upper mantle. An example of this dispersion can be seen in the Iranian seismogram (Fig. 11). Surface waves with longer periods arrive first and thus travel faster than surface waves at shorter periods. Measuring surface-wave dispersion is difficult but feasible, particularly for shallow earthquakes such as the Iranian event where the surface-wave arrivals dominate the seismogram. Seismologists began to measure dispersion in the late 1950s; some of the first applications of computers to seismological problems were the calculation of dispersion curves for oceanic and continental upper mantle and crust.

Measuring dispersion and other wave properties has evolved into a class of inversion procedures in which the entire seismogram is inverted directly for Earth structure. Such waveform inversion methods became feasible when computers became powerful enough to allow synthesis of most of the seismogram (**Fig. 12**). After many years of operation, even the sparse global digital arrays collect enough data to contemplate tomographic inversions. The most successful of these experiments have combined the seismogram-matching techniques of waveform inversion with a generalization of the tomographic approach to obtain models of the three-dimensional variation of upper mantle structure. The vertical integration shown in **Fig. 13** is a useful way of describing the geographic (two-dimensional) variation of a three-dimensional structure. Dark areas are fast relative to an average Earth, while light areas are slow. The darkest regions correspond to the oldest continental crust, known as cratons, while the lightest areas correspond to regions of active rifting either at mid-ocean ridges or at incipient oceans such as the Red Sea. *See* WAVEFORM.

Imaging the seismic source. The other half of the imaging problem in global seismology is constructing models of the seismic source. The simplest description of a so-called normal earthquake source requires two orthogonal force couples oriented in space. One force couple occurs on opposing sides of the rupture or fault surface and can be understood as the stress on either side of the fault induced by the earthquake rupture. The other couple is normal to the first and is required to conserve angular momentum. Implicit in this representation is the assumption that the earthquake is a point in both space and time (the point-source approximation), so that this type of source representation is known as a point-source

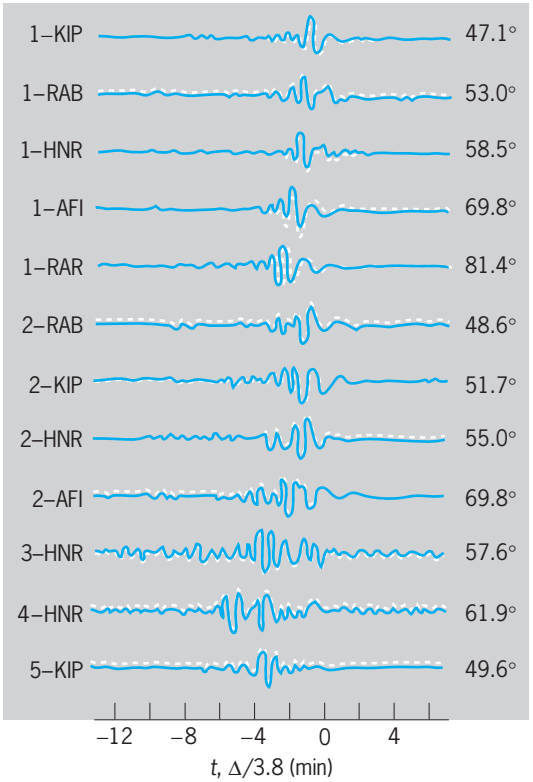


Fig. 12. Example of waveform inversion applied to fundamental and higher-mode surface waves crossing the western Pacific. The solid lines are data, and the broken lines represent synthetic seismograms computed for a model of the Earth obtained by waveform inversion. The numbers and letters to the left of each trace correspond respectively to earthquakes and stations used in the inversion. The distance between the epicenter and the recording station in degrees is given to the right of each trace. The seismograms are plotted as a function of reduced time, in which the time axis is translated by an amount proportional to the epicentral distance divided by a chosen velocity (Δ/3.8). This causes arrivals having the same velocity to line up vertically. The fits of the synthetic to the observed are very nearly "wiggle-for-wiggle" at nearly all distances, indicating the power of the waveform inversion technique.

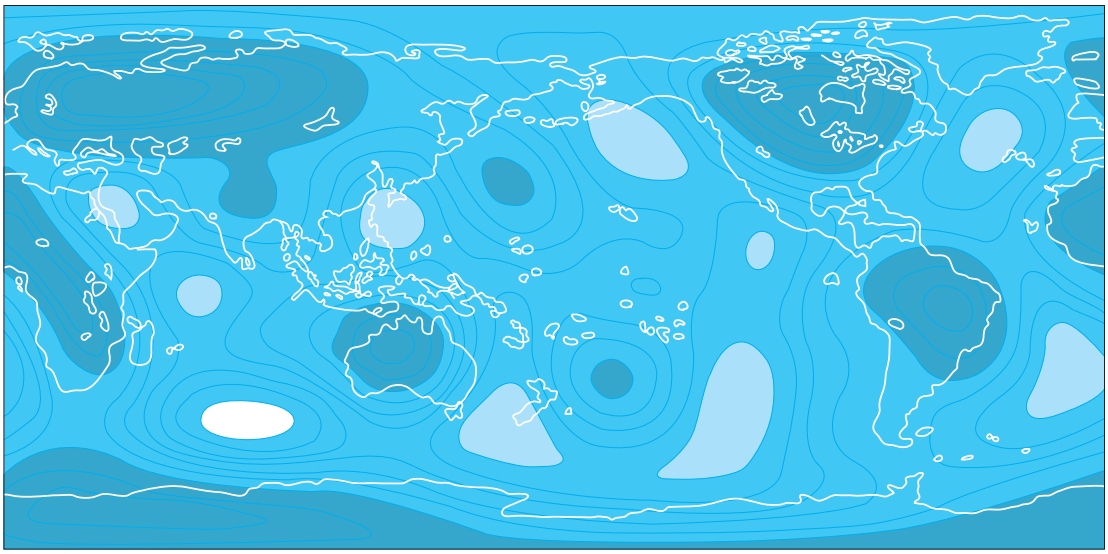


Fig. 13. Travel time anomaly from 400 to 40 km (240 to 25 mi), showing the deviation away from an average Earth of vertical shear-wave travel time derived by vertically integrating a three-dimensional model of shear–velocity variation. The model was developed by the inversion of several thousand long-period seismograms for upper-mantle velocity variations and then integrated. Dark areas are fast relative to the Earth, while light areas are slow. The darkest regions correspond to the oldest continental crust, known as cratons, while the lightest areas correspond to regions of active rifting either at mid-ocean ridges or incipient oceans such as the Red Sea.

double couple. Although there are some minor complications, the orientation of this double couple, and hence the orientation of the fault plane and the direction of rupture, can be inferred by analyzing the polarity of the very first P and S waves recorded worldwide. Thus this representation is known as a first-motion mechanism. It is not very difficult to make these observations provided that the instrument polarities are known, and the ensuing "beachball" diagrams are commonplace in the literature.

First-motion representations of seismic sources are the result of measurements made on the very first P waves or S waves arriving at an instrument; therefore they represent the very beginning of the rupture on the fault plane. This is not a problem if the rupture is approximately a point source, but this is true in practice only if the earthquake is quite small or exceptionally simple. An alternative is to examine only longer-period seismic phases, including surface waves, to obtain an estimate of the average point source that smooths over the space and time complexities of a large rupture. This so-called centroid-moment-tensor representation is an accurate description of the average properties of a source; it is routinely computed for events with magnitudes greater than about 5.5. Because an estimate for a centroid moment tensor is derived from much more of the seismogram than the first arrivals, it gives a better estimate of the energy content of the earthquake. This estimate, known as the seismic moment, represents the total stress reduction resulting from the earthquake; it is the basis for a new magnitude number M_W. This value is equivalent to the Richter body wave (m_b) or surface-wave magnitude (M_S) at low magnitudes, but it is much more accurate for magnitudes above about 7.5. The largest earthquake ever recorded, a 1960 event in Chile, is estimated to have had an M_W of about 9.5. For comparison, the 1906 earthquake in San Francisco had an estimated M_W of 7.9, and the 1964 Good Friday earthquake in Alaska had an M_W of 9.2.

While first-motion and centroid-moment-tensor representations are useful for general comparisons among earthquakes, there is still more information about the earthquake rupture process to be gleaned from seismograms. Types of analysis and instruments have been developed that demonstrate that some earthquakes, especially larger ones, are not adequately described by a first-motion or the averaged centroid-moment-tensor representations and require a more complex parametrization of the source process. These additional parameters may arise from relaxation of the point-source approximation in space or time or perhaps even of the double-couple restriction. In the former case, the earthquake is said to be finite, meaning that there is a spatial or temporal scale defining the rupture. The finiteness may also be manifested by multiple events occurring a few tens of seconds and a few tens of kilometers apart. In the latter case, the earthquake may have non-double-couple components that could be the result of explosive or implosive chemical phase changes, landslides, or volcanic eruptions. The size of the explosive non-double-couple component is one way of discriminating an earthquake from a nuclear explosion, for example.

One indication of source finiteness is an amplification of seismic signals in the direction of source rupture. Normally, the variation of seismic amplitudes with azimuth from a double-couple source varies in a predictable quadrupolar or bipolar pattern (the beach ball illustrates this best). For some earthquakes, however, the seismic waves that leave the source region at certain azimuths are strongly amplified. This has been ascribed to propagating ruptures

which "unzip" the fault along a fairly uniform direction. The best estimate is that most faults rupture at about two-thirds of the shear-wave velocity, but some faults may rupture even more slowly. Both unilateral and bilateral ruptures have been observed, and an important theoretical area in seismology is the examination of more complex ruptures and the prediction of their effects on observed seismic signals.

Another indication of source finiteness is the fact that some large events comprise smaller subevents distributed in space and time and contributing to the total rupture and seismic moment. The position and individual rupture characteristics of these subevents can be mapped with remarkable precision, given data of exceptional bandwidth and good geographical distribution. An outstanding problem is whether the location of these subevents is related to stress heterogeneities within the fault zone. These stress heterogeneities are known as barriers or asperities, depending on whether they stop or initiate rupture. The mappings of stress heterogeneities from seismological data is an active area of research in source seismology. *See* SEISMOGRAPHIC INSTRUMENTATION.

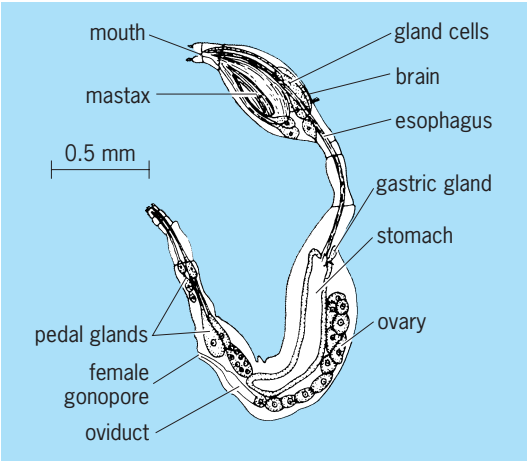
John Mutter; Art Lerner-Lam

Bibliography. J. Havskov and G. Alguacil, *Instrumentation in Earthquake Seismology (Modern Approaches in Geophysics: Vol. 22)*, 2004; C. H. Scholz, *The Mechanics of Earthquakes and Faulting*, 2002; S. Stein and M. Wysession, *An Introduction to Seismology, Earthquakes and Earth Structure*, 2002.

Seisonacea

A class of the phylum Rotifera which comprises a group of little-known marine animals. They form a single family with about seven species and are found only in Europe. The Seisonacea are epizoic or possibly ectoparasitic on crustacea, especially on members of the genus *Nebalia*. They have a very elongated jointed body with a small head; a long, slender neck region; a thicker, fusiform trunk; and an elongated foot, terminating in a perforated disk (see **illus.**). The Seisonacea differ from other rotifers in a number of characteristics. They have a very reduced corona, a quite aberrant type of mastax (fulcrate), paired ovaries but no vitellaria in females, and paired testes in the males. Furthermore, the sexes are of similar size and development; only one type of egg is produced, and it requires fertilization. The Seisonacea are also larger than other rotifers, attaining up to 0.12 in. (3 mm) in length. *See* ROTIFERA.

Elbert H. Ahlstrom



***Seison*, a rotifer.**

Seizure disorders

Conditions in which there are recurrent seizures. Such conditions are also known as epilepsy; the isolated occurrence of a seizure, however, is not designated as epilepsy. A seizure (ictus) is an event in which there is a sudden alteration in function of nerve cells, most commonly involving excessive electrical activity of the cells. This sudden change in nerve cell function is usually relatively brief, lasting seconds to minutes. Soon after a seizure, the brain may function quite normally. The manifestation of a seizure varies depending on which area of the brain is involved. *See* BRAIN; NERVOUS SYSTEM (VERTEBRATE).

Classification and symptomatology. A standard way to describe the seizures is outlined below. [The traditional nomenclature has been added in brackets.]

Nomenclature of the International Classification of Epileptic Seizures

I. Partial seizures (seizures beginning locally)
  A. Simple partial seizures (without impaired consciousness)
    1. With motor symptoms [focal motor seizures]
    2. With somatosensory or special sensory symptoms [focal sensory seizures]
    3. With autonomic symptoms
    4. With psychic symptoms [temporal lobe seizures]
  B. Complex partial seizures (with impaired consciousness)
    1. With impaired consciousness only [often confused with absence, or petit mal, seizures, from which it can be distinguished by an EEG]
    2. With automatisms [psychomotor seizures; temporal lobe seizures]
  C. Partial seizures secondary generalized [focal seizures progressing to grand mal (major motor) seizures, such as Jacksonian seizures]
II. Generalized seizures
  A. Generalized nonconvulsive seizures
    1. Absence seizures [petit mal seizures]
    2. Atonic seizures [atonic seizures; minor motor seizures; drop attacks]
  B. Generalized convulsive seizures
    1. Generalized tonic-clonic seizures [grand mal seizures; major motor seizures]

2. Myoclonic seizures [myoclonic seizures; minor motor seizures]
3. Tonic seizures [tonic seizures]
4. Clonic seizures [clonic seizures]

III. Unclassified epileptic seizures

Advances in the diagnostic evaluation of persons with epilepsy have also led to the classification of seizure disorders by epilepsy syndrome. The two major categories are considered as either generalized or localization-related (partial), with subclassifications of idiopathic, cryptogenic, symptomatic, or progressive for each syndrome.

The following are examples of the four common seizure types.

Focal motor epilepsy. This type of seizure, also known as a simple partial seizure with motor symptoms, is manifested by uncontrolled rhythmic jerking of the face, arm, or leg, caused by excessive abnormal discharges of nerve cells within the area of the brain, which under usual circumstances controls movement in that part of the body (**Fig. 1**). The abnormal nerve cell activity can be seen in the electroencephalogram recorded from electrodes on the scalp (**Fig. 2***b*). *See* ELECTROENCEPHALOGRAPHY.

Temporal lobe seizures. These seizures are known as simple partial seizures and may be manifested by a myriad of symptoms depending upon which part

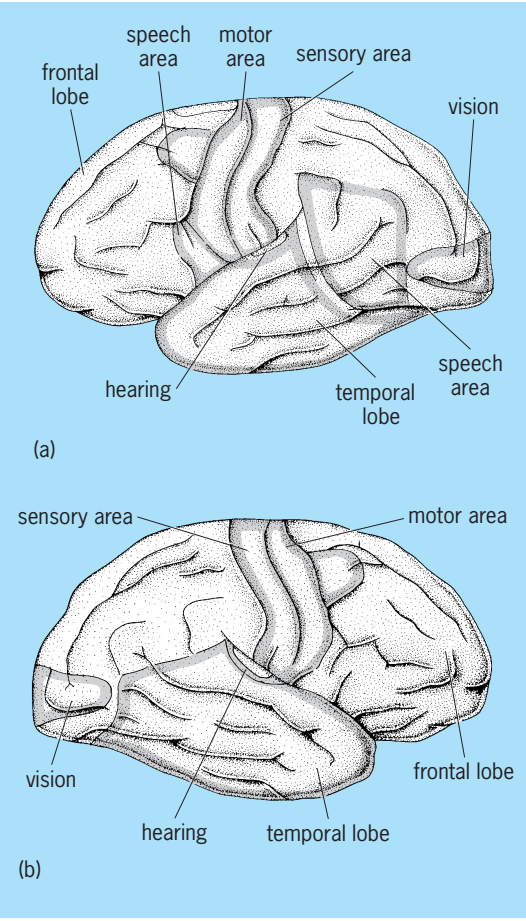


Fig. 1. Areas of the (*a*) left and (*b*) right brain.

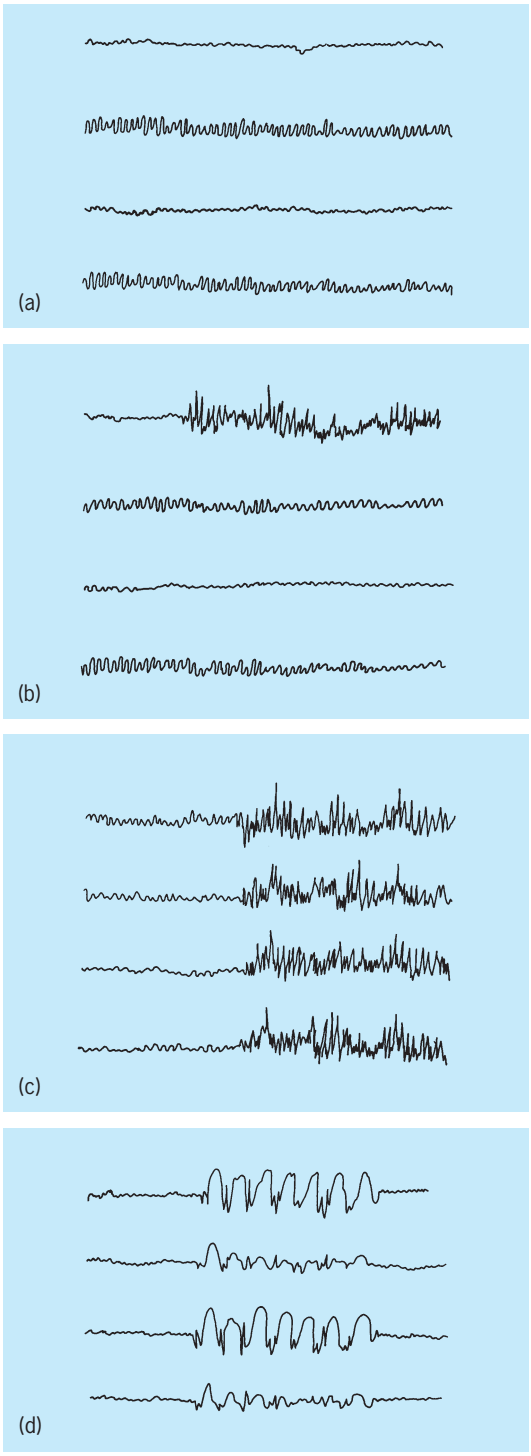


Fig. 2. Electroencephalograms: (*a*) Normal. (*b*) Partial (focal) seizure. (*c*) Generalized tonic-clonic (grand mal) seizure. (*d*) Absence (petit mal) seizure.

of the lobe is involved. Psychomotor, or complex partial, seizures are the most common type. Clinically, they are characterized by an alteration in the state of consciousness and performance of repetitive, patterned, non-goal-directed activity. Usually, the person is aware of having had the seizure but may not remember what happened during it. Occasionally, temporal lobe seizures produce visual, auditory, or olfactory hallucinations, or illusions,

such as feelings of fear, familiarity, or distortions of size. These types of seizures are characterized electrically by abnormal discharges occurring within the temporal lobe for the duration of the seizure (Fig. 2*b*).

Grand mal. These seizures, also referred to as generalized tonic-clonic convulsive seizures, major motor seizures, or convulsions, occur if the abnormal discharges involve the entire brain all at once. In this condition there are forceful, generalized, symmetrical musculature contractions accompanied by loss of consciousness, and, at times, by urinary incontinence and tongue biting. In some instances, characteristic experiences of a sensory or visceral nature precede the more dramatic motor manifestations. Such usually brief subjective phenomena are termed auras. Actually, however, they represent the very beginning of an ictal episode. They are important for diagnostic purposes since they may indicate the portion of the brain in which the epileptogenic process is located.

A seizure comes to an end abruptly or gradually, and, depending on the pattern and intensity of the attack, the person may return immediately to normal preictal conditions or may continue to show various physical or mental deficits or both for some time. Amnesia is rather common, and after major seizures the subject often falls into a sleeplike state or goes through a period of confusion. *See* AMNESIA.

The relatively uncommon but serious life-threatening condition in which one convulsion occurs after another without recovery between them and lasts for hours is called status epilepticus.

Electroencephalographically, generalized tonic-clonic seizures are characterized by generalized rhythmic spike discharges in all areas of the brain at once which last several seconds and are followed by other generalized abnormalities for the duration of the seizure and for varying times thereafter (Fig. 2*c*).

Generalized nonconvulsive seizures. These may be atonic seizures, which are characterized by a sudden loss of muscle tone, or they may be absences (petit mal), which consist of brief periods of loss of consciousness and immediate recovery. The subjects with absences are often unaware of having had the episode. Electroencephalographically, they are characterized by generalized bilaterally synchronous spike and wave discharges occurring at three cycles per second (Fig. 2*d*). This type of seizure is more frequently seen in children than in adults.

Etiology. Epilepsy is not a disease in itself. It is a symptom of an underlying disease process. That disease process may be metabolic, such as uremia, decreased brain oxygen, or low calcium levels; or structural brain damage, such as from head trauma at birth, brain injuries, brain tumors, strokes, or congenital malformations, or previous encephalitis or meningitis. In many instances, however, a cause is not found. The disorder is then referred to as idiopathic or primary epilepsy. In some types of epilepsy there is apparently also a genetic component. The modes of genetic transmission are likely varied and may include the tendency to develop seizures with acquired brain damages as well as some types of "idiopathic" epilepsy. The role of genetic transmission in epilepsy must be evaluated on an individual basis.

Prognosis. The prognosis in epilepsy differs with the type of seizure. Some seizure types, once established, may persist. Other seizure types may have a high degree of spontaneous remission, especially those of childhood onset, such as febrile seizures and some absences (petit mal seizures). Seizures that are posttraumatic in origin may also remit spontaneously. In persons with a tendency to have epilepsy, events such as puberty, menses, alcohol intake, or emotional stress may precipitate seizures.

Diagnosis. Seizures should be distinguished from other conditions which have some clinical similarities. These include fainting from hypoglycemia, cardiovascular disorders, or hysteria.

The tools necessary to make this differential diagnosis are the history of the illness, the general physical exam, and a neurologic exam, all of which may be entirely normal or which may demonstrate signs of underlying disease processes. An electroencephalogram may demonstrate abnormal electrical discharges during a time the patient is seizure-free. These discharges may indicate that portion of the brain from which the seizures arise. Computerized tomography scans or magnetic resonance imaging (MRI) may show any gross structural abnormality. Examination of the blood may demonstrate abnormal circulating chemicals. At times an examination of cerebrospinal fluid may give information regarding the underlying cause of the epilepsy. In persons with medically refractory epilepsy, long-term EEG-video monitoring is employed to document seizure types, and the intracarotid sodium amytal (Wada) test is used to determine side of language dominance. Functional magnetic resonance imaging (fMRI) may assist in lateralization of language, and when used in conjunction with the EEG, in localization of epileptic discharges. *See* COMPUTERIZED TOMOGRAPHY; MEDICAL IMAGING.

Treatment. The ideal treatment of epilepsy is removal of the cause, such as excision of a tumor, or correction of a metabolic disorder, such as uremia or hypocalcemia. In many instances, however, the cause cannot be established or may not be amenable to direct treatment. When the cause cannot be removed, the symptoms (seizures) are treated.

It is important to control seizures for biologic and social reasons. Because of metabolic demands made by the abnormally discharging nerve cells, the seizures themselves may cause damage to nearby nerve cells and render the individual more susceptible to further seizures. The social importance of controlling seizures is obvious: a person who has frequent seizures cannot maintain a driver's license and may have difficulties with employment. The initial method of obtaining seizure control is antiepileptic drugs. About 80% of the people with epilepsy obtain good control or elimination of seizures with

medication. These medications are chemicals of varying structures and may be effective by a number of different brain mechanisms.

Medication. The following drugs are commonly used to control seizures. Phenytoin, carbamazepine, and phenobarbital have been used for decades for individuals with partial (focal) or secondarily generalized convulsive seizures while ethosuximide or valproic acid may be beneficial in generalized, nonconvulsive seizures (absence, petit mal). A new generation of drugs, including topirimate, lamotrigine, levetiracetam, zonisimide, and oxcarbazepine, are now available, and most of these may have broad spectrum efficacy against a variety of seizure types. In individuals whose seizures are more difficult to control, combinations of these drugs might be used, or other agents may be added. All antiepileptic drugs when used in high enough doses may produce toxic effects. The most commonly observed toxic effects from one or more of these agents include impairment of balance, double vision, or lethargy. It is possible to measure the amount of a circulating drug in the blood by the use of gas-liquid chromatography. In this way therapeutic ranges have been established that are a guide to the adjustment of medication. There are substantial individual differences in the way these drugs are absorbed and metabolized, so considerably different oral doses may be needed to achieve the same drug levels in blood in different patients. Some of these drugs may have an adverse effect on the fetus, and counseling with a neurologist and obstetrician is desirable before and during pregnancy.

Surgery. Brain surgery is a therapeutic potential for some of the 20% of the people with epilepsy who do not achieve control on medication. Surgery may be considered if the abnormally discharging nerve cells which cause the epilepsy are in a dispensable area of brain, that is, one of the frontal or temporal lobes (Fig. 1). One of these regions may be excised in one cerebral hemisphere without significant deficit. About 10% of the individuals refractory to medicine may be candidates for excision of the epileptic focus, and of these about 70% will achieve seizure control. Thus, epilepsy is a symptom, not a disease in itself, and when treated often has a good prognosis. However, about 15–20% of epileptics cannot achieve seizure control with the present modes of therapy.

Prevention. The occurrence of epilepsy can be decreased by measures to reduce head injury and birth defects. There are some indications that epilepsy can also be reduced by prophylactic administration of antiepileptic drugs in conditions which, if untreated, may lead to development of chronically recurring seizures. These include febrile convulsions of childhood, traumatic head injuries, and postoperative conditions of patients who have had brain surgery for other reasons.

Epilepsy was once considered a sacred disease and was surrounded by many myths and superstitions. Many rules concerning social interactions were imposed on people with seizures. In recent times, understanding of the entity as an organic disorder and recognition that it is not an all-encompassing disease have led to changes in legislation and discriminatory practices. Modern therapy has rendered many people free of seizures with little or no medication toxicity so that they may lead normal lives. Public education is another area which has been productive in determining early recognition and treatment of seizures and in prevention of epilepsy.

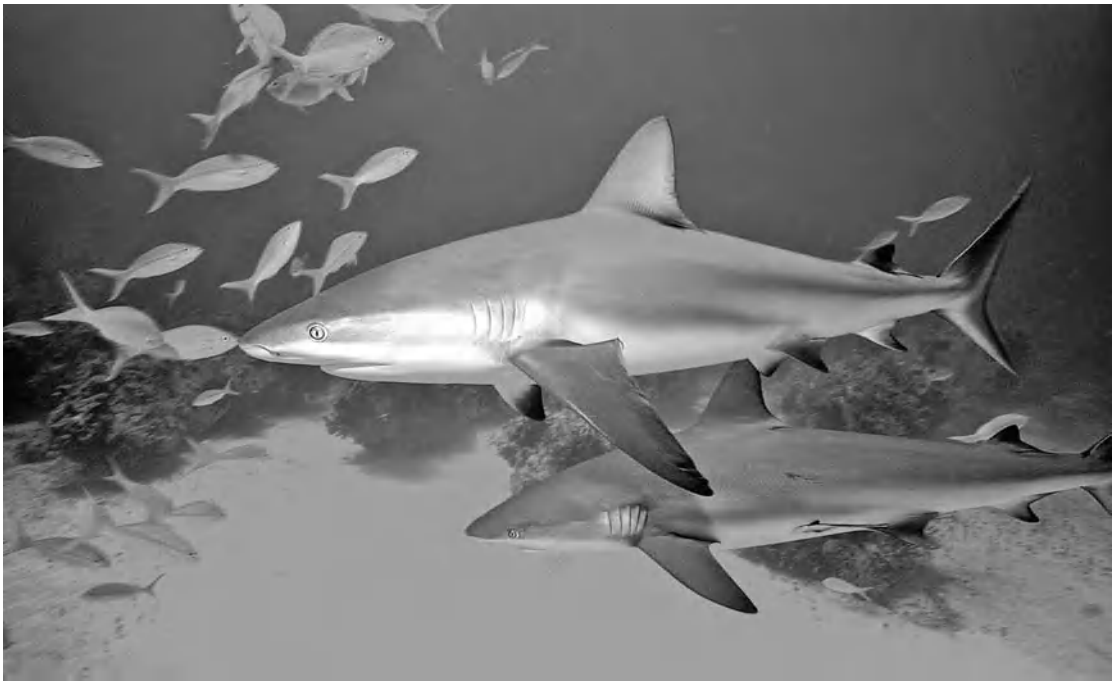
Linda Moretti Ojemann; Mark D. Holmes

Bibliography. F. E. Dreifuss, Proposal for revised clinical and electrographic classification of epileptic seizures, *Epilepsia*, 22:489–501, 1981; J. E. Engel and T. Pedley (eds)., *Epilepsy: A Comprehensive Textbook*, 3 vols., 1998; W. A. Hauser and D. C. Hesdorffer, *Epilepsy: Frequency, Causes and Consequences*, 1990; J. Laidlaw, A. Richens, and D. Chadwick (eds.), *A Textbook of Epilepsy*, 4th ed., 1993; G. A. Ojemann, Surgical therapy for medical intractable epilepsy, *J. Neurosurg.*, 66:489-499, 1987; C. P. Panayiotopoulos, *A Clinical Guide to Epileptic Syndromes and Their Treatment*, 2002; W. Penfield and H. Jasper, *Epilepsy and the Functional Anatomy of the Human Brain*, 1954; H. Weiser, Future aspects of epilepsy research, *Acta Neurochir. Suppl.*, 84:1–16, 2002.

Selachii

One of two subdivisions of the subclass Elasmobranchii, consisting of nine extant orders and 34 extant families, collectively known as sharks. The other subdivision, Batoidea, includes the sawfishes, guitarfishes, skates, and rays. In the classification system below, daggers indicate extinct orders. Each Recent order is covered by a separate article in the encyclopedia.

Class Chondrichthyes
  Subclass Elasmobranchii
    Subdivision Selachii
      †Order Cladoselachiformes
      †Order Symmoriiformes
      †Order Ctenacanthiformes
      †Order Hybodontiformes
      Order Heterodontiformes
        (bullhead sharks)
      Orectolobiformes
        (carpet sharks)
      Lamniformes
        (mackerel sharks)
      Carcharhiniformes
        (ground sharks)
      Hexanchiformes
        (frill shark and cow sharks)
      Echinorhiniformes
        (bramble sharks)
      Squaliformes
        (sleeper sharks and dogfish sharks)
      Squatiniformes
        (angel sharks)
      Pristiophoriformes
        (saw sharks)



Caribbean reef shark (*Carcharhinus perezi*), a species of ground shark. (*Getty Images © Digital Vision*)

There are several characters common to all sharks that distinguish them unequivocally from all other elasmobranchs: the gill openings are at least partly lateral (that is, they can be seen in lateral view); the edges of the pectoral fins are not attached to the sides of the head anterior to the gill openings; and the upper margin of the orbit is free from the eyeball, that is, there is a free lid. *See* BATOIDEA; ELASMOBRANCHII.

Morphology. Sharks vary considerably in body form. On one end of the physiognomic spectrum are the mackerel and ground sharks (see **illustration**), with sleek fusiform bodies built for speed and power, while on the other end are angel sharks, with a batoid-shaped body adapted for cruising on or near the sea floor. Sharks vary greatly in size, ranging from only 0.4 m (1.3 ft) in adult smooth dogfishes to, reputedly, 18 m (59 ft) in whale sharks.

Ecology and development. Sharks are carnivorous predators for the most part, feeding on any flesh available, be it fresh or carrion; however, the largest of all sharks, whale sharks and basking sharks, feed on small schooling fishes, as well as tiny crustaceans. The poor eyesight of sharks is compensated by a very high order of olfaction. In most sharks the teeth vary in size and shape from front to back and in the upper and lower jaws. The teeth occur in series of rows, with replacement teeth always in reserve, waiting to rotate outward as teeth are lost. Development is oviparous in a few sharks, ovoviviparous in most, and viviparous in yet others.

Distribution. The great majority of sharks are found in relatively shallow waters around the world, primarily tropical and subtropical latitudes of both hemispheres. Their numbers decrease in temperate waters and drop precipitously in north temperate zones.

Herbert Boschung

Fossil selachii. All fossil and Recent sharks, with the possible exception of Devonian cladoselachians, may be included in the Selachii. The selachian endoskeleton is entirely cartilaginous, but it is frequently calcified superficially with calcium phosphate prisms. Although Recent sharks have a single layer of such tesserae, many Paleozoic forms had much thicker calcification consisting of many layers. The body is covered in minute scales, which in Recent sharks are simple and nongrowing (placoid), but which in earlier sharks often fused together and became enlarged during life (zonal scales) and may have been more regularly arranged over the body.

Most extant shark families are represented by the Late Cretaceous, apart from whale sharks (family Rhinodontidae), false cat sharks (Pseudotriakidae), and the frilled shark, *Chlamydoselachus*, all of which have a very poor fossil record. Cow sharks (Hexanchoidea), Port Jackson sharks (Heterodontidae), nurse sharks (Orectoloboidea), angel sharks (Squatinoidea), dogfish (Scyliorhinidae), and rhinobatoids had already developed before the close of the Jurassic. One family (Palaeospinacidae) appears in the Late Triassic (*Palaeospinax*) and ranges into the early Tertiary (*Synechodus*) before becoming extinct.

Because sharks periodically shed their teeth, sometimes replacing several thousand in a lifetime, their teeth are common in the fossil record. It is, therefore, rather surprising that the earliest records of sharks are from the Late Devonian, when they were already highly diversified. Knowledge of these early forms has improved greatly in recent years, particularly from the Pennsylvanian of North America.

Cladoselachians. Cladoselachians (late Paleozoic) are not known to possess male intromittent organs (claspers). Cladoselachian pectoral fins have a

single basal endoskeletal cartilage (metapterygium) and numerous long, unjointed radials, of which the anteriormost articulate directly with the shoulder girdle. The eyes are surrounded by a ring of sclerotic plates. A bladelike spine projects from the back, anterior to the first dorsal fin. *See* CLADOSELACHII.

Symmoriids. These mainly Carboniferous selachians have pectoral fins like cladoselachians, but the anterior radials are jointed to form a basal series. Radials of the other fins are also jointed once, and there is only one dorsal fin, behind which is a peculiar median cartilage. Most forms lack spines, but *Stethacanthus* has a single dorsal spine.

Ctenacanthids. The pectoral fin of these late Paleozoic sharks is tribasal, with a pro-, meso-, and metapterygium. In addition, there are two dorsal fins, each with a deeply inserted fin spine, and jointed radials in all fins. Since these characters also occur in the hybodontids, there is every possibility that ctenacanths are not a monophyletic taxon, but simply represent a primitive assemblage in need of further study. The aberrant xenacanths of the late Paleozoic may be closely allied to some ctenacanths. *See* XENACANTHIDA.

Hybodontids. These sharks primitively resemble ctenacanths and have long been considered intermediates between them and Recent sharks. However, hybodontids have distinctive cranial morphology, fin spines, head spines, and teeth, and calcified pleural ribs. Therefore, they are probably not ancestral to Recent sharks, particularly since they lack characters which unite these forms.

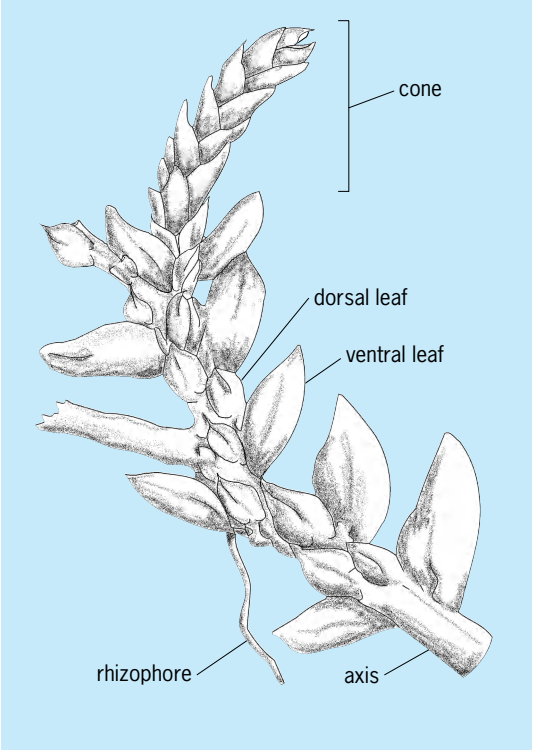
John Maisey

Bibliography. H. B. Bigelow and W. C. Schroeder, Sharks, in *Fishes of the Western North Atlantic*, Mem. Sears Found. Mar. Res. Mem. 1 (pt. 1), pp. 29–556, 1948; L. J. V. Compagno, Interrelationships of living elasmobranchs, in P. H. Greenwood, R. S. Miles, and C. Patterson (eds.), *Interrelationships of Fishes*, pp. 15–61, Zool. J. Linn. Soc. 53, Suppl. 1, Academic Press, New York, 1973; J. S. Nelson, *Fishes of the World*, 3d ed., Wiley, New York, 2006.

Selaginellales

An order of the class Lycopsida (club mosses), regarded as more advanced than the extinct Protolepidodendrales because they produce spores of two sizes, but less advanced than the Lepidodendrales and Isoetales because they lack wood and finite growth. All are perennial herbs, varying in growth habit from prostrate to climbing. *See* ISOETALES; LEPIDODENDRALES; PROTOLEPIDODENDRALES.

Although the approximately 700 extant species are traditionally assigned to a single genus, *Selaginella*, it is possible that this genus should include only the few species that possess undivided steles and microphyllous leaves of a single morphology. The bulk of the selaginellalean species are included in the genus *Stachygynandrum*. Fossil evidence suggests that the two lineages had already diverged by the early Carboniferous. Today, both genera are geographically widespread and occupy a wide range of habitats. Although selaginellaleans are rarely a significant element of the biomass in any of these communities, they dominate the ground flora in some tropical forests.



Characteristic features of a typical heterophyllous selaginellalean of the *Stachygynandrum* type. (*After T. R. Webster, Developmental problems in Selaginella in an evolutionary context, Ann. Mo. Bot. Gard., 79:633–647, 1992*)

Both selaginellalean genera show several remarkable biological features (see **illus.**). *Stachygynandrum* has leaves that are positioned both ventrally and dorsally. This character is regarded as an adaptation to efficient light capture; the orientation relative to light sources is precisely adjusted by differential growth, which is in turn controlled by auxins. Moreover, in many species the lower surface of the dorsal leaf and upper surface of the ventral leaf consist of more or less cubic cells that contain only one large chloroplast, whose orientation within the cell can be altered to maintain optimal levels of incident light for photosynthesis.

The origin of the rhizophores (specialized aerial rooting structures) is unclear.

Life history. Selaginellaleans have a free-sporing heterosporous life history. Sporangia are borne on sporophylls that are aggregated into cones. Each individual sporophyte is capable of producing both megasporangia, which on average contain four large megaspores, and microsporangia, which contain much larger numbers of much smaller microspores. Megasporangia generally develop in regions of the plant that experience the highest level of metabolic activity. Many species possess active

mechanisms for ejecting either microspores or entire microsporangia. *Selaginella selaginoides* disperses megaspores with a slingshot mechanism.

The megaspores and microspores germinate to produce independent, unisexual gametophytes. Both genders remain largely within the spore wall. The megagametophytes produce several archegonia that yield eggs, which chemically attract the swimming biflagellate spermatozoids released by the microspores. This requirement for an aquatic environment limits the habitats of sexually reproducing selaginellaleans.

Fossil record. Much of the fossil record of the Selaginellales is ambiguous. The more likely selaginellaleans are demonstrably both heterosporous and ligulate. *Selaginellites resimus*, from the early Carboniferous of England, is the earliest convincing record of the Selaginellales. Sporadic traces of possible selaginellaleans occur through the Mesozoic and Cenozoic. *See* LYCOPHYTA; LYCOPODIALES; LYCOPSIDA.

Richard M. Bateman; William A. DiMichele

Bibliography. E. E. Karrfalt, The comparative and developmental morphology of the root system of *Selaginella selaginoides, Amer. J. Bot.*, 68:244–253, 1981; J. E. Skog and C. R. Hill, The Mesozoic herbaceous lycopsids, *Ann. Mo. Bot. Gard.*, 79:648–675, 1992; B. A. Thomas, Paleozoic herbaceous lycopsids and the beginnings of extant *Lycopodium* and *Selaginella, Ann. Mo. Bot. Gard.*, 79:623–631, 1992; T. R. Webster, Developmental problems in *Selaginella* in an evolutionary context, *Ann. Mo. Bot. Gard.*, 79:632–647, 1992.

Selection rules (physics)

General rules concerning the transitions which may occur between the states of a quantum-mechanical physical system. They derive in almost all cases from the symmetry properties of the states and of the interaction which gives rise to the transitions. The system may have a classical (nonquantum) counterpart, and in this case the selection rules may often be related to the classical conserved quantities. A first use of selection rules is in determining the symmetry classes of the states, but in a great variety of ways they may yield other information about the system and the conservation laws. *See* QUANTUM MECHANICS; SYMMETRY LAWS (PHYSICS).

Angular momentum and parity rules. For an isolated system the total angular momentum is a conserved quantity; this fact derives from a fundamental fact of nature, namely, that space is isotropic. Each state is then classifiable by angular momentum J and its z component M $(= -J, -J + 1, \ldots, +J)$. Angular momenta combine in a vectorial fashion. Thus, if the system makes a particle-emitting transition $J_1,M_1 \rightarrow J_2,M_2$, the emitted particles must carry away angular momentum (j, μ), where $\mathbf{j} = \mathbf{j}_1 - \mathbf{j}_2$. This implies that $\mu = M_1 - M_2$ and that j takes on values $J_1 - J_2, J_1 - J_2 + 1, \ldots, J_1 + J_2$. Thus in transitions $(J = 4 \leftrightarrow J = 2)$ the possible j values comprise only 2, 3, 4, 5, 6, and, if it is also specified that $M_1 - M_2 = \pm 4$, only 4, 5, 6. Observe that J_z is additive. *See* ANGULAR MOMENTUM; QUANTUM NUMBERS.

Another fundamental symmetry, the parity, which determines the behavior of a system (or of its description) under inversion of the coordinate axes, is conserved by the strong and electromagnetic interactions, and gives a classification of systems as even $(\pi = +1)$ or odd $(\pi = -1)$. Under combination the parity combines multiplicatively. Thus, if the transition above is $4^{\pm} \rightarrow 2^{\mp}$, it follows that $j^{\pi} = 2^{-}, \ldots, 6^{-}$, while $4^{\pm} \rightarrow 2^{\pm}$ would give $j^{\pi} = 2^{+}, \ldots, 6^{+}$. The angular momentum $\mathbf{j}$ may be a combination of intrinsic spin $\mathbf{s}$ and orbital angular momentum $\mathbf{l}$. Scalar, pseudoscalar, vector, and pseudovector particles are respectively characterized by $s^{\pi_s} = 0^{+}$, 0^{-}, 1^{-}, 1^{+}, where π_s is the "intrinsic" parity, while l always carries $\pi_l = (-1)^l$. *See* NONRELATIVISTIC QUANTUM THEORY; PARITY (QUANTUM MECHANICS); SPIN (QUANTUM MECHANICS).

Electromagnetic transitions. The photon is a transverse vector particle; this implies that its electric field vector (classically) or its intrinsic spin vector (quantum-mechanically) lies in the plane normal to the propagation direction. Thus, while an ordinary vector particle with total angular momentum j may, for $l > 0$, have $l = j, j - 1, j + 1$, corresponding to the three internal degrees of freedom, a transverse particle may have $l = j$ or else a particular linear combination of $l = j \pm 1$ (two internal degrees). An $l = j$ photon ("magnetic" type, or Mj) then carries parity $(-1)^{j+1}$, while the $l = j \pm 1$ combination (electric type, or Ej) carries $\pi = (-1)^j$; the magnetic-electric nomenclature is appropriate because, in the long-wavelength (LWL) limit (that is, when the photon wavelength is much greater than the radius of the source) the radiation is generated by a magnetic-type interaction (for example, $\mathbf{L} \cdot \mathbf{H}$ or $\mathbf{S} \cdot \mathbf{H}$ for M1, with $\mathbf{H}$ the magnetic field) or electric-type interaction ($\mathbf{P} \cdot \mathbf{E}$ for E1, with $\mathbf{P}$ the dipole operator and $\mathbf{E}$ the electric field). For $j = 0$, the only combination of s and l is $(s, l, j) = (1, 1, 0)$, which describes an oscillating E0 field, which however vanishes outside the source and hence does not radiate. Thus, for transitions of the form $J^{\pm}_1 \leftrightarrow J^{\pm}_2$, photons of type M1, E2, M3, ... may be emitted, and, for $J^{\pm}_1 \leftrightarrow J^{\mp}_2$, E1, M2, E3, ..., the j values being restricted as above. If M_1, M_2 are also specified, as in Zeeman transitions, there are further restrictions. For dipole transitions $(j = 1)$, only $\mu = 0$ (π lines) and $\mu = \pm 1$ (σ lines) occur, the former not being radiated parallel to the magnetic field because of the photon transversality. If $M_1 = M_2 = 0$, the transition is forbidden unless $J_1 + J_2 + j$ is even; this inversion rule, not restricted to electromagnetic transitions, arises from the behavior of states and operators under a rotation through π radians about an axis in the x-y plane. It is important in discussing molecular symmetries and isospin selection rules. *See* MULTIPOLE RADIATION.

Approximate rules. The "exact" selection rules are to be supplemented by approximate rules which depend on the internal structure of the system. To the extent that L, S, the total orbital and spin angular

momenta, are good quantum numbers, as happens in many atoms and a few nuclei, the angular momentum radiated in a long-wavelength E1 or E2 transition (whose transition operators are spin-independent) must come from the orbital structure; thus $(\Delta\mathbf{L}, \Delta\mathbf{S}) = (1, 0), (2, 0)$, respectively. For the M1 operator, which has the form $\alpha\mathbf{L} + \beta\mathbf{S}$, $(\Delta\mathbf{L}, \Delta\mathbf{S})$ can have the values (1, 0) or (0, 1) but not (0, 0) or (1, 1); however, beyond that, since $\mathbf{L}$ and $\mathbf{S}$ can generate no orbital, spin, or radial excitations, M1 transitions in *LS*-coupling occur only between two levels of the same term (although this is not true for *Mj* with $j > 1$).

Many processes, including photon emission and beta decay, occur via one-body operators [for example, by

$$\mathbf{P} = \sum_i e_i \mathbf{r}(i)$$

for E1, where the sum is over particles, with the ith particle having charge e_i and position operator $\mathbf{r}(i)$]. These generate only single-particle transitions, in which at most one particle changes orbit. (The number of such particles is $\delta n = 1$ for odd parity, as in

$$p^2 \xrightarrow{E1} sp,\ pd;$$

$\delta n = 0, 1$ for even parity, as in

$$d^2 \xrightarrow{E2} d^2,\ sd,\ dg;$$

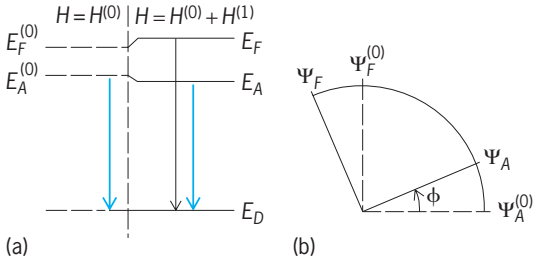
$\delta n = 0$ for M1 in *LS*-coupling, but not in *jj*-coupling, since, for example,

$$f_{5/2} \xrightarrow{M1} f_{7/2}$$

is allowed.) Because of this "$\delta n < 1$" rule, the $(J\pi)$ rules apply to individual particles (for example, $\Delta l^{\pi} = 1^-$ for E1) and thus greatly facilitate study of the microscopic structures of the states involved. *See* ATOMIC STRUCTURE AND SPECTRA; NUCLEAR STRUCTURE.

Forbidden transitions. Forbidden transitions are those which are hindered by a significant selection rule or combination of them (the term is often applied in a more restricted sense in some domains, for example, in atomic physics, to all photon transitions except E1). Several examples will be used to illustrate the remarkable variety of information which they yield.

Parity violation. The alpha-decay process $A(J^{\pi}) \rightarrow B(0^+) + \alpha(0^+)$ is allowed only if $\pi = (-1)^J$ since the final angular momentum J is orbital only. The forbidden process, with A a $2-$ state in ^{16}O and B the ^{12}C ground state, has been observed, with a lifetime on the order of 10^{13} larger than that of a nearby $2+$ state in ^{16}O. Since neither of the decay products has a nearby 0^- companion, the decay can be ascribed to a small 2^+ admixture in the $2-$ state, the admixing "intensity" (square of the 2^+ amplitude) being then on the order of 10^{-13}. This gives a measure of a fundamental symmetry breaking, the violation of parity by the weak interaction. This example, like several of those following, involves symmetry breaking in a two-level substructure, the simple quantum mechanics of which is explained in the **illustration**.



Symmetry admixing in a two-level substructure. (*a*) Energy-level diagram. (*b*) Rotation of states.

Two close-lying states Ψ_A and Ψ_F are eigenfunctions of a hamiltonian $H = H^{(0)} + H^{(1)}$ with energies E_A and E_F (illus. *a*), where $H^{(0)}$ preserves a symmetry and the small $H^{(1)}$ breaks it. Transitions occur to a distant state Ψ_D (with energy E_D), in which the symmetry admixing is negligible. $\Psi_A^{(0)}$ and $\Psi_F^{(0)}$, eigenfunctions of $H^{(0)}$, have different symmetries. The selection rules are such that the transition $\Psi_A^{(0)} \rightarrow \Psi_D$ is allowed, while $\Psi_F^{(0)} \rightarrow \Psi_D$ is forbidden. The effect of $H^{(1)}$, which has no diagonal matrix elements in the $\Psi^{(0)}$ states, is to "rotate" the states through a small angle ϕ (illus. *b*), thereby admixing the symmetries, generating a weak transition for Ψ_F, and slightly shifting the energies. The ratio of the transition strengths S determines ϕ, which in turn fixes the admixing intensity and the level shifts. Diagonalizing the 2×2 matrix gives $S_F/S_A = \tan^2 \phi$; $(\Psi_F^{(0)} H^{(1)} \Psi_A^{(0)}) = (E_F - E_A) \sin \phi \cos \phi$; $(E_F - E_F^{(0)}) = -(E_A - E_A^0) = (E_F - E_A) \sin^2 \phi$. The perturbation solution (small ϕ) is often adequate. *See* PERTURBATION (QUANTUM MECHANICS).

Multiple-photon transitions. Because the M1 operator has no radial dependence in the long-wavelength limit, the single-photon transition between the $2S_{1/2}$ and $1S_{1/2}$ states in hydrogen is strongly inhibited; the lifetime should be on the order of 10^5 s. The very low-energy E1 transition to $2P_{1/2}^-$ (which lies below $2S_{1/2}^+$ by the Lamb shift energy), followed by E1 to ground, would take on the order of 10^8 s. The dominant process instead, with a lifetime on the order of 10^{-1} s, is the simultaneous emission of two photons, which share the energy to produce a continuous photon spectrum. In nuclei the two-photon $0^+ \rightarrow 0^+$ transition has also been observed, as well as E0 deexcitation by ejection of a penetrating atomic electron ("complete internal conversion") and by the closely related emission of an e^+-e^- pair. All of these processes are forbidden for $0^+ \leftrightarrow 0^-$, which has not been observed. *See* LASER SPECTROSCOPY; NONLINEAR OPTICS; RYDBERG CONSTANT.

LS-forbidden transitions. Early examples discovered in optical astronomy involve $6s6p^3P_0 \leftarrow 6s^2\ ^1S_0 \rightarrow 6s6p^3P_2$ in HgI. The first is absolutely forbidden, while the second would require that highly improbable M2, with wave-function corrections (since $s \rightarrow p$ would require E1). But the decoupling effect of the hyperfine interaction between valence electrons

and the nuclear magnetic moment generates a small $^3P_1$ amplitude in the $^3P_0$ state (the total angular momentum being $F = {}^1/_2$ in each case) and thereby, as in the illustration, an allowed but greatly retarded $^1S_0 \rightarrow {}^3P_1$ transition. A similar process occurs for the second case. Astronomical observation of such forbidden lines indicates that particle densities in the source are low; otherwise deexcitation would occur via collisions.

Spherical shell-model orbits in nuclei. The process of adding a neutron in orbit (l, j) to a nucleus, $A + n(l, j) \rightarrow B$, is realizable in (d, p) "direct" reactions, whose angular distribution determines l and whose magnitude gives the intrinsic neutron-capture probability. If one of the nuclear states involved is 0^+, the j value must equal the angular momentum of the other. For example, if A is describable as $(j^{2n})_{J=0^+}$, there should be large cross sections to $(j^{2n+1})_{J=j}$ and $((j^{2n})_0 j')_{J=j'}$, in which the particles in orbit j are themselves coupled to zero. Forbidden transitions to $(j^{2n+1})_{J=j'\neq j}$ are, however, often observed. They arise via admixtures (on the order of a few percent) of $((j^{2n})_0 j')$ in the final state which, by the analysis associated with the illustration, give measures of the orbital "purity." *See* NUCLEAR REACTION.

K selection rules in deformed nuclei. There is a J_z splitting of single-particle orbits in many heavy nuclei; this is produced by the large permanent quadrupole deformation of the nucleus which gives rise to a noncentral single-particle potential. K labels the angular-momentum component along the symmetry axis, whose rotation then generates from each intrinsic state a band of levels with $J \geq K$, all with the same parity. If two levels belong to bands with very different K values (specifically if $\Delta = |K_1 - K_2| - j$ is large, where the multipolarity j follows from the J^π values), electromagnetic transitions between them will be very highly forbidden. This is so formally because the j-multipole transition operator can only generate $|\Delta K| \leq j$, and more physically because the very different rotations involved in the two states generate large differences between all the corresponding orbits which play a role in the rotation. Empirically one encounters Δ values as high as 8 and finds retardations on the order of $10^{2\Delta}$, the process itself occurring via Coriolis-type admixtures. Similar multiparticle forbiddenness would obtain for transitions between states built on the two potential minima observed in fission studies for many nuclei. *See* NUCLEAR FISSION.

Molecular selection rules. The example of K selection rules in deformed nuclei involves both single-particle and rotational motion in these nuclei. Vibrational motion, especially quadrupole, can be found also. The classification of the motions derives from molecular physics and is most developed in that domain. To the extent that the molecular motions can be well separated, the angular momentum transferred in a transition must come from one or another of them. Along with the selection rules implied by this (the rule for harmonic vibrations is $n_v \rightarrow n_v \pm 1$), the various coupling possibilities and any special symmetries displayed by the intrinsic structure must be considered. Let N, Λ, Σ be the projection onto the symmetry axis (the line joining the two nuclei for the diatomic molecules to which the discussion is now restricted) of the collective rotational, the electronic orbital, and the electronic spin angular momenta. Λ and Σ may combine to Ω, the total intrinsic angular momentum along the symmetry axis; then the rotation gives $J = \Omega, \Omega + 1, \ldots$ (Hund's case A). Alternatively (case B), $\Lambda + N = K$, which combines with S to give $\mathbf{J} = \mathbf{S} + \mathbf{K}$, so that $|K - S| \leq J \leq K + S$. In both cases the supplementary rules for E1 (with natural extensions to other multipoles) that $\Delta\Lambda = 0, \pm 1$ and $\Delta S = 0$ are valid, while for case A, $\Delta\Sigma = 0$, $\Delta\Omega = 0 \pm 1$, and for case B, $\Delta\mathbf{K} = 1$, these being valid if the spin-orbit interactions are weak.

The special symmetries of the intrinsic structure refer to its behavior under a reflection through the symmetry center and through a plane containing the symmetry axis. The first operation, $\mathscr{P}$, defines the intrinsic parity, labeled as g (for $\pi = +1$) and u ($\pi = -1$). The second, $\mathscr{S}$, defines a symmetry $\pm$; in nuclear physics one commonly uses $\mathscr{R} = \mathscr{P}\mathscr{S}$, with eigenvalue $r = \pm 1$, which generates a rotation through π radians about an axis normal to the symmetry axis. For homonuclear molecules the interchange of the two nuclei defines states as symmetric (s) or antisymmetric (a). Then for E1 the rules $g \leftrightarrow u$, $+ \leftrightarrow -$ are valid, and with homonuclear molecules $s \rightarrow s$, $a \rightarrow a$. *See* MOLECULAR STRUCTURE AND SPECTRA.

Further symmetries. The isospin symmetry of the elementary particles is almost conserved, being broken by electromagnetic and weak interactions. It is described by the group SU(2), of unimodular unitary transformations in two dimensions. Since the SU(2) algebra is identical with that of the angular momentum SO(3), isospin behaves like angular momentum with its three generators $\mathbf{T}$ replacing $\mathbf{J}$. For a nucleon $t = {}^1/_2$, the values $t_z = {}^1/_2, -{}^1/_2$ may be assigned for protons (p) and neutrons (n), respectively. Then, for nuclear states, $T_z = (Z - N)/2$ (where Z is the atomic number and N is the neutron number) and T ($\geq |T_z|$) is an almost good quantum number. The argument associated with the illustration can often be used to measure the isospin admixing; for example, with the pair of decays $^{15}\mathrm{N}$ ($T = {}^1/_2, {}^3/_2$) $\rightarrow {}^{14}\mathrm{N}$ ($T = 0 + n$), the second of which is forbidden by the selection rule $\Delta T = {}^1/_2$, one finds about 4% admixing. Since the nucleon charge is $|e|\{{}^1/_2 + t_z\}$, the electromagnetic transition operators split into an isoscalar part (for which $\Delta\mathbf{T} = 0$) and an isovector ($\Delta\mathbf{T} = 1$). Since the isoscalar part of the (long-wavelength) E1 operator

$$|e|\left\{{}^1/_2 \sum_i \mathbf{r}(i) + \sum_i t_z(i)\mathbf{r}(i)\right\}$$

is ineffective, being proportional to the center-of-mass vector, and the inversion rule described earlier forbids isovector transitions in self-conjugate nuclei ($T_z = 0$, $N = Z$), it follows that E1 transitions between states of the same T are forbidden in such nuclei.

The isospin group is a subgroup of SU(3) which defines a more complex fundamental symmetry of the elementary particles. Two of its eight generators commute, giving two additive quantum numbers, T_z and strangeness S' (or, equivalently, charge and hypercharge). The strangeness is conserved ($\Delta S' = 0$) for strong and electromagnetic, but not for weak, interactions. The selection rules and combination laws for SU(3) and its many extensions, and the quark-structure ideas underlying them, correlate an enormous amount of information and make many predictions about the elementary particles. *See* BARYON; ELEMENTARY PARTICLE; MESON; QUARKS; UNITARY SYMMETRY.

The fundamental beta-decay transition is $n \rightarrow p + e^- + \overline{\nu}$. For the Gamow-Teller decay mode (electron and antineutrino in a triplet state) and the Fermi mode (singlet), the transition operators are respectively components of

$$\sum_i \mathbf{s}(i)\, \mathbf{t}(i)$$

and of $\mathbf{T}$. When these decays are realized in a nucleus, in which also $p \rightarrow n + e^+ + \nu$ may occur, the rules $\Delta\pi = 0$, $\Delta\mathbf{T} = 1$, $\Delta T_z = \pm 1$ clearly hold, with $\Delta\mathbf{J} = 1, 0$ for Gamow-Teller and Fermi, respectively. But for Fermi decay there is also the very strong rule (which includes $\Delta T = 0$) that the decay takes place only to a single state, the isobaric analog, which differs from the parent only in T_z, the transition rate being then easily calculable. This special feature, analogous to that for M1 transitions in *LS* coupling, arises from the fact that the transition operator is a generator of an (almost) good symmetry. Besides superallowed Fermi transitions between $J = 0$ analog states (forbidden for Gamow-Teller), there are also forbidden Fermi decays between nonanalog states which, as in the illustration, arise from isospin admixing and have been used to give measures for it. Special rules, by no means as well satisfied, arise also for Gamow-Teller decays from the fact that the 15 components of the basic transition operators are the generators of an SU(4) group whose symmetry is fairly good in light nuclei. (For these same nuclei a different SU(3) realization than above gives a model for rotational bands.) Besides the allowed transitions, there is also a hierarchy of forbidden transitions in nuclei which, through recoil (retardation) effects, involve the orbital angular momentum and have therefore less restrictive selection rules. *See* NEUTRINO; RADIOACTIVITY.

A great variety of other groups have been introduced to define relevant symmetries for atoms, molecules, nuclei, and elementary particles. They all have their own selection rules, representing one aspect of the symmetries of nature. J. B. French

Bibliography. P. F. Bernath, *Spectra of Atoms and Molecules*, 1995; A. Bohr and B. R. Mottelson, *Nuclear Structure*, vol. 2, 1975; E. U. Condon and H. Odabasi, *Atomic Structure*, 1980; R. D. Cowan, *The Theory of Atomic Structure and Spectra*, 1981; K. S. Krane, *Introductory Nuclear Physics,* 1987.

Selectivity

The ability of a radio receiver to separate a desired signal frequency from other signal frequencies, some of which may differ only slightly from the desired value. Selectivity is achieved by using tuned circuits that are sharply peaked and by increasing the number of tuned circuits. With a sharply peaked circuit, the output voltage falls off rapidly for frequencies lower or higher than that to which the circuit is tuned. *See* Q (ELECTRICITY); RADIO RECEIVER; RESONANCE (ALTERNATING-CURRENT CIRCUITS). John Markus

Selenium

A chemical element, Se, atomic number 34, atomic weight 78.96. The properties of this element are similar to those of tellurium. *See* PERIODIC TABLE; TELLURIUM.

1												13	14	15	16	17	18
1 H	2																2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

Selenium burns in air with a blue flame to give selenium dioxide, SeO_2. The element also reacts directly with a variety of metals and nonmetals, including hydrogen and the halogens. Nonoxidizing acids fail to react with selenium, but nitric acid, concentrated sulfuric acid, and strong alkali hydroxides dissolve the element. The only important compound of selenium with hydrogen is hydrogen selenide, H_2Se, a colorless flammable gas possessing a distinctly unpleasant odor, and a toxicity greater and a thermal stability less than that of hydrogen sulfide. Selenium oxyhalide, $SeOCl_2$, is a colorless liquid widely used as a nonaqueous solvent. The oxybromide, $SeOBr_2$, is an orange solid having chemical properties similar to those of $SeOCl_2$. The oxyflouride, $SeOF_2$, a colorless liquid with a pungent smell, reacts with water, glass, and silicon, and also forms additional compounds. Compounds in which C-Se bonds appear are numerous and vary from the simple selenols, RSeH, to molecules exhibiting biological activity such as selenoamino acids and selenopeptides. *See* ORGANOSELENIUM COMPOUND.

The abundance of this widely distributed element in the Earth's crust is estimated to be about 7×10^{-5}% by weight, occurring as the selenides of heavy elements and to a limited extent as the free element in association with elementary sulfur. Examples

of the variety of selenide minerals are berzelianite (Cu_2Se), eucairite (AgCuSe), and jermoite [$As(S,Se)_2$]. Selenium minerals do not occur in sufficient quantity to be useful as commercial sources of the element.

Major uses of selenium include the photocopying process of xerography, which depends on the light sensitivity of thin films of amorphous selenium, the decolorization of glasses tinted by the presence of iron compounds, and use as a pigment in plastics, paints, enamels, glass, ceramics, and inks. Selenium is also employed in photographic exposure meters and solar cells and as a metallurgical additive to improve the machinability of certain steels. Minor uses include application as a nutritional additive for numerous animal species, use in photographic toning, metal-finishing operations, metal plating, high-temperature lubricants, and as catalytic agents, particularly in the isomerization of certain petroleum products. John W. George

The biological importance of selenium is well established, as all classes of organisms metabolize selenium. In humans and other mammals, serious diseases arise from either excessive or insufficient dietary selenium. The toxic effects of selenium have long been known, particularly for grazing animals. In soils with high selenium content, some plants accumulate large amounts of selenium. Animals that ingest these selenium-accumulating plants develop severe toxic reactions. *See* SOIL CHEMISTRY.

Although toxic at high levels, selenium is an essential micronutrient for mammalian species. The accepted minimum daily requirement of selenium for adult humans is 70 micrograms. Many types of food provide selenium, particularly seafood, meats, grains, and the onion family. Mammals and birds require selenium for production of the enzyme glutathione peroxidase, which protects against oxidation-induced cancers. Other seleno-proteins of unknown function are found in mammalian blood, various tissues, and spermatozoa. *See* AMINO ACIDS.

Milton J. Axley; Thressa G. Stadtman

Bibliography. M. A. Brook, *Silicon in Organic, Organometallic and Polymer Chemistry*, Wiley-Interscience, 1999; F. A. Cotton et al., *Advanced Inorganic Chemistry*, 6th ed., Wiley-Interscience, 1999; P. Jutzi and U. Schubert (eds.), *Silicon Chemistry: From the Atom to Extended Systems*, Wiley, 2003; U. Schubert (ed.), *Silicon Chemistry*, Springer, 1999.

Seligeriales

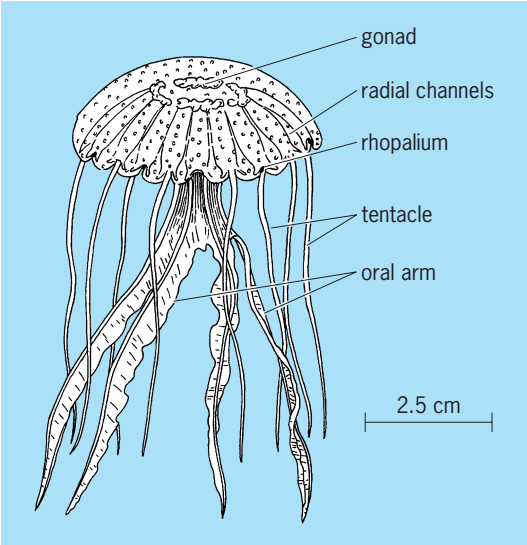
An order of the true mosses (Bryidae), consisting of one family and five or six genera. The plants grow on rocks, and may be exceedingly small and gregarious to moderate in size and tufted. Though the gametophytes resemble the Dicranales, the double structure of the peristome is distinctive.

The stems are erect and simple or forked. The linear to lance-subulate leaves have a single costa which often fills the subula, and short, smooth cells. The operculate capsules are terminal, and the peristome is usually present; though apparently single, it consists of an endostome of 16 narrow, smooth or rarely papillose teeth often cleft at the tips, the outer peristome being reduced to a membrane adherent to the endostome. The calyptra is generally cucullate. Chromosome numbers of 11, 12, 13, and 14 are known. *See* BRYIDAE; BRYOPHYTA; BRYOPSIDA; DICRANALES.

Howard Crum

Semaeostomeae

An order of the class Scyphozoa including most of the common medusae, such as *Aurelia aurita*. The umbrella of these medusae is more flat than high and is usually domelike. The margin of the umbrella is divided into many lappets. Sensory organs are situated between the lappets. The tentacles arise between the lappets in *Pelagia* (see **illus.**), on the exumbrella in



***Pelagia*, in the order Semaeostomeae. (*After L. Hyman, The Invertebrates, vol. 1, McGraw-Hill, 1940*)**

Aurelia, and on the subumbrella in *Cyanea*. They are generally well developed and very long except in a few forms such as *Aurelia*. The oral arms are well developed and either curtainlike, in *Cyanea*, or leaflike, in *Aurelia*. The stomach is cruciform, and several radial pockets or many radial canals which are sometimes connected with each other to form a network issue from it. The life history of this group shows the typical alternation of generations, with forms passing through the stages of planula, scyphopolyp, strobila, and ephyra. The Semaeostomeae are distributed mostly in temperate zones and are generally coastal forms with a few exceptions such as *Pelagia* and *Sanderia*. Some of them, such as *Dactylometra* and *Sanderia*, are known to have violent poison on their tentacles. This poison is not only harmful to the skin but also to the nerves, which are paralyzed temporarily. *Cyanea* is used as bait in fishing. Several fossils of this group were found in the strata of the Jurassic Period. *See* SCYPHOZOA.

Tohru Uchida

Semiconductor

A solid crystalline material whose electrical conductivity is intermediate between that of a metal and an insulator. Semiconductors exhibit conduction properties that may be temperature-dependent, permitting their use as thermistors (temperature-dependent resistors), or voltage-dependent, as in varistors. By making suitable contacts to a semiconductor or by making the material suitably inhomogeneous, electrical rectification and amplification can be obtained. Semiconductor devices, rectifiers, and transistors have replaced vacuum tubes almost completely in low-power electronics, making it possible to save volume and power consumption by orders of magnitude. In the form of integrated circuits, they are vital for complicated systems. The optical properties of a semiconductor are important for the understanding and the application of the material. Photodiodes, photoconductive detectors of radiation, injection lasers, light-emitting diodes, solar-energy conversion cells, and so forth are examples of the wide variety of optoelectronic devices. *See* INTEGRATED CIRCUITS; LASER; LIGHT-EMITTING DIODE; PHOTODIODE; PHOTOELECTRIC DEVICES; SEMICONDUCTOR DIODE; SEMICONDUCTOR RECTIFIER; THERMISTOR; TRANSISTOR; VARISTOR.

Conduction in Semiconductors

The electrical conductivity of semiconductors ranges from about 10^3 to 10^{-9} ohm^{-1} cm^{-1}, as compared with a maximum conductivity of 10^7 for good conductors and a minimum conductivity of 10^{-17} ohm^{-1} cm^{-1} for good insulators. *See* ELECTRIC INSULATOR; ELECTRICAL CONDUCTIVITY OF METALS.

The electric current is usually due only to the motion of electrons, although under some conditions, such as very high temperatures, the motion of ions may be important. The basic distinction between conduction in metals and in semiconductors is made by considering the energy bands occupied by the conduction electrons. *See* IONIC CRYSTALS.

A crystalline solid consists of a large number of atoms brought together into a regular array called a crystal lattice. The electrons of an atom can each have certain energies, so-called energy levels, as predicted by quantum theory. Because the atoms of the crystal are in close proximity, the electron orbits around different atoms overlap to some extent, and the electrons interact with each other; consequently the sharp, well-separated energy levels of the individual electrons actually spread out into energy bands. Each energy band is a quasicontinuous group of closely spaced energy levels. *See* BAND THEORY OF SOLIDS.

At absolute zero temperature, the electrons occupy the lowest possible energy levels, with the restriction that at most two electrons with opposite spin may be in the same energy level. In semiconductors and insulators, there are just enough electrons to fill completely a number of energy bands, leaving the rest of the energy bands empty. The highest filled energy band is called the valence band. The next higher band, which is empty at absolute zero temperature, is called the conduction band. The conduction band is separated from the valence band by an energy gap which is an important characteristic of the semiconductor. In metals, the highest energy band that is occupied by the electrons is only partially filled. This condition exists either because the number of electrons is not just right to fill an integral number of energy bands or because the highest occupied energy band overlaps the next higher band without an intervening energy gap. The electrons in a partially filled band may acquire a small amount of energy from an applied electric field by going to the higher levels in the same band. The electrons are accelerated in a direction opposite to the field and thereby constitute an electric current. In semiconductors and insulators, the electrons are found only in completely filled bands, at low temperatures. In order to increase the energy of the electrons, it is necessary to raise electrons from the valence band to the conduction band across the energy gap. The electric fields normally encountered are not large enough to accomplish this with appreciable probability. At sufficiently high temperatures, depending on the magnitude of the energy gap, a significant number of valence electrons gain enough energy thermally to be raised to the conduction band. These electrons in an unfilled band can easily participate in conduction. Furthermore, there is now a corresponding number of vacancies in the electron population of the valence band. These vacancies, or holes as they are called, have the effect of carriers of positive charge, by means of which the valence band makes a contribution to the conduction of the crystal. *See* HOLE STATES IN SOLIDS.

The type of charge carrier, electron or hole, that is in largest concentration in a material is sometimes called the majority carrier and the type in smallest concentration the minority carrier. The majority carriers are primarily responsible for the conduction properties of the material. Although the minority carriers play a minor role in electrical conductivity, they can be important in rectification and transistor actions in a semiconductor.

Electron distribution. The probability f for an energy level E to be occupied by an electron is given by the Fermi-Dirac distribution function (1), where k

$$f = \left[1 + \exp\left(\frac{E - W}{kT}\right)\right]^{-1} \tag{1}$$

is the Boltzmann constant and T the absolute temperature. The parameter W is the Fermi energy level; an energy level at W has a probability of 1/2 to be occupied by an electron. The Fermi level is determined by the distribution of energy levels and the total number of electrons. *See* FERMI-DIRAC STATISTICS.

In a semiconductor, the number of conduction electrons is normally small compared with the number of energy levels in the conduction band, and the probability for any energy level to be occupied is small. Under such a condition, the concentration

of conduction electrons is given by Eq. (2), where

$$N_n = \frac{2}{h^3}(2\pi m_n kT)^{3/2} \exp\left(\frac{W - E_c}{kT}\right) \qquad (2)$$

h is Planck's constant, E_c is the lowest energy of the conduction band, and m_n is called the effective mass of conduction electrons. The effective mass is used in place of the actual mass to correct the coefficient in the equation and to bring the results in line with experimental observations. This correction is necessary because the theory leading to these equations is based upon electrons moving in a field free space, which is not the exact picture. The electrostatic Coulomb potential throughout the crystal is varying in a periodic manner, the variation being due to the electric fields around the atomic centers. The concentration of holes in the valence band is given by Eq. (3), where m_p is the effective mass of a hole

$$N_p = \frac{2}{h^3}(2\pi m_p kT)^{3/2} \exp\left(\frac{E_v - W}{kT}\right) \qquad (3)$$

and E_v is the highest energy of the valence band.

Mobility of carriers. The velocity acquired by charge carriers per unit strength of applied electric field is called the mobility of the carriers. The velocity in question is the so-called drift velocity in the direction of the force exerted on the carriers by the applied field. It is added to the random thermal velocity. In semiconductors the carrier mobility normally ranges from 10^2 to 10^5 $cm^2/(s)(V)$. A material's conductivity is the product of the charge, the mobility, and the carrier concentration.

Electrons in a perfectly periodic potential field can be accelerated freely. Impurities, physical defects in the structure, and thermal vibrations of the atoms disturb the periodicity of the potential field in the crystal, thereby scattering the moving carriers. It is the resistance produced by this scattering that limits the carriers to only a drift velocity under the steady force of an applied field.

Intrinsic semiconductors. A semiconductor in which the concentration of charge carriers is characteristic of the material itself rather than of the content of impurities and structural defects of the crystal is called an intrinsic semiconductor. Electrons in the conduction band and holes in the valence band are created by thermal excitation of electrons from the valence to the conduction band. Thus an intrinsic semiconductor has equal concentrations of electrons and holes. The intrinsic carrier concentration N_i is determined by Eq. (4), where E_g is the energy

$$N_i = \frac{2}{h^3}(2\pi kT)^{3/2}(m_n m_p)^{3/4} \exp\left(-\frac{E_g}{2kT}\right) \qquad (4)$$

gap. The carrier concentration, and hence the conductivity, is very sensitive to temperature and depends strongly on the energy gap. The energy gap ranges from a fraction of 1 eV to several electronvolts. A material must have a large energy gap to be an insulator.

Extrinsic semiconductors. Typical semiconductor crystals such as germanium and silicon are formed by an ordered bonding of the individual atoms to form the crystal structure. The bonding is attributed to the valence electrons which pair up with valence electrons of adjacent atoms to form so-called shared pair or covalent bonds. These materials are all of the quadrivalent type; that is, each atom contains four valence electrons, all of which are used in forming the crystal bonds. *See* CRYSTAL STRUCTURE.

Atoms having a valence of +3 or +5 can be added to a pure or intrinsic semiconductor material with the result that the +3 atoms will give rise to an unsatisfied bond with one of the valence electrons of the semiconductor atoms, and +5 atoms will result in an extra or free electron that is not required in the bond structure. Electrically, the +3 impurities add holes and the +5 impurities add electrons. They are called acceptor and donor impurities, respectively. Typical valence +3 impurities used are boron, aluminum, indium, and gallium. Valence +5 impurities used are arsenic, antimony, and phosphorus.

Semiconductor material "doped" or "poisoned" by valence +3 acceptor impurities is termed *p*-type, whereas material doped by valence +5 donor material is termed *n*-type. The names are derived from the fact that the holes introduced are considered to carry positive charges and the electrons negative charges. The number of electrons in the energy bands of the crystal is increased by the presence of donor impurities and decreased by the presence of acceptor impurities. Let N be the concentration of electrons in the conduction band and let P be the hole concentration in the valence band. For a given semiconductor, the relation $NP = N_i^2$ holds, independent of the presence of impurities. The effect of donor impurities tends to make N larger than P, since the extra electrons given by the donors will be found in the conduction band even in the absence of any holes in the valence band. Acceptor impurities have the opposite effect, making P larger than N. *See* ACCEPTOR ATOM; DONOR ATOM.

At sufficiently high temperatures, the intrinsic carrier concentration becomes so large that the effect of a fixed amount of impurity atoms in the crystal is comparatively small and the semiconductor becomes intrinsic. When the carrier concentration is predominantly determined by the impurity content, the conduction of the material is said to be extrinsic. There may be a range of temperature within which the impurity atoms in the material are practically all ionized; that is, they supply a maximum number of carriers. Within this temperature range, the so-called exhaustion range, the carrier concentration remains nearly constant. At sufficiently low temperatures, the electrons or holes that are supplied by the impurities become bound to the impurity atoms. The concentration of conduction carriers will then decrease rapidly with decreasing temperature, according to either $\exp(-E_i/kT)$ or $\exp(-E_i/2kT)$, where E_i is the ionization energy of the dominant impurity.

Physical defects in the crystal structure may have similar effects as donor or acceptor impurities. They can also give rise to extrinsic conductivity.

An isoelectronic impurity, that is, an atom which has the same number of valence electrons as the host atom, does not bind individual carriers as strongly as a donor or an acceptor impurity. However, it may show an appreciable binding for electron hole pairs, excitons, and thereby have important effects on the properties. An example is nitrogen substituting for phosphorus in gallium phosphide; the impurity affects the luminescence of the material.

Hall effect. Whether a given sample of semiconductor material is *n*- or *p*-type can be determined by observing the Hall effect. If an electric current is caused to flow through a sample of semiconductor material and a magnetic field is applied in a direction perpendicular to the current, the charge carriers are crowded to one side of the sample, giving rise to an electric field perpendicular to both the current and the magnetic field. This development of a transverse electric field is known as the Hall effect. The field is directed in one or the opposite direction depending on the sign of the charge of the carrier. *See* HALL EFFECT.

The magnitude of the Hall effect gives an estimate of the carrier concentration. The ratio of the transverse electric field strength to the product of the current and the magnetic field strength is called the Hall coefficient, and its magnitude is inversely proportional to the carrier concentration. The coefficient of proportionality involves a factor which depends on the energy distribution of the carriers and the way in which the carriers are scattered in their motion. However, the value of this factor normally does not differ from unity by more than a factor of 2. The situation is more complicated when more than one type of carrier is important for the conduction. The Hall coefficient then depends on the concentrations of the various types of carriers and their relative mobilities.

The product of the Hall coefficient and the conductivity is proportional to the mobility of the carriers when one type of carrier is dominant. The proportionality involves the same factor which is contained in the relationship between the Hall coefficient and the carrier concentration. The value obtained by taking this factor to be unity is referred to as the Hall mobility.

Materials and Their Preparation

The group of chemical elements which are semiconductors includes germanium, silicon, gray (crystalline) tin, selenium, tellurium, and boron.

Elemental semiconductors. Germanium, silicon, and gray tin belong to group 14 (group IV) of the periodic table and have crystal structures similar to that of diamond. Germanium and silicon are two of the best-known semiconductors. They are used extensively in devices such as rectifiers and transistors. Gray tin is a form of tin which is stable below 13°C (55°F). White tin, which is stable at higher temperatures, is metallic. Gray tin has a small energy gap and a rather large intrinsic conductivity, about 5×10^3 ohm^{-1} cm^{-1} at room temperature. The *n*-type and *p*-type gray tins can be obtained by adding aluminum and antimony, respectively.

Selenium and tellurium both have a similar structure, consisting of spiral chains located at the corners and centers of hexagons. The structure gives rise to anisotropy of the properties of single crystals; for example, the electrical resistivity of tellurium along the direction of the chains is about one-half the resistivity perpendicular to this direction. Selenium has been widely used in the manufacture of rectifiers and photocells.

Semiconducting compounds. A large number of compounds are known to be semiconductors. Copper(I) oxide (Cu_2O) and mercury(II) indium telluride ($HgIn_2Te_4$) are examples of binary and ternary compounds. The series zinc sulfide (ZnS), zinc selenide (ZnSe), and zinc telluride (ZnTe), and the series zinc selenide (ZnSe), cadmium selenide (CdSe), and mercury(II) selenide (HgSe) are examples of binary compounds consisting of a given element in combinations with various elements of another column in the periodic table. The series magnesium antimonide (Mg_2Sb_2), magnesium telluride (MgTe), and magnesium iodide (MgI_2) is an example of compounds formed by a given element with elements of various other columns in the periodic table. *See* PERIODIC TABLE.

A group of semiconducting compounds of the simple type AB consists of elements from columns symmetrically placed with respect to column 14 of the periodic table. Indium antimonide (InSb), cadmium telluride (CdTe), and silver iodide (AgI) are examples of III–V, II–IV, and I–VI compounds, respectively. The various III–V compounds are being studied extensively, and many practical applications have been found for these materials. Some of these compounds have the highest carrier mobilities known for semiconductors. The compounds have zincblende crystal structure which is geometrically similar to the diamond structure possessed by the elemental semiconductors, germanium and silicon, of column 14, except that the four nearest neighbors of each atom are atoms of the other kind. The II–VI compounds, zinc sulfide (ZnS) and cadmium sulfide (CdS), are used in photoconductive devices. Zinc sulfide is also used as a luminescent material. *See* LUMINESCENCE; PHOTOCONDUCTIVITY.

Binary compounds of the group lead sulfide (PbS), lead selenide (PbSe), and lead telluride (PbTe) are sensitive in photoconductivity and are used as detectors of infrared radiation. The compounds, bismuth telluride (Bi_2Te_3) and bismuth selenide (Bi_2Se_3), consisting of heavy atoms, are found to be good materials for thermocouples used for refrigeration or for conversion of heat to electrical energy. *See* THERMOELECTRICITY.

The metal oxides usually have large energy gaps. Thus pure oxides are usually insulators of high resistivity. However, it may be possible to introduce into some of the oxides impurities of low ionization energies and thus obtain relatively good extrinsic conduction. Copper(I) oxide (Cu_2O) was one of the first semiconductors used for rectifiers and photocells: extrinsic *p*-type conduction is obtained by producing

an excess of oxygen over the stoichiometric composition, that is, the 2-to-1 ratio of copper atoms to oxygen atoms. A number of oxide semiconductors can be obtained by replacing some of the normal metal atoms with metal atoms of one more or less valency. The method is called controlled valence. An example of such a semiconductor is nickel oxide containing lithium.

Some compounds with rare-earth or transition-metal ions in their composition, such as EuTe and NiS_2, are semiconductors with magnetic properties. Another interesting type of semiconductor is characterized by layered structures. The interaction within a layer is significantly stronger than that between layers. A number of semiconductors of this type are known, such as PbI_2, GaSe, and various transition-metal dichalcogenides such as $SnSe_2$ and MoS_2.

Preparation of materials. The properties of semiconductors are extremely sensitive to the presence of impurities. It is therefore desirable to start with the purest available materials and to introduce a controlled amount of the desired impurity. The zone-refining method is often used for further purification of obtainable materials. The floating zone technique can be used, if feasible, to prevent any contamination of molten material by contact with crucible. *See* ZONE REFINING.

For basic studies as well as for many practical applications, it is desirable to use single crystals. Various methods are used for growing crystals of different materials. For many semiconductors, including germanium, silicon, and the III–V compounds, the Czochralski method is commonly used. The method of condensation from the vapor phase is used to grow crystals of a number of semiconductors, for instance, selenium and zinc sulfide. For materials of high melting points, such as various metal oxides, the flame fusion or Vernonil method may be used. *See* CRYSTAL GROWTH.

The introduction of impurities, or doping, can be accomplished by simply adding the desired quantity to the melt from which the crystal is grown. Normally, the impurity has a small segregation coefficient, which is the ratio of equilibrium concentrations in the solid and the liquid phases of the material. In order to obtain a desired impurity content in the crystal, the amount added to the melt must give an appropriately larger concentration in the liquid. When the amount to be added is very small, a preliminary ingot is often made with a larger content of the doping agent; a small slice of the ingot is then used to dope the next melt accurately. Impurities which have large diffusion constants in the material can be introduced directly by holding the solid material at an elevated temperature while this material is in contact with the doping agent in the solid or the vapor phase.

A doping technique, ion implantation, has been developed and used extensively. The impurity is introduced into a layer of semiconductor by causing a controlled dose of highly accelerated impurity ions to impinge on the semiconductor. *See* ION IMPLANTATION.

An important subject of scientific and technological interest is amorphous semiconductors. In an amorphous substance the atomic arrangement has some short-range but no long-range order. The representative amorphous semiconductors are selenium, germanium, and silicon in their amorphous states, and arsenic and germanium chalcogenides, including such ternary systems as Ge-As-Te. Some amorphous semiconductors can be prepared by a suitable quenching procedure from the melt. Amorphous films can be obtained by vapor deposition.

Rectification in Semiconductors

In semiconductors, narrow layers can be produced which have abnormally high resistances. The resistance of such a layer is nonohmic; it may depend on the direction of current, thus giving rise to rectification. Rectification can also be obtained by putting a thin layer of semiconductor or insulator material between two conductors of different material.

Barrier layer. A narrow region in a semiconductor which has an abnormally high resistance is called a barrier layer. A barrier may exist at the contact of the semiconductor with another material, at a crystal boundary in the semiconductor, or at a free surface of the semiconductor. In the bulk of a semiconductor, even in a single crystal, barriers may be found as the result of a nonuniform distribution of impurities. The thickness of a barrier layer is small, usually 10^{-3} to 10^{-5} cm.

A barrier is usually associated with the existence of a space charge. In an intrinsic semiconductor, a region is electrically neutral if the concentration n of conduction electrons is equal to the concentration p of holes. Any deviation in the balance gives a space charge equal to $e(p - n)$, where e is the charge on an electron. In an extrinsic semiconductor, ionized donor atoms give a positive space charge and ionized acceptor atoms give a negative space charge. Let N_d and N_a be the concentrations of ionized donors and acceptors, respectively. The space charge is equal to $e(p - n + N_d - N_a)$.

A space charge is associated with a variation of potential. A drop in potential $-\Delta V$ increases the potential energy of an electron by $e\Delta V$; consequently every electronic energy level in the semiconductor is shifted by this amount. With a variation of potential, the electron concentration varies proportionately to $\exp(eV/kT)$ and the hole concentration varies as $\exp(-eV/kT)$. A space charge is obtained if the carriers, mainly the majority carriers, fail to balance the charge of the ionized impurities.

A conduction electron in a region where the potential is higher by ΔV must have an excess energy of $e\Delta V$ in order for it to have the minimum energy on reaching the low potential region. Electrons with less energy cannot pass over to the low potential region. Thus a potential variation presents a barrier to the flow of electrons from high to low potential regions. It also presents a barrier to the flow of holes from low to high potential regions.

Surface barrier. A thin layer of space charge and a resulting variation of potential may be produced

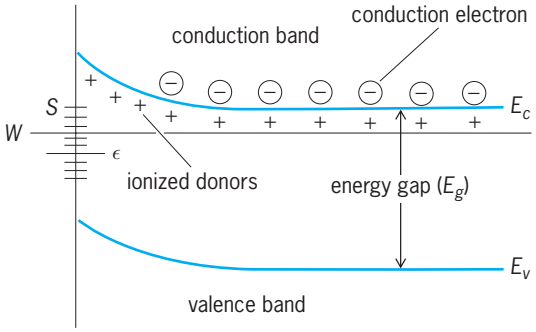


Fig. 1. Energy diagram of a surface barrier as employed in an *n*-type semiconductor.

at the surface of a semiconductor by the presence of surface states. Electrons in the surface states are bound to the vicinity of the surface, and the energy levels of surface states may lie within the energy gap. Surface states may arise from the adsorption of foreign atoms. Even a clean surface may introduce states which do not exist in the bulk material, simply by virtue of being the boundary of the crystal.

The surface is electrically neutral when the surface states are filled with electrons up to a certain energy level ϵ in the energy gap E_g, which is the energy difference between the bottom of the conduction band E_c and the top of the valence band E_v. If the Fermi level W in the bulk semiconductor lies higher in the energy gap, more surface states would be filled, giving the surface a negative charge. As a result the potential drops near the surface and the energy bands are raised for *n*-type material (**Fig. 1**). With the rise of the conduction band, the electron concentration is reduced and a positive space charge due to ionized donors is obtained. The amount of positive space charge is equal to the negative surface charge given by the electrons in the surface states between ϵ and the Fermi level.

Contact barrier. The difference between the potential energy E_o of an electron outside a material and the Fermi level in the material is called the work function of the material. **Figure 2** shows the energy diagram for a metal and a semiconductor, the work functions of which differ by eV. Upon connecting the two bodies electrically, charge is transferred between them so that the potential of the semiconductor is raised relative to that of the metal; that is, the electron energy levels in the semiconductor are lowered. Equilibrium is established when the Fermi level is the same in the two bodies. In this case, the metal is charged negatively and the semiconductor is charged positively. The negative charge on the metal is concentrated close to the surface, as is expected in good conductors. The positive charge on the semiconductor is divided between the increase of space charge in an extension of the barrier and the depopulation of some of the surface states. The charging of the semiconductor is brought about by a change of eV_2 in the barrier height ϕ. The sum of eV_2 and the potential energy variation eV_1 in the space between the two bodies is equal to the original difference eV between the work functions.

With decreasing separation between the two bodies, the division of eV will be in favor of eV_2. However, if there is a very large density of states, a small eV_2 gives a large surface charge on the semiconductor due to the depopulation of surface states.

It is possible that eV_2 is limited to a small value even at the smallest separation, of the order of an interatomic distance in solids. In such cases, the barrier height remains nearly equal to the value ϕ of the free surface, irrespective of the body in contact. This situation has been found in germanium and silicon rectifiers. Before the explanation was given by J. Bardeen, who postulated the existence of surface states, it had been assumed that the height of a contact barrier was equal to the difference of the work functions.

The understanding and the application of metal-semiconductor contacts have been extended to various kinds of contacts, such as that between different semiconductors, heterojunctions, and metal oxide semiconductor (MOS) junctions.

Single-carrier theory. The phenomenon of rectification at a crystal barrier can be described according to the role played by the carriers. Where the conduction property of the rectifying barrier is determined primarily by the majority carriers, the single-carrier theory is employed. Such cases are likely to be found in semiconductors with large energy gaps, for instance, oxide semiconductors. **Figure 3** shows the energy diagrams of metal-semiconductor contact rectifiers under conditions of equilibrium. The potential variation in the semiconductor is such as to reduce the majority carrier concentration near

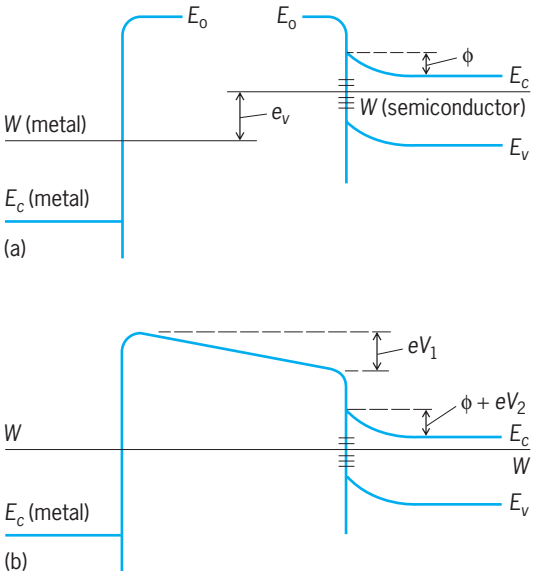


Fig. 2. Energy diagram for a metal (left) and an *n*-type semiconductor (right). E_0 is the potential energy of an electron outside the material, E_c is the energy at the bottom of the conduction band, and Ev is the energy at the top of the valence band. (*a*) Semiconductor and metal isolated. (*b*) Semiconductor and metal in electrical contact, $eV_1 + eV_2 = eV$.

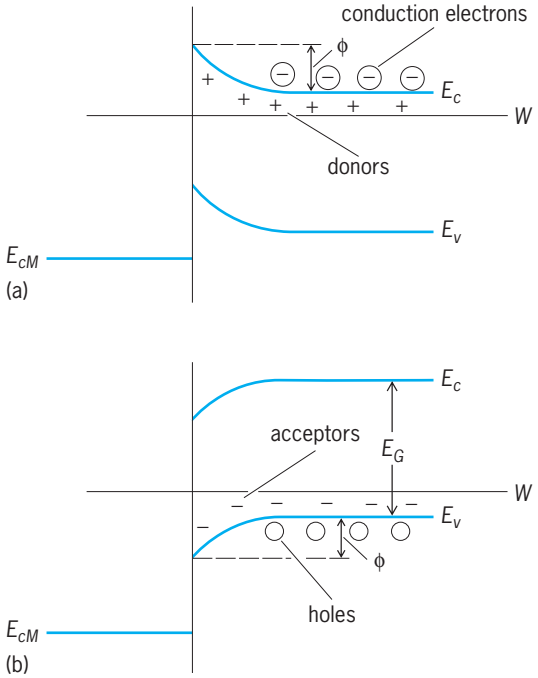


Fig. 3. Energy diagrams of a rectifying contact between a metal and a semiconductor: (*a*) *n*-type semiconductor; (*b*) *p*-type semiconductor.

the contact. If the energy bands were to fall in the case of an *n*-type semiconductor or to rise in the case of a *p*-type semiconductor, the majority carrier concentration would be enhanced near the contact, and the contact would not present a large and rectifying resistance. It is clear that in the cases shown in Fig. 3, the minority carrier concentration increases near the contact. However, if the energy gap is large, the minority carrier concentration is normally very small, and the role of minority carriers may be still negligible even if the concentration is increased.

Under equilibrium conditions, the number of carriers passing from one body to the other is balanced by the number of carriers crossing the contact in the opposite direction, and there is no net current. The carriers crossing the contact in either direction must have sufficient energies to pass over the peak of the barrier. The situations under applied voltages are shown in **Fig. 4** for the case of an *n*-type semiconductor. When the semiconductor is made positive, its energy bands are depressed and the height of the potential barrier is increased, as shown in Fig. 4*a*. Fewer electrons in the semiconductor will be able to cross over into the metal, whereas the flow of electrons across the contact from the metal side remains unchanged. Consequently, there is a net flow of electrons from the metal to the semiconductor. The flow of electrons from the metal side is the maximum net flow obtainable. With increasing voltage, the current saturates and the resistance becomes very high. Figure 4*b* shows the situation when the semiconductor is negative under the applied voltage. The energy bands in the semiconductor are raised.

The flow of electrons from the semiconductor to the metal is increased, since electrons of lower energy are able to go over the peak of the barrier. The result is a net flow of electrons from the semiconductor to the metal. There is no limit to the flow in this case. In fact, the electron current increases faster than the applied voltage because there are increasingly more electrons at lower energies. The resistance decreases, therefore, with increasing voltage. The direction of current for which the resistance is low is called the forward direction, while the opposite direction is called the reverse or blocking direction. A general expression for the current can be written in the form of Eq. (5), where j is the current density, n is

$$j = enC\left(\exp\frac{-\phi}{kT}\right)\left[\exp\left(\frac{eV}{kT}\right) - 1\right] \qquad (5)$$

the carrier concentration in the bulk of the semiconductor, ϕ is the barrier height, and V is the applied voltage taken as positive in the forward direction. The factor C depends on the theory appropriate for the particular case.

Diffusion theory. When there is a variation of carrier concentration, a motion of the carriers is produced by diffusion in addition to the drift determined by the mobility and the electric field. The transport of carriers by diffusion is proportional to the carrier concentration gradient and the diffusion constant. The diffusion constant is related to the mobility, and both are determined by the scattering suffered by

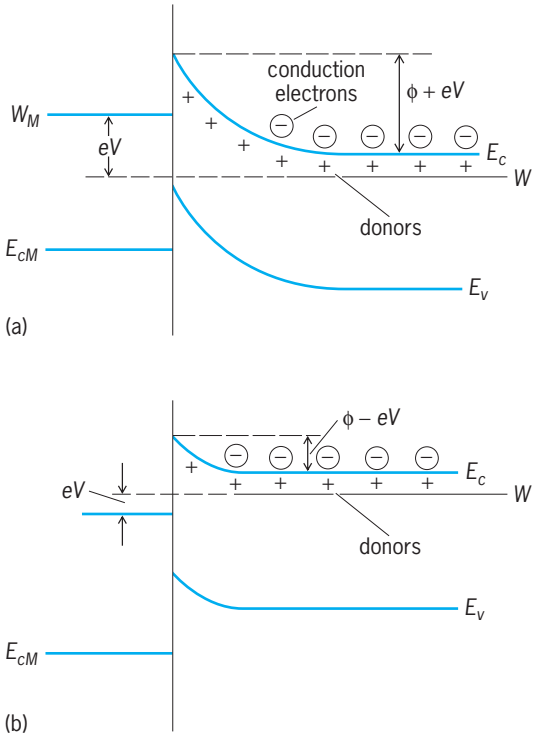


Fig. 4. Energy diagrams of a rectifying contact between a metal and an *n*-type semiconductor under an applied voltage *V*. (*a*) Positive semiconductor. There is a net flow of electrons from metal to semiconductor. (*b*) Negative semiconductor. There is a net flow of electrons from semiconductor to metal.

moving carriers. The average distance traveled by a carrier in its random thermal motion between collisions is called the mean free path. If barrier thickness is large compared to mean free path of carriers, motion of carriers in the barrier can be treated as drift and diffusion. This viewpoint is the basis of the diffusion theory of rectification. According to this theory, the factor C in Eq. (5) depends on the mobility and the electric field in the barrier.

Diode theory. When the barrier thickness is comparable to or smaller than the mean free path of the carriers, then the carriers cross the barrier without being scattered, much as in a vacuum tube diode.

According to this theory, the factor C in the rectifier equation is $v/4$, where v is the average thermal velocity of the carriers.

Tunneling theory. Instead of surmounting a potential barrier, carriers have a probability of penetrating through the barrier. The effect, called tunneling, becomes dominant if the barrier thickness is sufficiently small. This effect is important in many applications. *See* TUNNELING IN SOLIDS.

Two-carrier theory. Often the conduction through a rectifying barrier depends on both electron and hole carriers. An important case is the *pn* junction between *p*- and *n*-sections of a semiconductor material. Also, in metal-semiconductor rectifiers, the barrier presents an obstacle for the flow of majority carriers but not for the flow of minority carriers, and the latter may become equally or more important.

Rectification at pn junctions. A *pn* junction is the boundary between a *p*-type region and an *n*-type region of a semiconductor. When the impurity content varies, there is a variation of electron and hole concentrations. A variation of carrier concentrations is related to a shift of the energy bands relative to the constant Fermi level. This is brought about by a variation of the electrostatic potential which requires the existence of a space charge. If the impurity content changes greatly within a short distance, a large space charge is obtained within a narrow region. Such is the situation existing in a rectifying *pn* junction.

When a voltage is applied to make the *n*-region negative relative to the *p*-region, electrons flow from the *n*-region, where they are abundant, into the *p*-region. At the same time, holes flow from the *p*-region, where holes are abundant, into the *n*-region. The resistance is therefore relatively low. The direction of current in this case is forward. Clearly, the resistance will be high for current in the reverse direction.

With a current in the forward direction, electrons in the *n*-region and holes in the *p*-region flow toward the junction and there must be continuous hole-electron recombination in the neighborhood of the junction. The minority carrier concentration in each region is increased near the junction because of the influx of the carriers from the other region. This phenomenon is known as carrier injection. When there is a current in the reverse direction, there must be a continuous generation of holes and electrons in the neighborhood of the junction, from which electrons flow out into the *n*-region and holes flow out into the *p*-region. Thus current through a *pn* junction is controlled by the hole-electron recombination or generation in the vicinity of the junction.

The transistor consists of two closely spaced *pn* junctions in a semiconductor with an order *pnp* or *npn*.

Contact rectification. If the height of a rectifying contact barrier is high, only a very small fraction of majority carriers can pass over the barrier. The fraction may be so small as to be comparable with the concentration of the minority carriers, provided the energy gap is not too large. The current due to the minority carriers becomes appreciable if the barrier height above the Fermi level approaches the energy difference between the Fermi level and the top of the valence band (Fig. 3).

The concentration of minority carriers is higher at the contact than in the interior of the semiconductor. With a sufficiently high barrier, it is possible to obtain at the contact a minority carrier concentration higher than that of the majority carriers. The small region where this condition occurs is called the inversion layer.

As in the case of a *pn* junction, a forward current produces injection of minority carriers. With the presence of an inversion layer, the injection can be so strong as to increase appreciably the conductivity in the vicinity of the contact. Ordinarily, contact rectifiers consist of a semiconductor in contact with a metal whisker. For large forward currents, the barrier resistance is small, and the resistance of the rectifier is determined by the spreading resistance of the semiconductor for a contact of small area. By increasing the conductivity in the vicinity of the contact where the spreading resistance is concentrated, carrier-injection may reduce considerably the forward resistance of the rectifier.

Surface electronics. The surface of a semiconductor plays an important role technologically, for example, in field-effect transistors and charge-coupled devices. Also, it presents an interesting case of two-dimensional systems where the electric field in the surface layer is strong enough to produce a potential wall which is narrower than the wavelengths of charge carriers. In such a case, the electronic energy levels are grouped into subbands, each of which corresponds to a quantized motion normal to the surface, with a continuum for motion parallel to the surface. Consequently, various properties cannot be trivially deduced from those of the bulk semiconductor. *See* CHARGE-COUPLED DEVICES; SURFACE PHYSICS.

H. Y. Fan

Bibliography. M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors,* 2d ed., 1988; K. Seeger, *Semiconductor Physics*, 7th ed., 1999; M. Shur, *Physics of Semiconductor Devices*, 1996; J. Singh, *Semiconductor Devices: An Introduction*, 1994; S. Sze, *Physics of Semiconductor Devices*, 2d ed., 1981; S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, 1989; R. K. Willardson et al. (eds.), *Semiconductors and Semimetals*, vols. 1–42, 1966–1994; M. Zambuto, *Semiconductor Devices*, 1989.

Semiconductor diode

A two-terminal electronic device that utilizes the properties of the semiconductor from which it is constructed. In a semiconductor diode without a *pn* junction, the bulk properties of the semiconductor itself are used to make a device whose characteristics may be sensitive to light, temperature, or electric field. In a diode with a *pn* junction, the properties of the *pn* junction are used. The most important property of a *pn* junction is that, under ordinary conditions, it will allow electric current to flow in only one direction. Under the proper circumstances, however, a *pn* junction may also be used as a voltage-variable capacitance, a switch, a light source, a voltage regulator, or a means to convert light into electrical power. *See* SEMICONDUCTOR.

Silicon and germanium are the semiconductors most often used in diodes. However, other materials may be used for special purposes: cadmium sulfide and cadmium selenide in photoconductors; gallium phosphide, gallium arsenide–phosphide, and silicon carbide in light-emitting diodes; and gallium arsenide in microwave generators. *See* GALLIUM; GERMANIUM; SILICON.

Diodes without pn junctions. The band structure of each semiconductor contains a forbidden energy gap. That is, there is a range of energies, ΔE, in which there are no quantum states that electrons may occupy. When the semiconductor is in thermal equilibrium, its electrons are distributed among the allowed quantum states according to Fermi statistics: The free electrons tend to reside near the minimum in the conduction band, and the unoccupied states in the valence band, called free holes, tend to be concentrated near the maximum in the valence band. **Figure 1** shows a schematic of the band structure

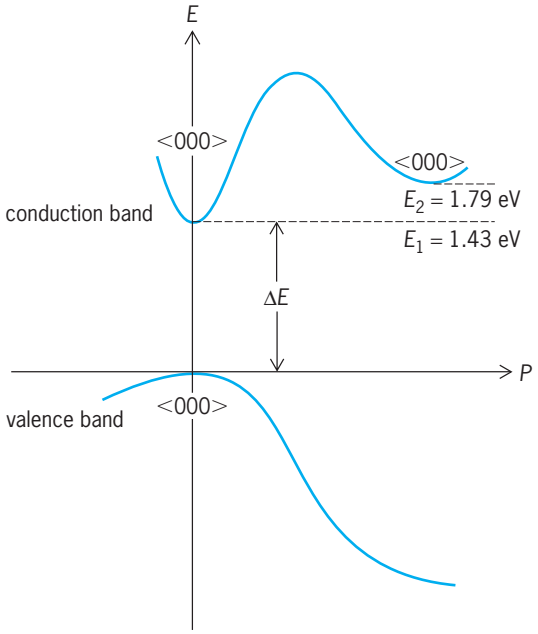


Fig. 1. Band structure of gallium arsenide; *E* = electron energy, *P* = momentum.

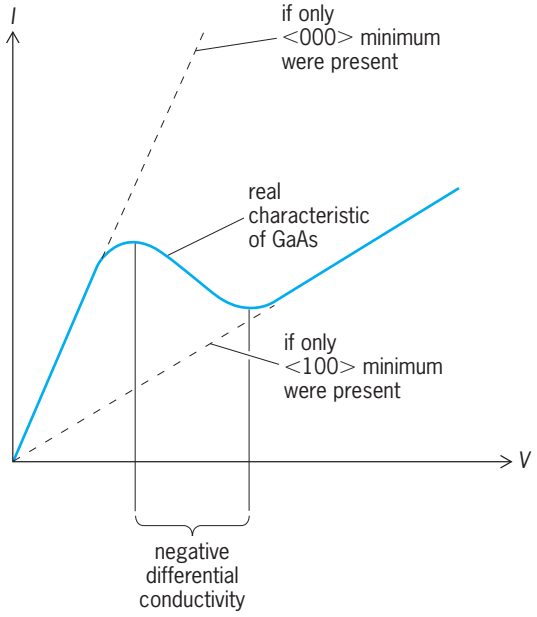


Fig. 2. Current-voltage characteristic bulk of GaAs.

in gallium arsenide, GaAs. According to Fermi statistics, the holes are located mainly at $E = 0, p = 0$; and the electrons are concentrated near $E = E_1, p = 0$.

The conductivity of a semiconductor is proportional to the number of electrical carriers (electrons and holes) it contains. In a temperature-compensating diode, or thermistor, the number of carriers changes with temperature. For example, in an intrinsic diode the number of carriers and the conductivity are proportional to $e^{-\Delta E/kT}$. As temperature T increases, the conductivity increases. This effect may be used to cancel out other conductivity changes in an electrical circuit so that the net change is zero. *See* THERMISTOR.

In a photoconductor the semiconductor is packaged so that it may be exposed to light. Light photons whose energies are greater than ΔE can excite electrons from the valence band to the conduction band, increasing the number of electrical carriers in the semiconductor. Thus the conductivity of a semiconductor is a measure of the light intensity striking the semiconductor. *See* PHOTOCONDUCTIVITY.

In some semiconductors the conduction band has more than one minimum. An example is gallium arsenide, as shown in Fig. 1. Here the ⟨000⟩ minimum is located 1.43 eV above the valence band, and the ⟨100⟩ minimum is located 0.36 eV above the ⟨000⟩ minimum. When an electric field is applied to GaAs, the electrons start out in the ⟨000⟩ minimum and the current-voltage relationship initially follows Ohm's law, as shown in **Fig. 2**. In this and other figures, I represents current and V represents voltage. When the field approaches 3200 V/cm, a significant number of electrons are scattered into the ⟨100⟩ minimum. Since carriers in the ⟨100⟩ minimum have lower mobilities than those in ⟨000⟩, the conductivity must decrease. This results in a region of negative differential conductivity, and a device operated in this

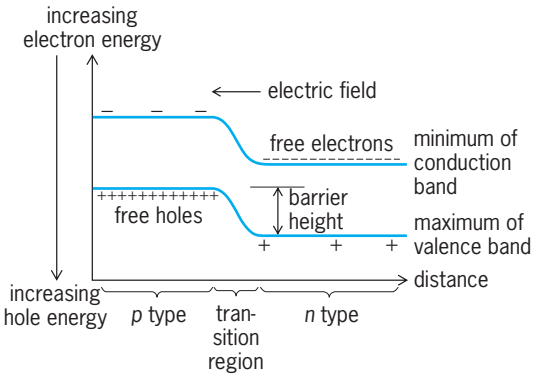


Fig. 3. Energy bands at the potential barrier for *pn* junction of a diode.

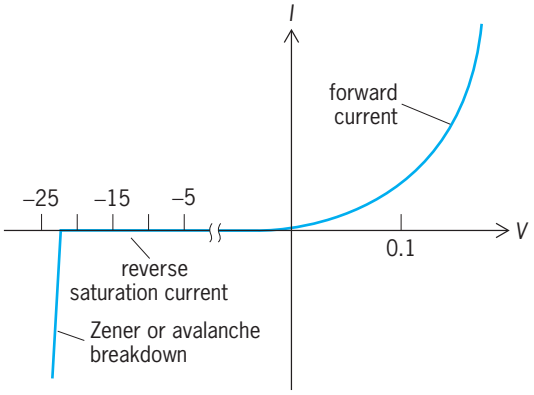


Fig. 4. Current-voltage characteristic of a *pn* junction.

region is unstable. The current pulsates at microwave frequencies, and the device, a Gunn diode, may be used as a microwave power source. *See* MICROWAVE SOLID-STATE DEVICES.

Diodes with pn junctions. A rectifying junction is formed whenever two materials of different conductivity types are brought into contact. Most commonly, the two materials are an *n*-type and a *p*-type semiconductor, and the device is called a junction diode. However, rectifying action also occurs at a boundary between a metal and a semiconductor of either type. If the metal contacts a large area of semiconductor, the device is known as a Schottky barrier diode; if the contact is a metal point, a point-contact diode is formed. *See* SCHOTTKY EFFECT.

The contact potential between the two materials in a diode creates a potential barrier which tends to keep electrons on the *n* side of the junction and holes on the *p* side. This barrier is shown in the energy band diagram of **Fig. 3**. When the *p* side is made positive with respect to the *n* side by an applied field, the barrier height is lowered and the diode is forward-biased. Majority electrons from the *n* side may flow easily to the *p* side, and majority holes from the *p* side may flow easily to the *n* side. When the *p* side is made negative, the barrier height is increased and the diode is reverse-biased. Then, only a small leakage current flows: Minority electrons from the *p* side flow into the *n* side, and minority holes from the *n* side flow into the *p* side. The current-voltage characteristic of a typical diode is shown in **Fig. 4**. Rectifying diodes can be made in a variety of sizes, and much practical use can be made of the fact that such a diode allows current to flow in essentially one direction only. At one extreme, single devices may be used to handle thousands of watts of power in changing alternating current to direct current. At the other extreme, a small diode may be used to detect an amplitude-modulated (AM) radio signal that has only microwatts of power.

The potential barrier at a *pn* junction consists of an electric dipole made up of charged impurity atoms, positively charged on the *n* side and negatively charged on the *p* side. Because the ends of the dipole are separated by a small distance, the junction acts like a capacitor. Moreover, the capacitance is voltage-variable since the barrier height is voltage-dependent. A diode which is specifically designed to utilize this capacitance characteristic is called a varactor.

When light photons with energies greater than ΔE strike a semiconductor near a *pn* junction, the potential barrier will sweep the conduction electrons that were generated by the light into the *n* side of the diode, and the newly created holes into the *p* side. If the diode, now called a solar cell, is connected to an external circuit, the diode will supply electrical power to the circuit as long as light strikes the *pn* junction. If, on the other hand, the diode is reverse-biased, the magnitude of the reverse saturation current can be used to measure light intensity. In this configuration the diode is called a photodiode. *See* PHOTOVOLTAIC EFFECT; SOLAR CELL.

When a junction diode is forward-biased, the number of electrons on the *p* side and the number of holes on the *n* side are increased far above their equilibrium values. This makes the recombination of electrons with holes highly probable, and the recombination energy may be released as photons of light.

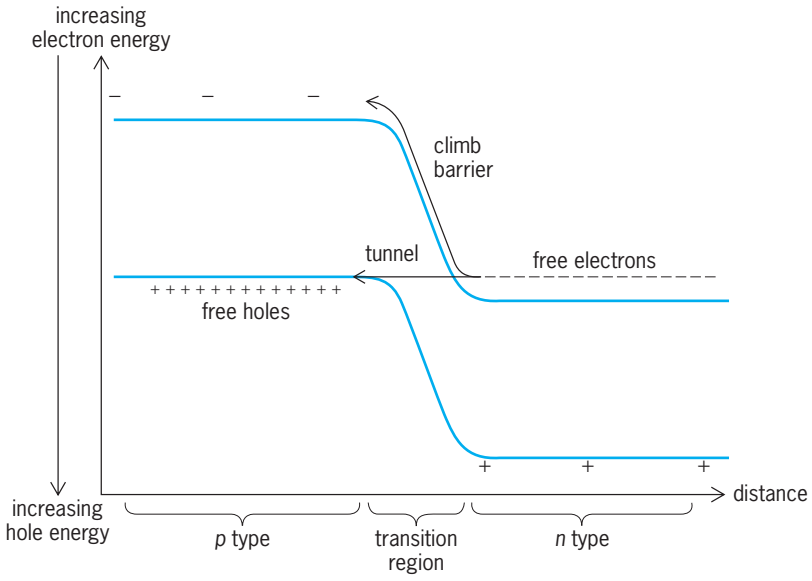


Fig. 5. Energy bands at *pn* junction in a tunnel diode.

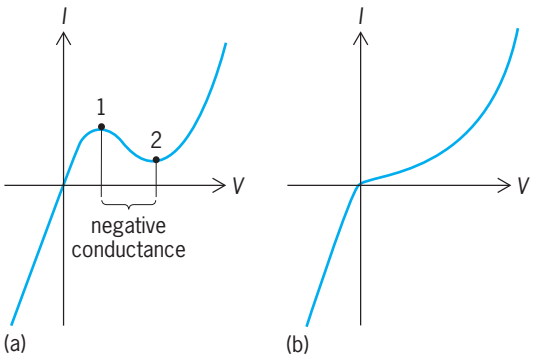


Fig. 6. Current-voltage characteristic of (*a*) a tunnel diode and (*b*) a backward diode.

In some materials, notably GaP for red or green light and GaAs for infrared light, the conversion of electrical energy directly to light can be made quite efficient. And in materials like GaAs, where the valence band maximum and the conduction band minimum are at the same value of crystal momentum, laser action may be achieved.

An Esaki tunnel diode is formed when both the *n* and *p* sides of a junction diode are very heavily doped. In this case, as shown in **Fig. 5**, an electron may move from the *n* side to the *p* side under forward bias in either of two ways: It may climb over the barrier, as in an ordinary diode; or it may quantum-mechanically tunnel through the barrier, since there are some available states on the *p* side that are at the same energy as states on the *n* side. Tunneling is highly probable at low and medium reverse voltages and at low forward voltages. As shown in **Fig. 6***a*, a tunnel diode has an unstable negative-resistance region. The diode may be operated either at point 1 or point 2, but may not be operated in between. This feature may be used to advantage in logic circuits, where the diode is made to switch from one point to the other. *See* TUNNEL DIODE.

A backward diode is similar to a tunnel diode except that it has no forward tunnel current (Fig. 6*b*). It is useful as a low-voltage rectifier.

Although simple diode theory predicts a constant reverse saturation current no matter how high the reverse voltage, all real diodes eventually break down either by tunneling or by avalanche (Fig. 4).

Avalanching occurs when the electrons or the holes gather so much energy that they can create new hole-electron pairs by collision. This process is cumulative and can result in a rapid buildup of current carriers. A diode that has a sharp breakdown provides a virtually constant voltage over a wide range of current. Popularly known as a Zener diode, this device may be used to provide a reference voltage or to regulate the voltage at some point in a circuit.

Stephen Nygren

Bibliography. J. Millman and A. Grabel, *Microelectronics*, 2d ed., 1987; G. W. Neudeck, *The PN Junction Diode*, 2d ed., 1989; J. Singh, *Semiconductor Devices: An Introduction*, 1994; E. S. Yang, *Microelectronic Devices*, 1988.

Semiconductor heterostructures

Structures consisting of two different semiconductor materials in junction contact, with unique electrical or electrooptical characteristics. A heterojunction is a junction in a single crystal between two dissimilar semiconductors. The most important differences between the two semiconductors are generally in the energy gap and the refractive index. In semiconductor heterostructures, differences in energy gap permit spatial confinement of injected electrons and holes, while the differences in refractive index can be used to form optical waveguides. Semiconductor heterostructures have been used for diode lasers, light-emitting diodes, optical detector diodes, and solar cells. In fact, heterostructures must be used to obtain continuous operation of diode lasers at room temperature. Heterostructures also exhibit other interesting properties such as the quantization of confined carrier motion in ultrathin heterostructures and enhanced carrier mobility in modulation-doped heterostructures. Structures of current interest utilize III–V and IV–VI compounds having similar crystal structures and closely matched lattice constants. *See* BAND THEORY OF SOLIDS; LASER; LIGHT-EMITTING DIODE; OPTICAL DETECTORS; REFRACTION OF WAVES; SOLAR CELL.

Carrier and optical field confinement. The most intensively studied and thoroughly documented materials for heterostructures are gallium arsenide (GaAs) and aluminum gallium arsenide ($Al_xGa_{1-x}As$). Several other III–V and IV–VI systems also are used for semiconductor heterostructures. The variation of the energy gap E_g and the refractive index $\overline{n}$ with aluminum arsenide (AlAs) mole fraction x are shown in **Fig. 1**. The lattice constant a_0 is also noted in Fig. 1 to emphasize that a_0 for GaAs and AlAs differs, but by an amount less than 0.14%. A close lattice match is necessary in heterostructures in order to obtain high-quality crystal layers by epitaxial growth and thereby to prevent excessive carrier recombination at the heterojunction interface. *See* CRYSTAL STRUCTURE.

A GaAs-$Al_{0.3}Ga_{0.7}As$ double heterostructure is illustrated in **Fig. 2**. The energy gap E_g versus x given in Fig. 1*a* shows that E_g for $Al_{0.3}Ga_{0.7}As$ is 0.37 eV greater than for GaAs. When a positive potential is connected to the *p*-type $Al_{0.3}Ga_{0.7}As$, electrons are injected into the GaAs layer from the wider-energy-gap *n*-type $Al_{0.3}Ga_{0.7}As$, and holes are injected into the GaAs from the wider-energy-gap *p*-type $Al_{0.3}Ga_{0.7}As$. The wider-energy-gap $Al_{0.3}Ga_{0.7}As$ layers also confine the injected electrons and holes to the GaAs layer, where they can recombine radiatively. This confinement is illustrated in Fig. 2*a*.

Figure 1*b* shows that the refractive index for the $Al_{0.3}Ga_{0.7}As$ layers will be less than for the GaAs layer. This refractive index step is shown in Fig. 2*b*. Solution of the reduced wave equation shows that the optical field will be confined to the larger-refractive-index GaAs layer. The resulting optical field distribution is shown in Fig. 2*c*. When this heterostructure is formed as a rectangular bar with parallel reflecting

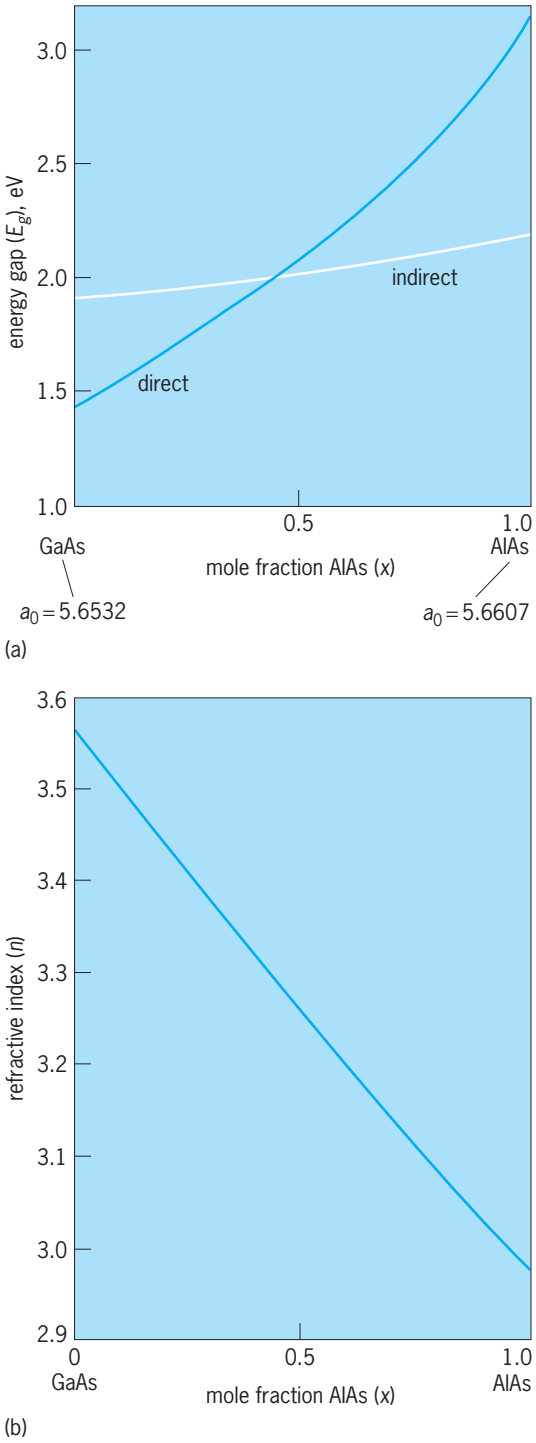


Fig. 1. Compositional dependence in $Al_xGa_{1-x}As$ of (*a*) the energy gap and (*b*) the refractive index at a photon energy of 1.38 eV. In both cases, temperature = 297 K (75°F). (*After S. E. Millet and A. G. Chynoweth, eds., Optical Fiber Telecommunications, Academic Press, 1979*)

surfaces, it can become a diode laser at current densities near 2×10^3 A/cm^2 at a forward voltage of 1.6 V.

Quantum well effects. When the narrow energy gap layer in heterostructures becomes a few tens of nanometers or less in thickness, new effects that are associated with the quantization of confined carriers are observed. These ultrathin heterostructures are referred to as superlattices, artificially layered structures, or quantum well structures, and they consist of alternating layers of GaAs and $Al_xGa_{1-x}As$ (**Fig. 3***a*). These structures are generally prepared by molecular-beam epitaxy. Each layer has a thickness in the range of 5 to 40 nanometers. The energy band diagram for the quantum well structure is shown in Fig. 3*b*. The abrupt steps in the energy gaps form potential wells in the conduction and valence bands. *See* ARTIFICIALLY LAYERED STRUCTURES.

In the GaAs layers, the motion of the carriers is restricted in the direction perpendicular to the heterojunction interfaces, while they are free to move in the other two directions. The carriers can therefore be considered as a two-dimensional gas. The Schrödinger wave equation shows that the carriers moving in the confining direction can have only discrete bound states. As the thickness of the quantum wells gets large, a continuum of states then results. The discrete states in the undoped GaAs quantum wells are illustrated in Fig. 3*b* and are accurately predicted by the well-known quantum-mechanical problem of the particle-in-a-box. *See* NONRELATIVISTIC QUANTUM THEORY.

The presence of the discrete quantum states may readily be observed by low-temperature optical absorption measurements. The observation

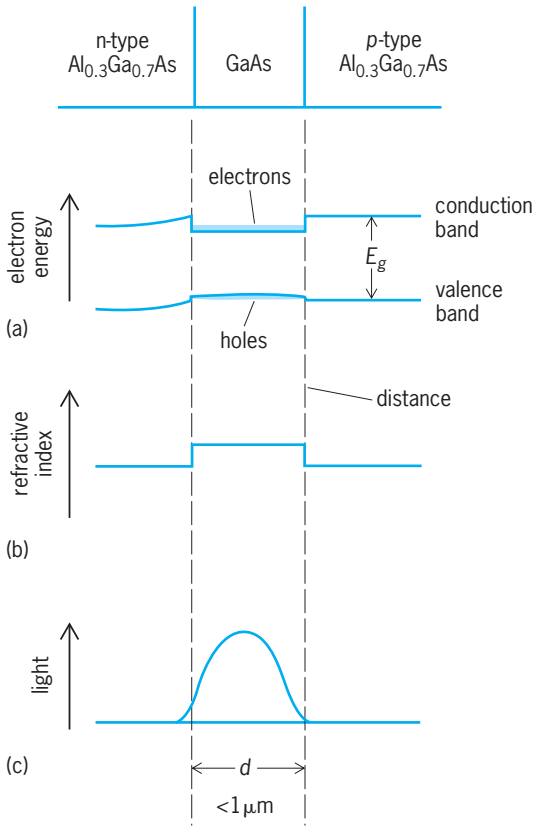


Fig. 2. Schematic representation of an $Al_{0.3}Ga_{0.7}As$-GaAs-$Al_{0.3}Ga_{0.7}As$ double heterostructure laser showing (*a*) energy band diagram at high forward bias, (*b*) refractive index profile, and (*c*) optical field distribution. (*After H. C. Casey, Jr., and M. B. Panish, Heterostructure Lasers, part A, Fundamental Principles, Academic Press, 1978*)

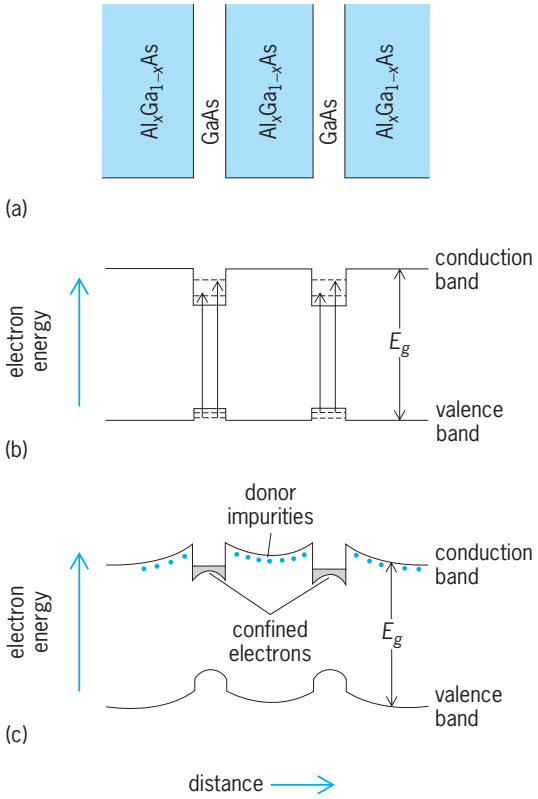


Fig. 3. GaAs-$Al_{0.3}Ga_{0.7}As$ superlattice. (*a*) Schematic representation. (*b*) Quantum wells, with the discrete conduction and valence band states for undoped GaAs. (*c*) Quantum wells with modulation doping.

wavelength is from 0.6 to 0.9 micrometer. These measurements are generally made at liquid helium temperatures with 20 or more GaAs quantum wells separated by $Al_xGa_{1-x}As$ layers. The absorption of photons takes place in the GaAs quantum wells by the transition of electrons in quantized states of the valence band to the quantized states of the conduction band. Sharp peaks in the absorption spectrum occur at the energies that separate the quantized states in the valence and conduction bands, and the photon energies of these absorption peaks vary with the thickness of the GaAs layer as predicted.

Modulation doping. Another property of semiconductor heterostructures is illustrated by a modulation doping technique that spatially separates conduction electrons and their parent donor impurity atoms. As illustrated in Fig. 3*c*, the donor impurities are incorporated in the wider-energy-gap $Al_xGa_{1-x}As$ layer but not in the GaAs layer. The GaAs conduction band edge is lower in energy than the donor states in the $Al_xGa_{1-x}As$, and therefore electrons from the donor impurities will move into the GaAs layers.

The useful feature of modulation doping is that the mobile carriers are the electrons in the GaAs layers and that these mobile carriers are spatially separated from their parent donor impurities in the $Al_xGa_{1-x}As$ layers. Since the carrier mobility in semiconductors is decreased by the presence of ionized and neutral impurities, the carrier mobility in the modulation-doped GaAs is larger than for a GaAs layer doped with impurities to give the same free electron concentration. Higher carrier mobilities should permit preparation of devices that operate at higher frequencies than are possible with doped layers.

Heteroepitaxy. In order to prepare optical sources and detectors that operate at a desired wavelength or to achieve high-mobility semiconductors, often it is necessary to use quaternary solid solutions. To obtain high-quality layers, the epitaxial layers must be grown on binary substrates that have the same lattice constant as the solid solutions. An example of considerable interest is the growth of $Ga_xIn_{1-x}P_yAs_{1-y}$ layers lattice-matched to InP. In this case, the heterojunction is used to obtain a similar lattice constant for dissimilar semiconductors at a desired bandgap. The heteroepitaxial growth techniques are chemical-vapor deposition, liquid-phase epitaxy, and molecular-beam epitaxy.

H. C. Casey, Jr.

Chemistry. The III–V and II–VI compounds and alloys commonly used in heterostructures are usually formed either by direct synthesis from the elements or by vapor-phase decomposition reactions using compounds containing hydrogen, halogen, or organic radicals. Substrates used for the growth of heterostructures are usually gotten from single-crystal boules synthesized directly from the elements. Most heterostructure materials can be doped either *n* or *p* type. However, in some cases one conductivity type can be difficult to obtain due to a "self-compensation" mechanism in which the formation of one type of vacancy is energetically favorable and compensates the opposite doping type. For example, the *p*-type beryllium (Be) doping of aluminum nitride (AlN) is compensated by the formation of *n*-type nitrogen vacancies. In general, variations in stoichiometry will greatly affect electrical and optical properties. Also, the properties of heterostructures greatly depend on the lattice constant match and chemical abruptness of the interface. A lattice mismatch of greater than 0.5% usually results in an interface with a significant density of both electron and hole traps. The chemical rule that determines the allowable alloy compositions is that the sum of the atom fraction of the group III (or group II) elements must equal the sum of the atom fractions of group V (or group VI) elements in the crystal, for example, $Ga_{1-x}Al_xAs$, $0 \leq x \leq 1$.

Fabrication. Semiconductor heterostructures are usually fabricated as single-crystal structures using thin-film epitaxial crystal growth techniques such as liquid-phase epitaxy (LPE), chemical vapor deposition (CVD), and molecular-beam epitaxy (MBE).

Liquid-phase epitaxy. The liquid-phase epitaxy method has been the most widely used technique for fabricating heterostructures. The principal advantage is the ease of fabrication of high-purity heterostructures with good electrical, optical, and interface properties as a result of impurity segregation from the growing layer into the melt during growth. The main disadvantage is the poor control of layer thickness and interface morphology and abruptness. Epitaxial growth by the liquid-phase

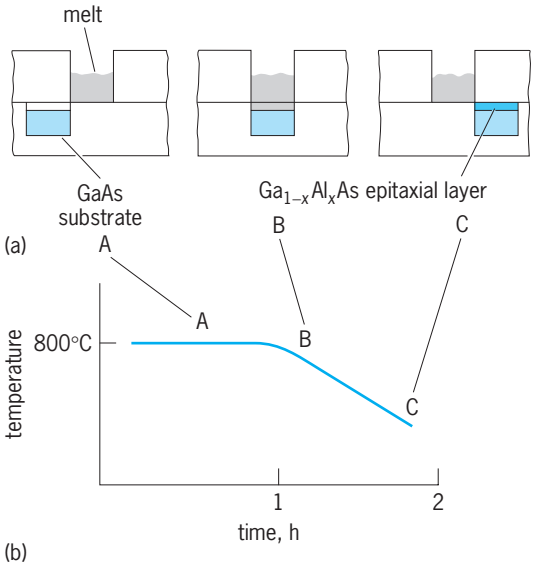


Fig. 4. Liquid-phase epitaxy method for GaAs-$Ga_{1-x}Al_xAs$ heterostructures. (*a*) Simplified diagrams of successive stages of the process. (*b*) Variation of temperature during the process. 800°C = 1500°F. (*After J. M. Woodall, III-V compounds and alloys: An update, Science, 208:908–915, 1980*)

epitaxy method occurs when a melt or solvent becomes supersaturated, usually by supercooling, with respect to a solid phase in presence of a single-crystal substrate. For the growth of III–V materials, the melt usually has a large concentration of one of the group III components in epitaxial layer. For example, epitaxial growth of $Ga_{1-x}Al_xAs$ for $0 \leq x \leq 1$ can be obtained from melts containing about 90% Ga and 10% Al plus As. A "typical" schedule for the liquid-phase epitaxy growth of a $Ga_{1-x}Al_xAs$ layer on GaAs is shown in **Fig. 4**. The important features are a "bottomless" bin to contain the melt and a fixture with a slightly recessed substrate which can be moved with respect to the melt bin. The first step is to thermally equilibrate the melt prior to contact with the substrate (A in Fig. 4). This means that the melt is either at solid-liquid equilibrium or slightly undersaturated. Next, the melt is cooled, usually at constant rate, while the substrate is in contact with it (B of Fig. 4). The desired layer thickness is obtained by programming either a change in temperature or elapsed time and then removing the substrate from the melt (C of Fig. 4). The growth of heterostructures can also be accomplished by using multimelt fixtures, moving the substrate between the various melts during cooling. The term liquid-phase epitaxy originally referred to growth from supercooled melts. The technique now includes growth from melts supersaturated by other techniques such as electroepitaxy, in which the solid–liquid interface region becomes supersaturated due to the flow of electric current across the interface and isothermal melt mixing techniques in which solid–liquid equilibrium melts of different compositions are mixed together and become supersaturated.

Chemical vapor deposition. The chemical vapor deposition method refers to the formation of thin solid films as the result of thermochemical vapor-phase reactions. When the films are epitaxial, the method is sometimes called vapor phase epitaxy (VPE). For III–V materials two different chemistries have been widely studied: the group III and group V halogen compounds and group V hydrogen compounds; and the group III metal-organic compounds and V–hydrogen compounds such as $Ga(CH_3)_3$ and AsH_3. The halogen transport reactions are of the "hot" to "cold" type in which the III-halogen is produced in a high-temperature zone by the reaction of the III element with HCl. The III-halogen then diffuses to the low-temperature zone, where it combines with the V species to form an epitaxial layer of III–V material. Vapor-phase epitaxy using metal-organic compounds (metal-organic chemical vapor deposition; MOCVD) occurs when the organic radical R of the III-R compound is "cracked" or pyrolyzed away at a "hot wall" or hot substrate in the presence of the V-H_3 compound. A schematic illustration of the metal-organic chemical vapor deposition technique applied to GaAs is shown in **Fig. 5**. Both the halogen and metal-organic chemistries have been successfully used to fabricate heterostructures. It is thought that the metal-organic chemical vapor deposition method may have long-term advantages such as better purity, composition, and thickness control and the ability to produce a wide range of III–V compounds and alloys.

Molecular-beam epitaxy. The molecular-beam epitaxy technique is relatively new compared with liquid-phase epitaxy and chemical vapor deposition. For heterostructure fabrication, it is capable of a layer thickness and interface abruptness control on a monoatomic scale. This represents almost two orders of magnitude improvement in structural resolution in the growth direction compared to liquid-phase epitaxy and chemical vapor deposition. The molecular-beam epitaxy technique is similar to

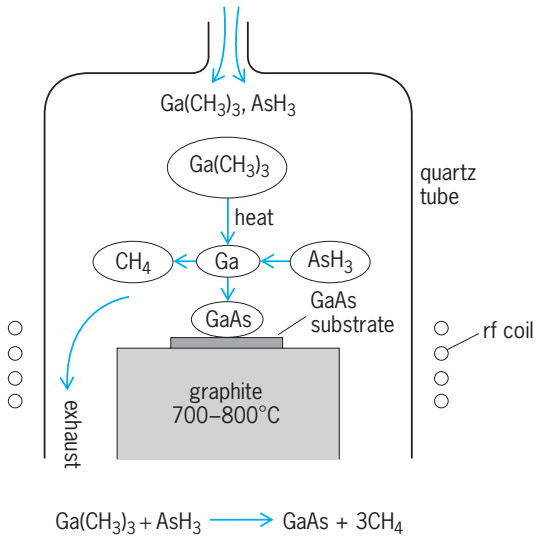


Fig. 5. Simplified diagram of the metal-organic method for chemical vapor deposition. 700–800°C = 1300–1500°F.

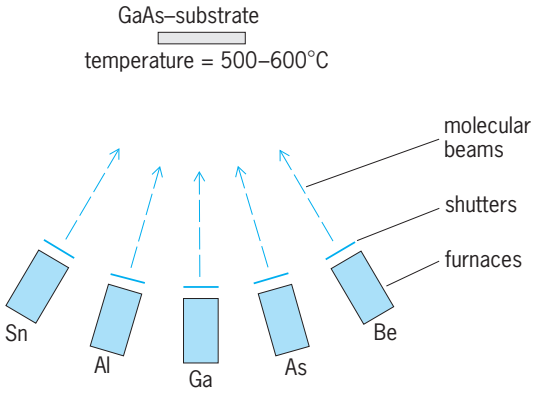


Fig. 6. Schematic illustration of a system configuration for growth of doped $Ga_{1-x}Al_xAs$ by molecular-beam epitaxy. 500–600°C = 900–1100°F. (*After P. E. Luscher, Crystal growth by molecular beam epitaxy, Solid State Technol., 20(12):43–52, 1977*)

evaporation techniques used to deposit thin metal films. The key features of the molecular-beam epitaxy method are the use of an ultrahigh-vacuum environment and epitaxial growth by the reaction of multiple molecular beams of differing flux and chemistry with a heated single-crystal substrate. A schematic illustration of molecular-beam epitaxy applied to doped $Ga_{1-x}Al_xAs$ is shown in **Fig. 6**. Each furnace heats a crucible which is charged with one of the constituent elements or compounds of the desired film. Tin (Sn) functions as an *n*-type dopant, and beryllium (Be) functions as a *p*-type dopant. The furnace temperature is chosen so that the vapor pressure is sufficient to produce the desired beam flux at the substrate surface. The furnaces are arranged so that the flux from each is maximum at the substrate position. The quality and composition of the epitaxial film are determined by the quality of the ultrahigh-vacuum system, the substrate temperature, and the furnace temperature. The shutters interposed between each furnace and the substrate allow the beams to be modulated. This feature, coupled with typical growth rates of a few tenths of a nanometer per second, facilitates the fabrication of very thin and very abrupt heterostructures. *See* CRYSTAL GROWTH; MOLECULAR BEAMS; SEMICONDUCTOR.

Jerry M. Woodall

Bibliography. M. G. Astles, *Liquid-Phase Epitaxial Growth of III–V Compound Semiconductor Materials and Their Device Applications*, 1990; N. G. Einspruch and W. R. Frensley (eds.), *Microstructure Science: Heterostructures and Quantum Devices*, 1994; M. A. Herman and H. Sitter, *Molecular Beam Epitaxy*, 2d ed., 1997; E. V. Ivchenko and P. E. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena*, 2d ed., 1997; R. A. Kiehl and G. Sollner (eds.), *Semiconductors and Semimetals*, vol. 41: *High Speed Heterostructure Devices*, 1994; M. B. Panish and H. Temkin, *Gas Source Molecular Beam Epitaxy: Growth and Properties of Phosphorus-Containing III–V Heterostructures*, 1993; J. Singh, *Physics of Semicondutors and Their Heterostructures*, 1993; S. Sivaram, *Chemical Vapor Deposition*, 1995.

Semiconductor memories

Devices for storing digital information that are fabricated by using integrated circuit technology. Semiconductor memories are widely used to store programs and data in almost every digital system, and have replaced core memory as the main active computer memory.

Organization. Many different types of semiconductor memories are used in digital systems to perform various functions—bulk data storage, program storage, temporary storage, and cache (or intermediate) storage. Almost all of the memories are a form of random-access memory (RAM), in which any storage location can be accessed in the same amount of time.

The storage locations are set up as an array of rows and columns (**Fig. 1**). To the user, the storage array is organized as *M* words deep by *N* bits wide, even though the actual physical arrangement may be different. To access the data, the system supplies addresses—multiple-bit parallel words—that specify the locations to be accessed. Each address pinpoints the storage cell or group of cells (the memory bit or word, respectively), very much in the same way as a pair of grid coordinates specifies a location on a two-dimensional map. Various control lines then permit the system to either read data out from or store new data into the memory, or both, depending on the memory type.

There are also some specialized memory circuits that can be accessed only sequentially. These devices either are based on long strings of shift registers or use an internal RAM array but have a built-in address sequencer that automatically increments or decrements through all the memory locations when the circuit receives timing and control signals (**Fig. 2**). There is another type of serial memory that does permit random access. In this unique memory, the multibit address is loaded one bit at a time rather than in parallel, as with most other memories. After the address is loaded, a control signal tells the memory to access the data, and then the desired data are serially shifted in or out. Such serial memories are typically much slower to access since it is necessary to load in each address with many clock cycles, rather than a single cycle, as with the standard memory chips.

Read/write RAMs. Even though most semiconductor memories can be randomly accessed, they are not all referred to as RAMs. RAMs are memory chips that cannot retain data without power but permit data to be both read from or written into the memory chip's storage locations.

Dynamic and static types. Within the category of read/write RAMs, many subdivisions have been created to satisfy the performance and system architecture requirements of the various applications. Basically there are two types of read/write RAMs—dynamic and static (DRAMs and SRAMs). The terms "dynamic" or "static" refer to the structure of the actual storage circuit (the cell structure) used to hold each data bit within the memory chip. A dynamic memory uses a storage cell based on a transistor

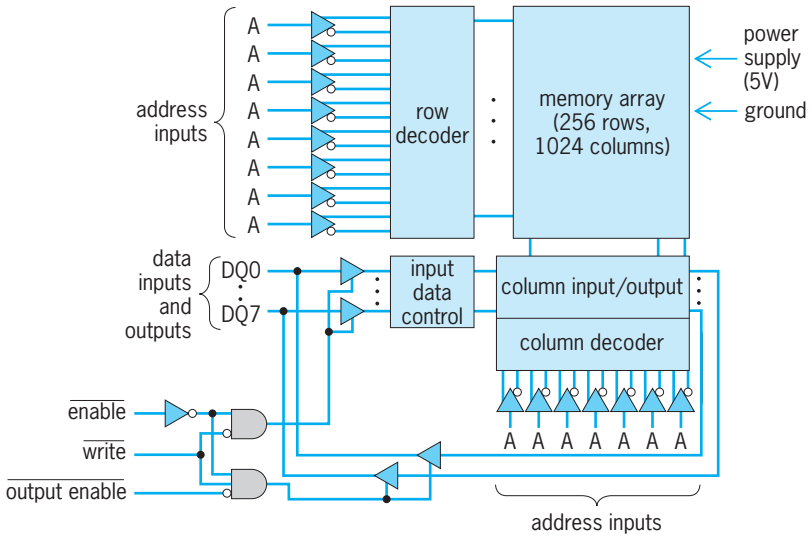


Fig. 1. Basic memory chip with array of memory cells. In this case, a 256-kbit static random-access memory is set up as 256 rows by 1024 columns, which are accessed when addresses are presented at the address inputs and are decoded to select the proper row and column. The 8-bit-wide data path, DQ0 to DQ7, is bidirectional, and just three active-low control lines—enable, write, and output enable—coordinate chip operation.

and capacitor combination, in which the digital information is represented by a charge stored on each of the capacitors in the memory array (**Fig. 3**). The memory gets the name "dynamic" from the fact that the capacitors are imperfect and will lose their charge unless the charge is repeatedly replenished (refreshed) on a regular basis (every few milliseconds) by externally supplied signals. Static memories, in contrast, do not use a charge-storage technique; instead, they use either four transistors and two resistors to form a passive-load flip-flop (**Fig. 4***a*), or six transistors to form a flip-flop with dynamic loads (Fig. 4*b*), for each cell in the array. Once data are loaded into the flip-flop storage elements, the flip-flops will indefinitely remain in that state until the information is intentionally changed or the power to the memory circuit is shut off.

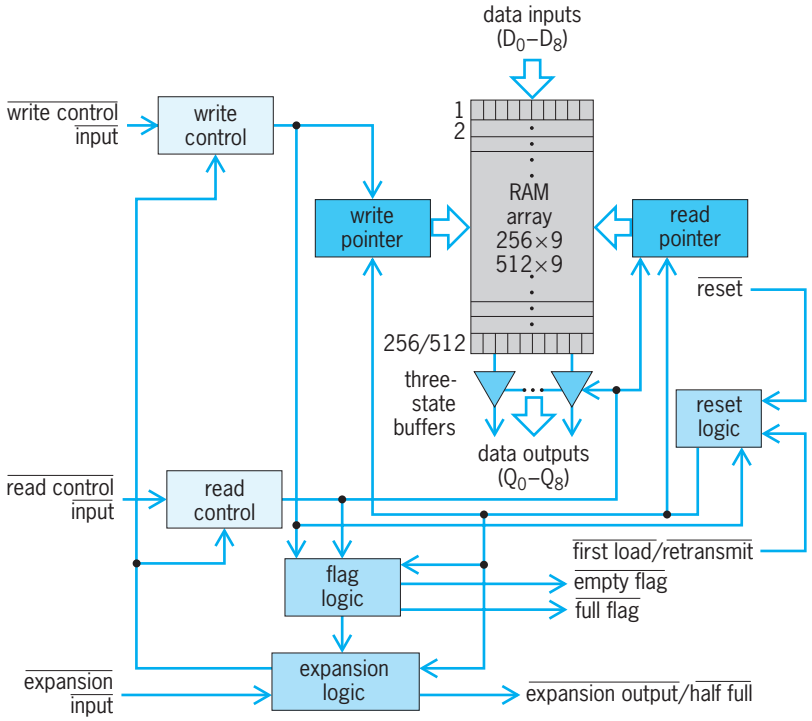


Fig. 2. A 9-bit-wide first-in/first-out (FIFO) memory chip. Inside the memory, read and write pointers provide sequential addresses that determine which memory location is being accessed. Separate 9-bit data paths allow data writes and data reads to take place simultaneously. Additional circuits on the chip provide information as to whether the memory array is empty, full, or half full.

The simple cell structure of dynamic RAMs (two elements—a transistor and a capacitor—that can be fabricated either side-by-side or one on top of the other) makes possible exceptionally small storage-cell areas. (As of 2001, the smallest memory cell area reported for a DRAM was just over 0.1 square micrometer. The small cell was fabricated using minimum feature dimensions of 0.1 μm.) Static RAMs, which require at least four or six transistors for each cell, consume almost four times the area to achieve the same storage capacity as dynamic RAMs. (In 2001, however, researchers crafted a loadless static memory cell that requires an area of just 0.99 μm^2 when fabricated using 0.13 μm design rules.) However, the continual refresh required by the dynamic RAMs demands that memory arrays be surrounded by support circuits (both on- and off-chip to provide the refresh timing signals), and that causes both the DRAM cells and refresh circuits to consume a few milliwatts of power even when the memories are not being accessed. That power drain must be taken into account when dynamic RAMs are used in battery-powered systems.

The interest in low-power system design that started in the early 1990s, coupled with the move downward in power supply voltage, has created the need for dynamic RAMs that operate with 3.3-V and lower power supplies. In the late 1990s, designers were working with even lower supply voltages. Many of the memories designed after 1998 use a 2.5-V power supply, further lowering the power consumption of the memory chips as densities go beyond 256 Mbits on a chip. Still lower power supply voltages are in demand—many designs are now targeted to operate from supply voltages as low as 1 V for cellular phones, digital personal assistants, and other portable systems.

Furthermore, the low-power dynamic RAMs must have special capabilities to retain the data during long standby periods with a minimal power drain. To accomplish this, memory designers have created special slow-refresh operating modes on low-power dynamic RAMs so that fewer refresh cycles are needed to retain the data in the dynamic memory cells, and thus operating power demands are reduced.

To keep dynamic RAM package pin counts as low as possible, the address information is divided into both a row address and a column address. The row and column addresses are time multiplexed with each other and sent to the memory over the same set of pins, thus doubling the amount of time required to supply the complete memory location address. As a result, the best access time for a traditional dynamic RAM can be five times slower than that of a

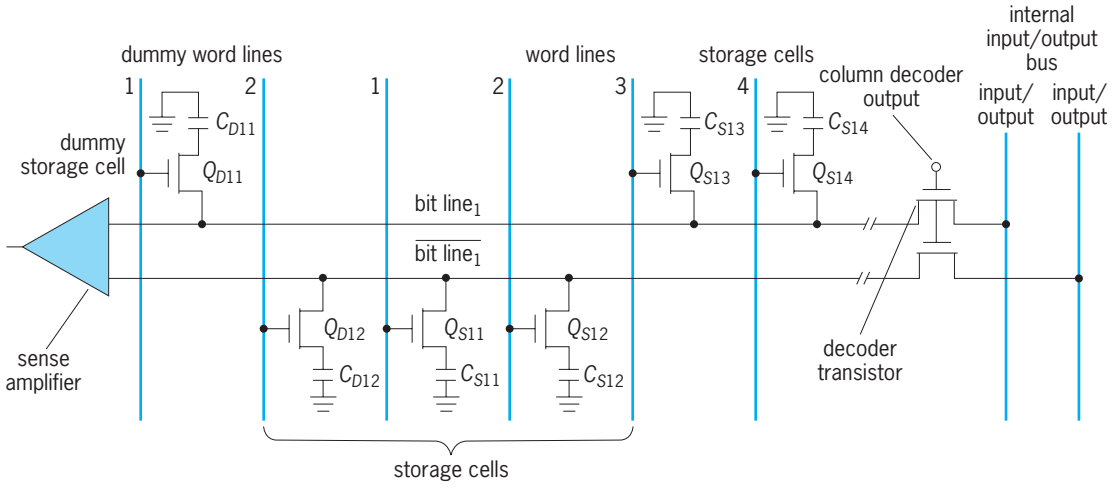


Fig. 3. Internal structure of a dynamic RAM. A typical bit line has many storage cells attached to it, with each cell consisting of a gating transistor Q_S controlled by the word line and a storage capacitor C_S connected to ground. When a particular cell is accessed on the bit line, a matching dummy cell (with gating transistor Q_D and capacitor C_D) is activated so that the sense amplifier receives a clean differential signal that represents the stored value.

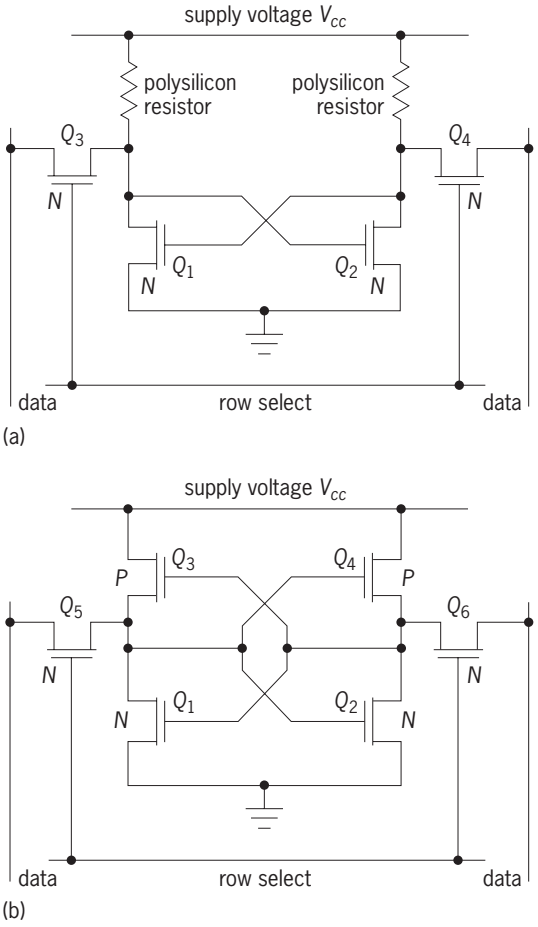


Fig. 4. Static RAM storage cells. (*a*) Cell with four NMOS transistors ($N = n$-channel), which form a simple cross-coupled flip-flop, and two large-value resistors, formed in an overlying polysilicon layer, that serve as the passive load elements. (*b*) Cell with PMOS transistors ($P = p$-channel) instead of resistors as the load elements. Leakage current in the cell is reduced since the off-resistance of the transistors is higher than that of polysilicon resistors.

high-performance static memory fabricated with the same process and design rules.

Alternative types. In addition to static and dynamic RAMs, there is an attempt to combine both technologies, thus merging the high storage density of dynamic memory cells with the simplicity of use of static RAMs. Referred to as pseudostatic or pseudodynamic RAMs, these memories include circuits on the chip to automatically provide the refresh signals needed by the dynamic cells in the memory array. Since the signals do not have to be supplied by the external system, the memory appears to function like a static RAM.

Dynamic RAMs and static RAMs continue to evolve, driven by the need for extremely fast access times and high memory data bandwidths to feed high-performance microprocessors. The standard DRAM has given way to devices with interfaces that incorporate fast-page, nibble, or extended-data-out (EDO) access modes, synchronous bus interfaces, and other more-dedicated interfaces based on standards such as Rambus (developed by Rambus Inc.) or Sync-Link (developed by the SLDRAM consortium), or novel devices such as the multibank DRAM architecture developed by Mosys Inc.

Most personal computers and workstations incorporate synchronous DRAMs (SDRAMs) for their main memory since the chips offer higher data transfer speeds (10 ns or less per transfer after the initial access) than previous memory types and thus allow higher-performance computer systems to be implemented. Fast-page, nibble mode, and EDO DRAMs are all fading in popularity since their performance cannot keep pace with the faster systems.

The synchronous dynamic RAM employs a standard dynamic RAM core, but the system interface timing is lock-stepped to the rest of the system through the use of a master clock input signal. Initial 16-Mbit and 64-Mbit synchronous memories were able to operate at system clock rates of 100 MHz, while the latest 256-Mbit SDRAMs can operate at system bus speeds of 133 MHz or faster. (1 kbit = 1024 bits;

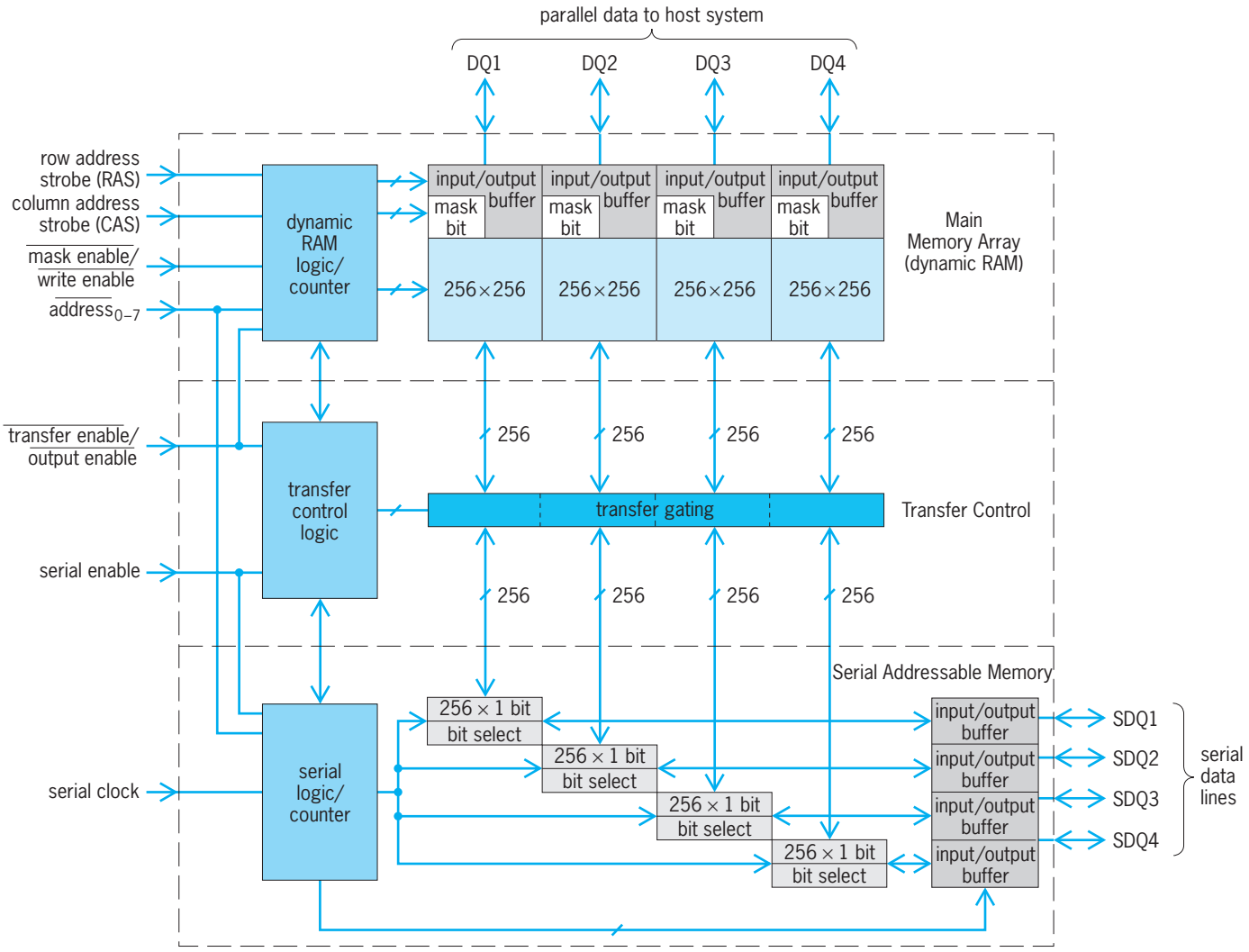


Fig. 5. Video memory organized as 64 kwords by 4 bits, with two data paths. On the top, a 4-bit-wide parallel data path to the 64-kword storage array is divided into four banks of 64 kbits each. The RAS and CAS signals, along with the eight address lines, permit any of the 64-k locations to be randomly accessed. From each of the four banks, a 256-bit string of bits can be transferred to separate bidirectional shift registers and serially shifted out at high speed over the four serial data lines. Conversely, data can be serially shifted into the four serial data lines and transferred into the chip's main memory array.

1 Mbit = 1024 × 1024 bits = 1,048,576 bits.)

New clocking techniques promise to rapidly reduce the transfer speeds of synchronous memories still further by using both the leading and trailing edges of the clock signal to transfer data. (Most synchronous memories prior to 1998 used only one edge of the clock signal to do the transfer.) This effective doubling of the transfer rate is called double-data-rate (DDR), and as of 2001 such DDR interfaces were available on SDRAM chips from just about every DRAM supplier. Although initially developed for use in SDRAMs, the technique can also be applied to synchronous static memories and can effectively double the data transfer speed of those memories as well.

Additionally, design automation tools, which make it much easier to design or modify a circuit, have opened up many new hardware options, especially for video and graphics. Specialized DRAMs that incorporate second and even third data ports provide the high bandwidths needed by high-performance graphics controllers and multimedia subsystems. The first of these variants was the video RAM which, in addition to the standard DRAM memory array, incorporated a high-speed 4- or 8-bit wide buffer memory that transfers a sequential string of data to the graphics controller (**Fig. 5**).

Several other memory types were developed for graphics applications during the 1990s, including devices such as the Rambus interface, Window RAMs, triple-port RAMs, and synchronous graphics RAMs (SGRAMs). For mass-market applications, such as the graphic-adapter cards used in most personal computers, the SGRAM became the most popular option. (The SGRAM is similar to the SDRAM, but has a few additional functions that better suit it for graphics applications.)

Computer systems in 2001 started to incorporate a variation of the original Rambus interface into main memory subsystems. The Direct Rambus interface, an enhanced memory-to-CPU (central processing unit) bus defined by Intel Corp. and Rambus Inc., allows data transfers of 800–1000 Mbytes/s and faster (**Fig. 6**). The Sync-Link DRAM (SLDRAM) interface, developed jointly by a consortium of companies to

serve as a standard, high-performance memory interconnect option, was dropped as an alternative to the SDRAM or Rambus DRAM (**Fig. 7**); however, parts of the SLDRAM definition will persist as designers finalize the definition of a high-performance second-generation DDR interface that incorporates some aspects of the SLDRAM interface.

When Direct Rambus DRAMs (RDRAMs) perform a read operation, a bank of memory in the DRAM core is activated (the core is actually split into multiple banks—16 in the case of the first Direct RDRAMs) and, based on the input address, transfers the requested data from the bank into the internal input/output buffer. That buffer, though, is 128 bits wide and actually receives data from all 16 banks so that after the first access, sequential bytes in the buffer can be accessed at the clock speed. Thus, when clocked at 333 MHz and both edges of the clock are used, the effective data transfer rate is 666 Mbytes/s for a Direct RDRAM device with an 8-bit interface, and over 1300 Mbytes/s for a device with a 16-bit data interface. If the desired data are not in the buffer, the latency encountered to reach the first piece of the data runs about 20 ns; however, subsequent accesses to other data in the buffer take place at the approximate 1.6 ns/clock transfer rate. Even faster versions that could clock at 400 MHz (800 Mtransfers/s) were available in 2000, and 500- and 600-MHz versions were sampled by suppliers in 2001.

Specialty dynamic RAMs with small blocks of static RAM, forming a unique type of cached dynamic RAM, were introduced in 1993. These unusual memories allow system designers to implement memory subsystems with the cache subsystem distributed across the dynamic RAM array. The advantage of the cache-dynamic RAM (CDRAM) integration is that in many smaller systems it can eliminate the need for a separate secondary cache and thus reduce the system cost or size.

In addition to the various DRAM options available for main memory, there are many specialized memory components that provide data buffering in video systems and data communications applications, or speed adjustments between two systems that operate at different bus speeds. First-in/first-out (FIFO) memories provide buffering for burst data transfers, while dual-port or multiport memories allow multiple subsystems, each operating at different clock rates, to read or write data to a central memory and thus provide rate buffering. Although most FIFO memories and multiport memories are fabricated with static RAM cells, a few of these specialty memories are based on arrays of DRAM cells surrounded by logic that turns the memory array into the desired memory type.

Static memories. Just as diversity has become the norm for DRAMs, the same is true for static memories. In addition to the basic asynchronous SRAM, the industry has crafted a variety of higher-performance SRAM approaches to meet the access time requirements of high-speed computers, telecommunications systems, and many other applications. Standard asynchronous SRAMs, which operate with the straightforward application of address inputs and a few simple control signals and data, are the most commonly used memory type and are found in many moderate-performance and low-power applications. Access times for asynchronous devices tend to be in the 60–200 ns range.

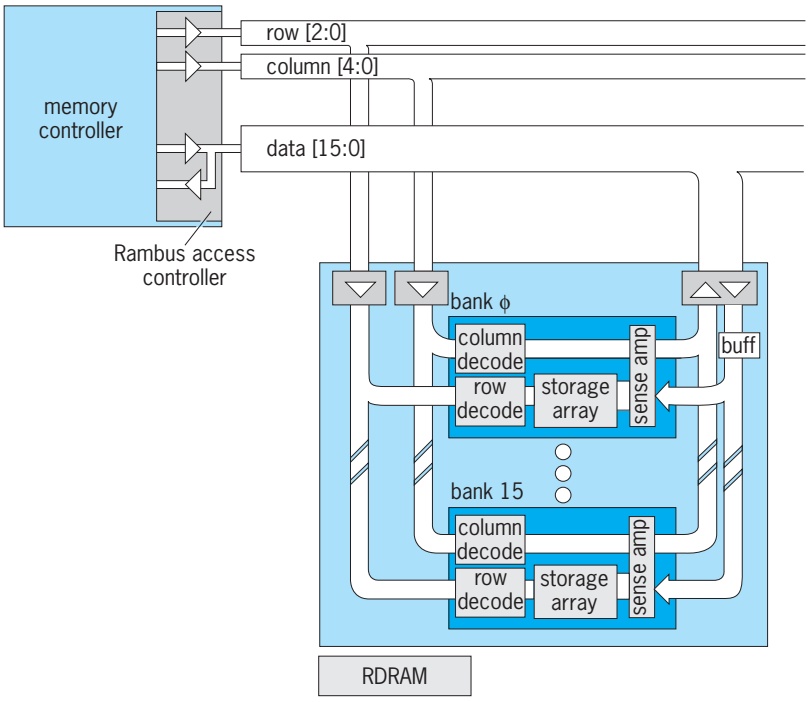


Fig. 6. Multiple memory banks are used in the Direct Rambus DRAM (RDRAM) to provide high-speed access to data. The 16-bit-wide interface and fast bank access allows the Direct RDRAMs to deliver a bandwidth of more than 4 Gbytes/s. (*Rambus Inc.*)

The huge growth in portable equipment, from cell phones to battery-powered medical instruments, has created a demand for low-power asynchronous memories. These devices consume less than a few hundred milliwatts of power when active, and microwatts of power when idling. To achieve that, especially for the higher-density devices (512 kbits and larger), memory designers are moving the memory cell structures from the commonly used four-transistor cell to the six-transistor cell to reduce leakage currents, simplify manufacturing, and improve the cell tolerance to a phenomenon called soft errors (random changes in bit values due to unpredictable impacts of alpha-particle radiation). *See* RADIATION HARDENING.

High-speed personal computer and workstation cache-memory subsystems typically employ synchronous devices. Some derivative versions of the synchronous memories include small counters on the chip that will let the memories deliver four words in a burst, such that the computer supplies the first address, and then upon each of three subsequent clock cycles, the counter increments and the next data word is sent to the computer, thus eliminating the need for the computer to send a new address for each word. These devices are called pipelined synchronous-burst SRAMs and are widely used with microprocessors such as the Intel Pentium and the IBM/Motorola PowerPC to implement level-2

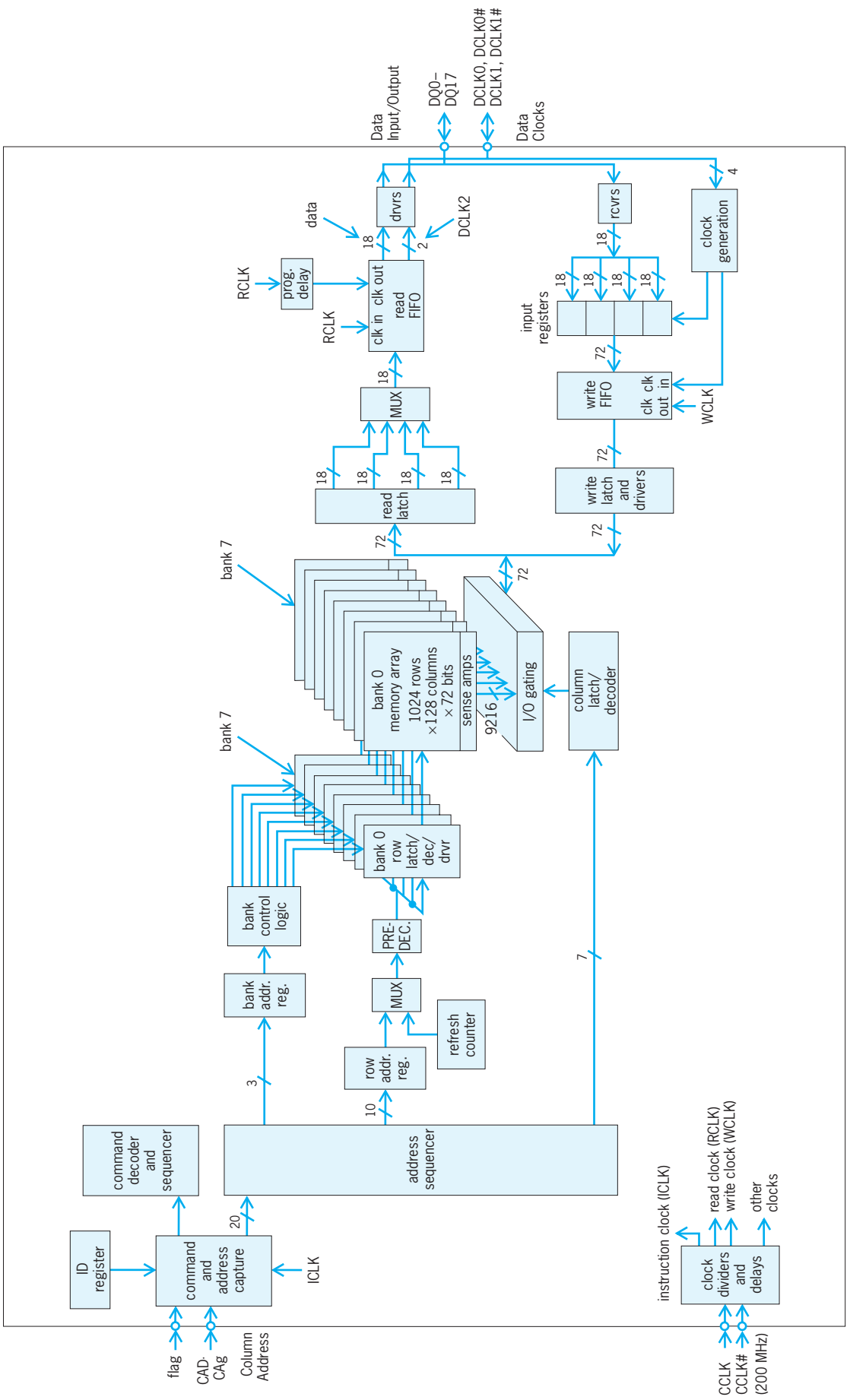


Fig. 7. Synchronous-link DRAM uses multiple memory banks and a packet-based transfer protocol. It decouples the address, control, and data paths so that multiple internal operations can take place concurrently, which speeds memory operation and allows data transfers of 800 Mbytes/s and faster. (*SLDRAM consortium*)

cache subsystems that supplement the on-chip level-1 caches of the processors. The on-chip counters can either be set to count in the special order required by x86 family processors or in standard linear sequence used by other processors.

Yet another variation to the synchronous interface, called zero-bus turnaround or no-bus latency, eliminates the short delay time required by the burst-type parts when the direction of data on the bus must be reversed. For example, when a read operation follows a write operation, typical burst memories must reset some internal circuits, and the adds a delay of a full memory cycle (several nanoseconds) to the response time.

With the no-latency devices, a read can follow a write or vice versa, with no performance penalty. Such memories provide a higher-performance alternative to equipment designers who must deal with reads and writes that occur in a totally unpredictable sequence rather than the typical sequential reads and writes in a desktop computer or workstation cache. Personal computer and workstation cache subsystems typically transfer four or more sequential words into internal processor caches. Those transfers are most often reads that are done by the central processing unit rather than writes to the memory, and thus the time it takes for the bus to reverse direction is of minimal consequence.

In 2001 a new high-performance SRAM interface was developed. Known as a quad-data-rate (QDR) interface, this scheme uses the standard DDR signals to transfer data. However, it also employs separate read and write memory ports. The read and write ports can be simultaneously active, thus doubling the maximum data transfer rate possible with standard single-port DDR memories. That doubling of the double-data-rate transfer speed results in a fourfold overall improvement over the standard single-data-rate memory.

Redundancy. As memory chips increase in storage capacity, manufacturing defects can often cause several storage cells to fail and thus render the chip useless. To counter that problem and thus improve production yield, many manufacturers have incorporated redundant storage cells that can be swapped into the memory array during unpackaged chip testing via the blowing of electrical fuses or the use of a laser to burn away microscopic fuses. Most static and dynamic RAMs now incorporate some degree of redundancy, and all memory types except the read-only memories typically include a small number of redundant cells as a hedge against defects.

Capacity and special features. In the early 1970s, dynamic memories with a capacity (density) of 1024 bits per chip were introduced. Since then, improvements in semiconductor processing and circuit design have made practical continual increases in density.

Traditional dynamic RAMs were introduced in organizations with single-bit word widths and depths that started at 1024 words. As the memory depths increased, the single-bit-wide memory chips did not provide the optimum increments (granularity) in all but the largest of systems. Thus, dynamic RAMs with 4-, 8-, 9- (the ninth bit is used for parity), 16-, and 32-bit-wide data paths have appeared to best match the architectures of diverse systems.

By the mid-1990s DRAMs with densities of 16 Mbits were commercially available, and although samples of 256-Mbit devices have been available since 1996, many manufacturers found it a little harder than expected to mass-produce such chips, with mass production starting in late 1999. In the interim, the 16- and 64-Mbit chips became the workhorse DRAMs, and in 1998 and 1999 a number of manufacturers introduced the first intermediate density, 128-Mbit DRAM. [In past generations manufacturers typically increased capacity by increments of 4X (1-k, 4-k, 16-k, 64-k, 256-k, 1-M, 4-M, 16-M, 64-M, 256-M, 1 Gbit, and so on); since one row and one column address signal must be added, the two extra address bits represent a fourfold increase in address range.]

Blocks of memory (notably DRAM and SRAM), functions such as FIFO buffers and multiport SRAMs, and flash EEPROMs are also available as building blocks in the design libraries of most manufacturers of application-specific integrated circuits (ASICs). Although SRAMs, FIFOs, and multiport SRAMs have been part of the design libraries since the 1980s, the mid-1990s saw the first commercial offerings of custom-designed logic chips that embedded large blocks of DRAM and large blocks of flash memory. Since then, DRAM blocks of up to 64 Mbits, SRAM blocks over 1 Mbit, flash memory blocks of up to about 4 Mbits, and blocks of FIFO and multiport memories have been available to designers of custom circuits.

The increased use of cache memories in the late 1980s spawned a new class of static memory chips with special features to ease their use in fast systems and eliminate the need for off-chip circuits to implement the functions. Some typical examples include individual bit-enable control lines on multibit outputs, parity bits carried through the memory, special cache-tag memories to carry the tag bits for the cache, special self-timed logic to permit extremely fast access times in pipelined systems, and very wide word sizes to reduce the chip count in very wide word but low word depth systems.

Access times for the fastest cache memory chips in the late 1990s were well below 5 ns, allowing them to keep pace with processors that operate at clock speeds of 300 MHz and higher. Still higher-speed processors introduced after 2000 actually eliminated the need for stand-alone cache memories since CPU designers incorporated both first- and second-level caches onto the same chip as the CPU. In cache subsystems, memory bandwidth is a key consideration, while memory depth for a second-level cache is typically limited to about 4 Mbytes. Thus memories with wide words are very popular for cache subsystems. In the cache-memory area, companies offer a range of memories with word widths of 8, 9, 16, 18, 32, 36, and 64 bits, and memory capacities of up to 8 Mbits (1 Mbyte) on a single chip. For small cache subsystems, a single cache-memory chip can provide all the desired memory.

Nonvolatile memories. There are many other forms of semiconductor memories in use—maskprogrammable read-only memories (ROMs), fuse-programmable read-only memories (PROMs), ultraviolet-erasable programmable read-only memories (UV EPROMs), electrically alterable read-only memories (EAROMs), electrically erasable programmable read-only memories (EEPROMs), flash EPROMs, nonvolatile static RAMs (NV RAMs), and ferroelectric memories. Most of these memory types are randomly accessible, but their main distinguishing feature is that once information has been loaded into the storage cells, the information stays there even if the power is shut off.

Some of these memories also come in specialty versions that were developed for a narrow range of applications. For instance, there are ROMs, UV EPROMs, and EEPROMs that are designed for serial interfaces to minimize chip pin count. There are also special UV EPROMs with small cachelike registers on the chip to deliver data at fast static-RAM-like speeds.

ROM. The ROM is programmed by the memory manufacturer during the actual device fabrication. Here, though, there are two types of ROMs: one is called late-mask or contact-mask programmable, and the other is often referred to as a ground-up design. In the late-mask type of ROM, the masks used in the final stages of the fabrication process determine the connections to the internal transistors. The connections, in turn, determine the data pattern that will be read out when the cell is accessed. The ground-up type of ROM is designed from the bottom up—all fabrication masks used in the multiple mask process are custom-generated—and that makes the chip slightly smaller and, in large volumes, less expensive than the late-mask version.

ROMs can be had with capacities of up to about 32 Mbits, with almost all chips available with an 8-bit word width. The increased use of 16-, 32-, and 64-bit microprocessors, however, has moved ROMs toward 16-bit-wide word organizations, and chips with such structures are available at the 1-Mbit level and above. Additionally, at such densities the cost of manufacturing the chip and the manufacturing losses due to imperfections that destroy stored data bits require that some method to correct bad bits be used. Thus, ROMs with built-in error detection and correction circuits have been put into limited production. The use of ROMs, however, has been declining since the late 1990s as the cost of flash memory devices and other nonvolatile storage options has decreased.

Fuse PROM. As an alternative to the mask-programmable memories, all the other nonvolatile memory types permit the users to program the memories themselves. The fuse PROM is a one-time programmable memory—once the information is programmed in, it cannot be altered. One type uses microscopic fuse links that are blown open to define a logic one or zero for each memory cell; the other type causes metal to short out base-emitter transistor junctions. The latter type that forms a short in the base-emitter junction connection is the predecessor to what today is referred to as antifuse technology. Antifuses are one-time programmable elements used to configure some field-programmable logic devices.

Most fuse PROMs were fabricated with bipolar transistors, and a few were implemented with complementary metal-oxide semiconductor (CMOS) technology. However, advances in CMOS technology and stored-charge-based nonvolatile memory structures have caused an almost total decline in the use of fuse-based memories. *See* TRANSISTOR.

Reprogrammable memories. The birth of the microprocessor in the early 1970s brought with it nonvolatile memory types that offered reusability. Information stored in the memory can be erased—in the case of the UV EPROM, by an ultraviolet light, and in the case of the EAROM, EEPROM, flash EPROM (often referred to as just a flash memory device), nonvolatile (NV) RAM, or ferroelectric memory, by an electrical signal. Then the circuit can be reprogrammed with new information that can be retained indefinitely. All of these memory types are starting to approach the ideal memory element for the computer, an element that combines the flexibility of the RAM with the permanence of the ROM when power is removed. *See* MICROPROCESSOR.

The ideal memory circuit should allow unlimited read and write operations without any unusual voltage levels or extra circuitry. At present, UV EPROMs require an external ultraviolet lamp to erase the stored information and a programming voltage of about 12.5 V for the newer versions, compared to a 5- or 3.3-V supply for the read-only mode. Most of the newer EAROMs and EEPROMs have been able to place control and timing circuitry on the same chip to boost the standard 5-V or 3.3-V supply voltage up to the level necessary to program or erase the cell, as well as provide the other necessary support functions. Low operating voltage is also important for nonvolatile memories used in portable systems, and many of the latest versions of these memories can operate from 3.3-V or lower power supplies.

An offshoot of the EEPROM and UV EPROM technologies is the flash EEPROM and flash EPROM. Unlike most EEPROMs, which allow the user to selectively change a single word, the flash memory permits the user to erase only the contents of the entire chip before storing new data into the chip. Some of the flash memories allow some degree of block erasure, typically one-half or one-sixteenth of the memory array. The limited electrical alterability makes possible significantly simpler circuit designs and chips that are higher in capacity, smaller in area, and thus less expensive.

Flash memories have evolved along two architectural paths. The first path consists of a standard RAM-like architecture that provides full random access to any data word within the memory chip. Such memories are based on a form of NOR-type storage cell (a few manufacturers employ an OR-type structure which is very similar) [**Fig. 8***a*]. The second architectural path consists of a disk-drive-like access scheme that incurs some system overhead to access the first byte (or word) of data specified by the address and then allows the remaining data in the desired block

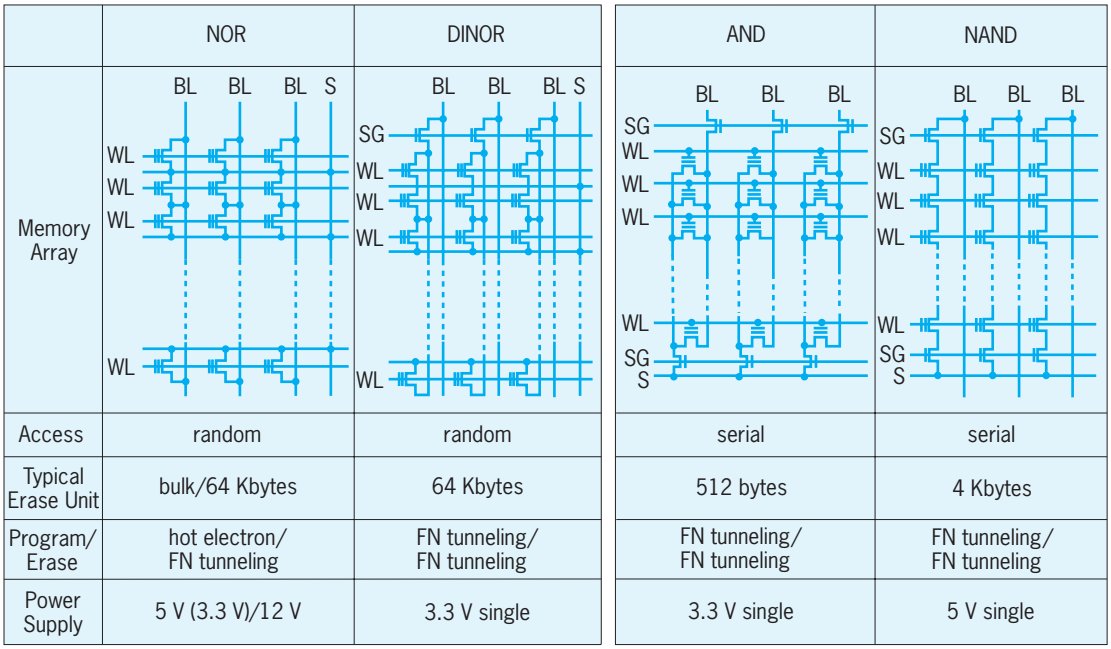


Fig. 8. Flash memory architectures have taken two major directions: (*a*) Random access. A NOR-type (occasionally, OR) memory cell structure is used at the heart of a full random-access memory architecture. (*b*) Serial access. The byte-serial architecture is based on NAND-gate-like strings of series-connected MOS transistors. In some architectures an AND array is employed instead of NAND. BL = bit line; S = select; SG = select gate; WL = word line; FN = Fowler-Nordheim. (*Mitsubishi Electronics America Inc.*)

to be read out from or written into the memory in a byte (word) serial fashion using just timing and control signals. Since this second architectural approach does not employ a full random-access architecture, the internal memory array is much denser than the NOR/OR-type array. Part of that density is attributable to the very different memory cell approach used in the heart of the arrays. *See* LOGIC CIRCUITS.

Most of the byte-serial devices employ a NAND-based memory cell in which eight transistors are serially connected such that the source of one is connected to the drain of the next, and so on (Fig. 8*b*). [Alternatively, several companies have developed serial memories that are based on an AND-connected multitransistor cell structure but at the chip level perform much like the NAND-based memories.] The serial NAND (or AND) architecture is well suited for applications that might otherwise rely on electromechanical disk drives, but can operate 10–100 times faster than the disk drives. Memory cards using such chips are in widespread use in digital cameras, personal digital assistants, solid-state dictation systems, and many other applications.

Coming closest to the ideal specification is the NV RAM, a memory that combines a static RAM with a nonvolatile memory array, so that for every stored bit there are two memory cells, one of which is volatile and the other nonvolatile. During normal system operation, the NV RAM uses the volatile memory array, but when it receives a special store signal, information held in the RAM area is transferred into the nonvolatile section. Thus the RAM section provides unlimited read and write operations, while the nonvolatile section provides back-up when power is removed.

Another method employs a ferroelectric storage element deposited on top of a static RAM cell (or even a dynamic RAM cell). The potential advantages of the ferroelectric technology include a large number of storage cycles with little increase in chip area since the ferroelectric material can be deposited on top of the memory cells. However, the technology is still in the advanced experimental stage, with only a couple of companies offering devices with relatively low densities (16 kbits to 1 Mbit).

However, the EAROM, EEPROM, flash, NV RAM, and ferroelectric RAMs suffer one common failing that keeps them from reaching the ideal—they wear out. The electrical process used to store information in the nonvolatile array causes a steady deterioration in the ability of the memory to retain data for a guaranteed time period. Advances in materials technology are improving the number of store (write) cycles that the EAROM, EEPROM, flash, and ferroelectric memories can endure. Most commodity EEPROM and flash memories, for instance, can typically endure at least 100,000 write cycles, although a large number of suppliers can offer endurances of 1 million store cycles. Ferroelectric-based memories can offer still higher numbers of store cycles, often 10–100 times higher than flash or EEPROM-type devices.

Another new storage technology disclosed in 2000 is the magnetoresistive-RAM. The nonvolatile technology used to form the memory cells consists of two

ultrathin layers (tens of angstroms) of magnetically responsive materials separated by a thin dialectric layer. The memory cells can retain data indefinitely and have extremely high endurance levels since this technology does not have a wearout mechanism such as is found in flash memories.

To overcome the limited number of storage cycles possible with the true nonvolatile memory technologies, ultralow-power-consumption static RAMs (with a standby current of just tens of nanoamperes) have been combined with some power control circuitry and a button-type long-life battery in a single package. When built into a system, this form of NV RAM operates with the characteristics of a static RAM when the system supplies power—fast access times and unlimited reads and writes. However, when the power supply line drops below a preset level, the control circuitry in the memory package detects the drop, disables the memory's ability to store new data, and then switches over to the built-in battery to keep the stored data alive until main power returns.

Semiconductor technology. The semiconductor technology used to fabricate all the different memory types spans the entire range of available commercial processes (see **table**). All dynamic and static RAMs and most nonvolatile memories are fabricated only in CMOS (and in a few cases a merged bipolar-CMOS process). The use of CMOS

Summary of major memory types, performance levels, and technologies

Memory type	Maximum capacity for commercial chips, 2001–2002	Typical access speed/cycle time	Programmability	Technology
Dynamic RAM (standard, fast page, and extended data out)	No longer in widespread use	40 ns access, 60 ns cycle	Read/write, volatile	CMOS
Dynamic RAM (synchronous, single data rate and double data rate)	256 Mbit/512 Mbit (plus limited 1 Gbit production)	20–40 ns first access, subsequent sequential access at 4–8 ns/access, 125 MHz SDR, 250 MHz DDR	Read/write, volatile	CMOS
Dynamic RAMs (Rambus)	256 Mbit	1.6 ns access, 660 MHz clock	Read/write, volatile (8- or 16-bit sequential access)	CMOS
Dynamic RAMs (Sync Link)	64 Mbit (never commercialized)	2 ns access, 500 MHz clock	Read/write, volatile	CMOS
Cache DRAM	16 Mbit	5 ns access, 200 MHz clock	Read/write, volatile	CMOS
Multibank DRAM	64 Mbit	5 ns access, 200 MHz clock	Read/write, volatile	CMOS
Video RAM	16 Mbit (being phased out)	8 ns, serial port, 50 ns, random access port	Read/write, volatile	CMOS
Synchronous graphics RAM	32 Mbit (being phased out)	5 ns access, 200 MHz clock	Read/write, volatile	CMOS
Embedded DRAM	64 Mbits	5 ns access, 200 ns clock	Read/write, volatile	CMOS
Asynchronous static RAMs (no bus latency)	16 Mbits	20 ns access, 20 ns cycle	Read/write, volatile	CMOS
Synchronous static RAMs	16 Mbits	5 ns access, 8 ns cycle	Read/write, volatile	CMOS
Synchronous burst SRAM	8 Mbits	3 ns access, 5 ns cycle	Read/write, volatile	CMOS
FIFO buffers	4 Mbits	5 ns access, 5 ns cycle	Read/write, volatile	CMOS
Mulitport SRAMs	4 Mbits	5 ns access, 5 ns cycle	Read/write, volatile	CMOS
PROM	Being phased out and replaced with electrically reprogrammable memories	10 ns access	Electrically one-time programmable, nonvolatile	Bipolar, CMOS
UV EPROMs	32 Mbits (no new development)	60 ns access	UV erasable, electrically programmable cell	CMOS
EEPROM	1 Mbit	100 ns access, less than 10 μs/word for storage	Electrically programmable and erasable on a word-by-word basis, nonvolatile	CMOS
Flash EPROM (NOR type)	256 Mbits	50 ns access, less than 10 μs/word for storage	Bulk or block erasable, programmable on a word-by-word random basis, nonvolatile	CMOS
Flash EPROM (NAND type)	512 Mbits	60–100 ns for the first access, 20 ns for subsequent sequential accesses	Bulk or block erasable, programmable in word sequential order	CMOS

memory structures (the combination of both *p*- and *n*-channel devices to form the memory cells and supporting logic on the memory chip) permits a reduction of orders of magnitude in the standby power consumption, and often a reduction by a factor of 2–5 in the active power requirements of a memory chip, as compared to the older *n*-channel MOS (NMOS) equivalent.

The speed at which dynamic RAMs can be accessed has been shortening as each generation has been introduced. Static RAM access times have also improved over the years, and CMOS static RAMs with access times of as little as 10 ns are considered commodity products. Devices with access times as short as 6 ns are also readily available, with most of them targeted at applications such as cache subsystems and high-speed signal processing. Of course, not every application demands the minimum access time, and many speed grades are available, up to several hundred nanoseconds.

For even faster access times, merged bipolar-CMOS processes and full bipolar technologies such as emitter-coupled logic (ECL) have traditionally been used to offer access times of 3–5 ns for commercial memories. However, except for a few very specialized applications, CMOS technology has been able to catch up with the performance of most bipolar memories and to offer higher capacities as well. The use of gallium arsenide processes has led to the development of memory chips with access times of less than 3 ns at density levels of 1, 4, and 16 kbits.

Access times for the various nonvolatile memories range from less than 35 ns in their read modes to about 10 μs when they store data. The flash-style EEPROMs typically have access times in the 50–100-ns range, while memory capacities have reached levels of 256 Mbits. In 2001, prototypes of 512-Mbit flash memory chips were described at several technical symposia. In addition to the basic evenly sectored architecture, specialty flash versions referred to as boot-block flash memories and special serial-file flash memories have been developed.

The boot-block approach was developed specifically for use in personal computers to hold the basic input/output software control (BIOS) programs. The address space is divided into several unequal regions, so that part of the memory remains secured while other parts of the memory can be overwritten to update the control software.

Serial file memories were developed for use as a solid-state alternative to the electromechanical hard-disk drives. File memories access data words serially and are thus optimized for use with data that come in large blocks where all the words in the block can be transferred sequentially (such as when loading a program file or transferring a large table of data). Thus, rather than allowing random access to any stored data word, the file memories are designed to operate somewhat like a disk drive. The first access to a desired block of data contains several microseconds of overhead (much like the positioning of the read/write head over the right track in a disk drive), but then sequential accesses to the rest of the data block can be done at typicial random-access-like speeds (35–100 ns per transfer). When multiple chips are combined into an array, solid-state drive alternatives with capacities of 2 to more than 500 Mbytes can be implemented in a credit-card-sized format.

Two other innovative flash-based memory architectures first appeared at the 16-Mbit density level. The first employs a dynamic RAM-like interface so that it can be used in systems that would otherwise transfer code from slow ROMs into a faster dynamic or static RAM and then execute the code from the RAM to take advantage of the shorter access times of the RAMs. The second unique architecture takes the same base 16-Mbit flash memory array, but then integrates a synchronous interface on the chip so that it appears to function much like a clocked static RAM, thus reducing the access time to about 35 ns, about half that of the standard flash memory.

A novel storage scheme that was introduced in 1994 employs a charge-sensing scheme to double the number of bits that a chip can store. The scheme divides the charge held in the nonvolatile memory cell into four levels and, through the use of on-chip voltage comparators, will differentiate the charge values to allow two bits to be stored in each cell. Thus a 32-Mbit memory chip can be built on roughly the same size chip as a 16-Mbit memory. This double-density technique has been applied to produce flash memory chips with capacities up to 512 Mbits, and still higher-capacity chips can leverage this approach as process improvements allow the fabrication of chips with more than 256 million actual storage cells.

Additionally, a unique memory chip that combines UV EPROM, on-chip static RAM, and programmable logic was released in 1988. The programmable logic allows portions of the UV EPROM to be mapped into a different portion of the system's address space. The remapping can be used to improve addressing speed or to provide security by scrambling the address. *See* COMPUTER STORAGE TECHNOLOGY; INTEGRATED CIRCUITS.

David Bursky

Bibliography. C. Hu, *Nonvolatile Semiconductor Memories: Technologies, Design, and Applications*, 1991; Institute of Electrical and Electronics Engineers, *IEEE Standard Terminology of SSemiconductor Memories*, IEEE Stand. 662–1992, 1992; B. Prince, *Semiconductor Memories: A Handbook of Design, Manufacture, and Application*, 2d ed., 1991; A. K. Sharma, *Semiconductor Memories: Technology, Testing, and Reliability*, 1996; A. J. Van de Goor, *Testing Semiconductor Memories: Theory and Practice*, 1991.

Semiconductor rectifier

A semiconductor diode that is used in rectification and power control. The semiconductor diode conducts current preferentially in one direction and inhibits the flow of current in the opposite direction by utilizing the properties of a junction formed from two differently doped semiconductor materials (**Fig. 1**). Doped silicon is by far the most widely used semiconductor. Semiconductor diodes are intrinsic

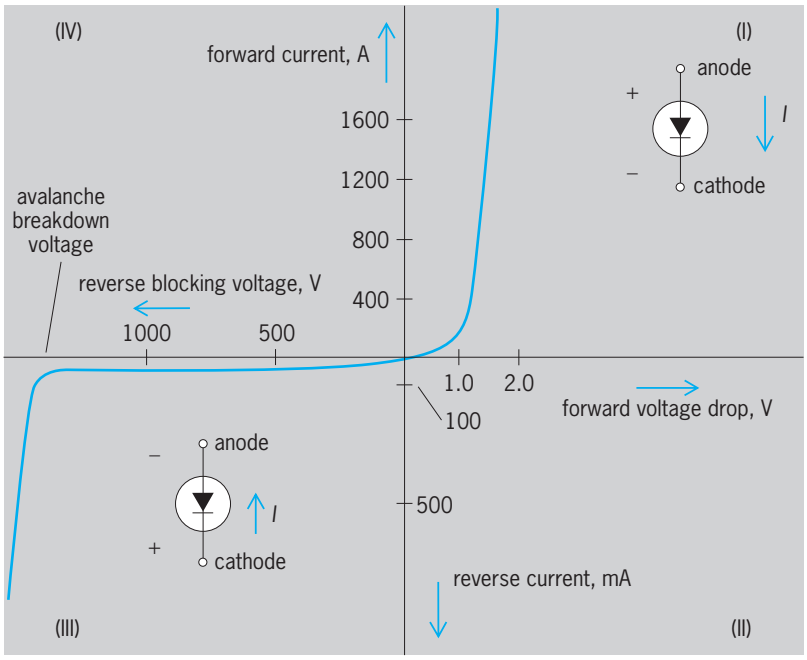


Fig. 1. Current-voltage characteristic of a commercially available 250-A silicon rectifier diode. Different voltage and current scales are used for positive and negative values.

to integrated circuits and discrete device technology and are used to perform a wide variety of isolation, switching, signal processing, level shifting, biasing, control, and alternating-current (ac) to direct-current (dc) conversion functions. This article focuses on using the semiconductor diode for ac-to-dc conversion, called rectification, as well as ac and dc power-control applications. *See* CONTROLLED RECTIFIER; RECTIFIER; SEMICONDUCTOR; SEMICONDUCTOR DIODE.

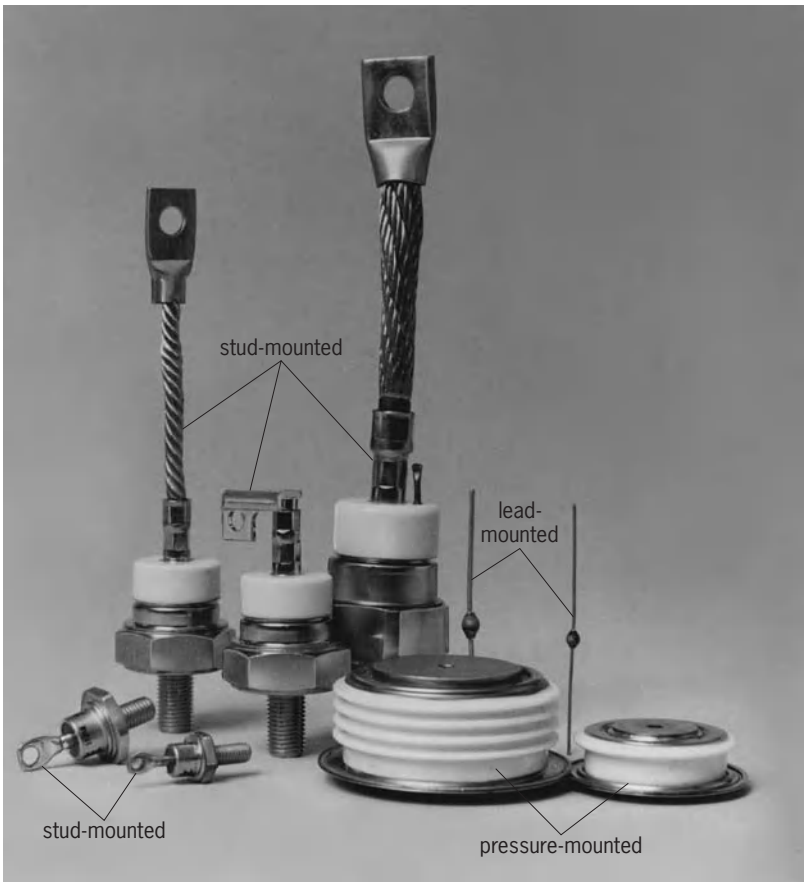


Fig. 2. Typical rectifier diode packages which allow heat dissipation and proper electrical contact. (*General Electric Co.*)

Either as a key element of an integrated circuit or as a discrete packaged part, the silicon rectifier diode is used in a plethora of applications from small power supplies for consumer electronics to very large power-rectification industrial installations. Many semiconductor diodes are used in non-power-conversion applications in signal processing and communications. These include avalanche or Zener diodes; diodes used for amplitude-modulation radio detection, mixing, and frequency translation; IMPATT, PIN, and step-recovery diodes, used at microwave frequencies; diodes fabricated from gallium arsenide and related compounds, used in optoelectronics; and light-emitting diodes (LEDs) and solid-state lasers. *See* AMPLITUDE-MODULATION DETECTOR; LASER; LIGHT-EMITTING DIODE; MICROWAVE SOLID-STATE DEVICES; MIXER; ZENER DIODE.

Silicon rectifier diodes. The electrical heart of the semiconductor diode is the junction between *p*-type and *n*-type doped silicon regions. The primary processing technology used to form a *pn* junction is high-temperature diffusion. In the diffusion process, a chemically very pure wafer of single-crystal silicon is exposed to a hot vapor of a dopant material. *See* JUNCTION DIODE.

The die, as a portion cleaved from a wafer, has a thickness on the order of 0.5 mm for a 100-mm-diameter (4-in.) wafer. For integrated circuit diodes, the current levels are low, in the range of a few microamperes, so that the junction dimensions are as small as a few micrometers on a side. For high-power applications rated in the thousands of amperes, the die could be as large as the entire wafer. The final die dimension is determined predominantly by current-handling requirements and packaging considerations. Advanced semiconductor processing technology utilizes 150-mm (6-in.) and even 200-mm (8-in.) wafers, although these are not used in whole-wafer form for semiconductor rectifiers because of economic factors. The die must be mounted in a suitable hermetically sealed protective package which allows the heat to be dissipated to the external environment and allows for proper electrical contact. Usually packages are lead-mounted for currents up to a few amperes, stud-mounted and flat-base mounted for currents up to a few hundred amperes, and external-pressure-mounted for the largest devices capable of conducting thousands of amperes (**Fig. 2**). The lead-mounted package dissipates the losses directly into the environment, while the other package styles are mounted on heat sinks for natural-convection-air, forced-air, or liquid cooling. *See* INTEGRATED CIRCUITS.

Discrete silicon diodes are commercially available with forward-current specifications from under 1 A to several thousands of amperes. Diodes may be connected in parallel for greater current capability as long as the design provides for the current being uniformly distributed between the parallel diodes.

This is usually done with a ballast resistor in series with each diode. *See* BALLAST RESISTOR.

Ideally, the current through a reverse-biased diode, called the saturation current (I_S) or reverse current (I_R), approaches zero (Fig. 1). Practically speaking, this current is several orders smaller than the forward current (I_F). The maximum value of the reverse blocking voltage is limited primarily by the structure and doping of the semiconductor layers. (The package design must be appropriate for the maximum design voltage.) This maximum voltage is referred to as the avalanche breakdown voltage, or the peak reverse voltage (PRV) or peak inverse voltage (PIV). It is a very important parameter for power supply and power conversion designs. Exceeding the peak inverse voltage is usually destructive unless the circuit design provides for limiting the avalanche current and resultant heating. In summary, at positive voltages and currents (quadrant I of the voltage-current characteristic; Fig. 1), the silicon rectifier diode shows the on-state conducting characteristic, with high current and low forward voltage drop; at negative voltages and currents (quadrant III), it shows the reverse-blocking or reverse-bias, off-state characteristic, with high blocking voltage and low (ideally zero) reverse blocking current. *See* ELECTRICAL BREAKDOWN.

Integrated-circuit diode-junction avalanche breakdown voltages are of the order of several tens of volts. Single silicon rectifier diodes designed for power conversion applications are available with ratings from a few hundred to a few thousand volts. Several diodes can be connected in series for greater voltage capability. Prepackaged series diode strings can be rated to tens of thousands of volts at several amperes. This series connection must ensure equal voltage division across each diode to guard against catastrophic failure of the entire series. Typically this is done by including a high-value equal-value resistor in parallel with each diode to obtain equal voltages, and a parallel capacitor to provide a low-impedance path for high-voltage transients that are often present in industrial environments.

Schottky silicon rectifier diodes. Unlike a silicon diode formed from a *pn* junction, the Schottky diode makes use of the rectification effect of a metal-to-silicon interface and the resultant barrier potential. The Schottky diode, sometimes called the Schottky-barrier diode, overcomes the major limitation of the *pn* junction diode; being a majority carrier device, it has a lower forward voltage drop (0.2–0.3 V, compared to 0.7–1.0 V) and faster switching speed than its minority-carrier *pn* junction counterpart. However, other factors confine its use to low-voltage power applications, chiefly the relatively small breakdown voltage, typically 45 V. Secondary shortcomings include a high reverse current and restricted temperature of operation, with commercial devices providing a maximum of 175°C (347°F) compared with 200°C (392°F) for *pn* junction diodes.

Integrated circuits used in computer and instrument systems commonly require voltages less than 15 V and as low as 3.3 V. Thus the advantage of low forward-voltage drop and faster switching favors the Schottky diode. This is particularly true for high-frequency switching voltage regulator power supply applications where voltages at 20–50 kHz must be rectified. The higher reverse current can be tolerated. However, cooling or heat sinking is more critical because of the higher reverse-current temperature coefficient and lower maximum operating temperature. *See* SCHOTTKY BARRIER DIODE.

Rectifier circuits. The greatest usage of rectifier diodes is the conversion of ac to dc. The single diode of a half-wave rectifier for a single-phase ac voltage (**Fig. 3***a,b*) conducts only on the positive half-cycle.

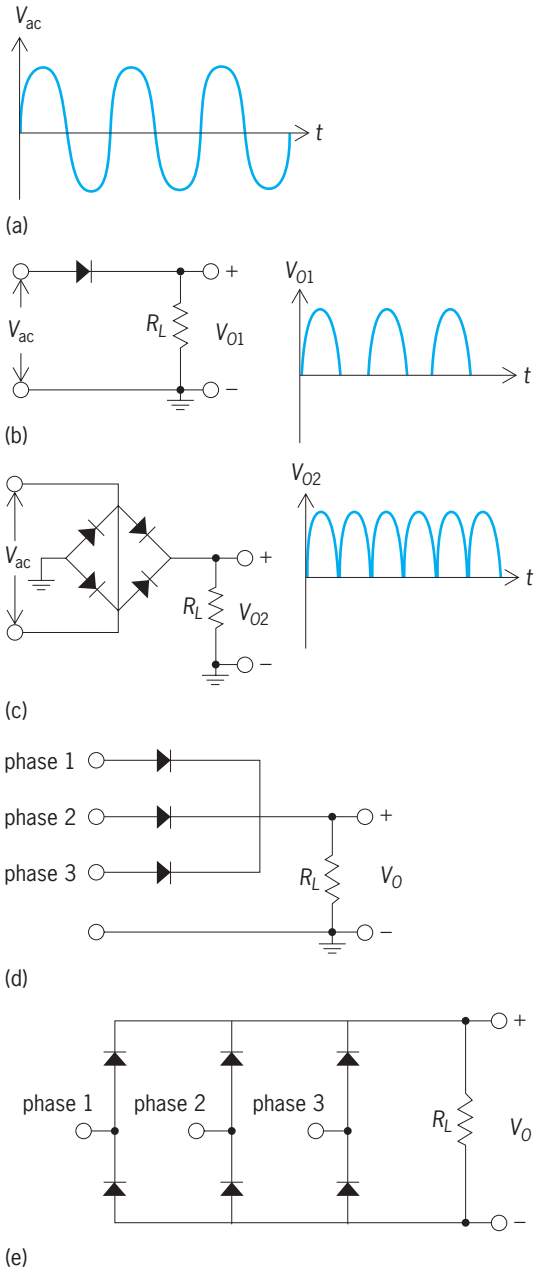


Fig. 3. Rectifier circuits. (*a*) Waveform of single-phase ac input voltage, V_{ac}, for both half-wave rectifier and full-wave rectifier circuits. (*b*) Half-wave rectifier circuit, showing rectification of input on the positive half-cycles to a pulsating dc output voltage, V_{O1}. R_L = load resistance. (*c*) Full-wave bridge rectifier circuit, showing rectification of input on both the negative and positive half-cycles to a pulsating dc output voltage, V_{O2}. (*d*) Half-wave and (*e*) full-wave rectifier circuits for three-phase ac inputs.

Because of this, the output voltage across the load resistance is unidirectional and has a nonzero average value. This output waveform is called a pulsating dc. Therefore the input ac voltage has been rectified to a dc voltage. For most applications, a filter, usually consisting of large electrolytic capacitors, must be employed at the output to smooth the ripple present on the pulsating dc voltage to come close to a constant dc voltage value. *See* CAPACITOR; ELECTRIC FILTER; ELECTRONIC POWER SUPPLY; RIPPLE VOLTAGE.

A widely used diode circuit for rectification of both the positive and negative half-cycles of the single-phase ac supply voltage is the full-wave bridge rectifier (Fig. 3*c*). It is commercially available as a single package. The output voltage is still a pulsating dc voltage, but the average value will be double that of the half-wave rectifier.

Many high-power industrial applications require the use of three-phase power (Fig. 3*d,e*). The output waveform exhibits less ripple voltage and is easier to filter. *See* ALTERNATING CURRENT.

In lower-power applications from a few watts to a few hundred watts, such as used in computers, television receivers, and laboratory instruments, a switching voltage regulator is commonly used to generate a 10-kHz–50-kHz ac signal from the high-ripple ac power supply voltage. The advantage is the ease and lower cost in filtering the ripple resulting from rectifying high-frequency ac as opposed to filtering low-frequency ac. *See* VOLTAGE REGULATOR.

Thyristors. Whereas the basic semiconductor rectifier has two terminals, an anode and cathode, a silicon controlled rectifier (SCR) has three terminals: an anode, cathode, and control electrode called the gate (**Fig. 4**). The silicon controlled rectifier is a four-layer device modeled as two interconnected *pnp* and *npn* transistors.

Normally, there is no current flow from the anode to the cathode. Both transistors are off; that is, they are blocking any current flow. By applying a relatively small trigger pulse control signal to the gate electrode, the *npn* transistor is switched on. When the *npn* transistor is switched on, the *pnp* transistor is also switched on. Consequently the silicon controlled rectifier is turned on and a current flows through the silicon controlled rectifier and external circuit. The resultant internal voltages keep both the *npn* and *pnp* transistors on even when the gate voltage is removed. The device is said to exhibit regenerative, positive-feedback, or latching-type switching action. There is a voltage drop of about 1 V across the on-state silicon controlled rectifier. The power dissipation rating required in specifying a silicon controlled rectifier is given by this 1-V drop multiplied by the peak current flowing through the device. *See* TRANSISTOR.

Current continues to flow even when the gate signal is reduced to zero (Fig. 4*d*). To reset the silicon-controlled rectifier, the external current must be reduced below a certain value. Thus, the thyristor can be switched into the on state (conducting condition) by applying a signal to the gate, but must be restored to the off state by circuit action. If the anode current momentarily drops below some holding current or if the anode voltage is reversed, the silicon controlled rectifier reverts to its blocking state and the gate terminal regains control. Typical silicon controlled rectifiers turn on in 1–5 microseconds and require 10–100 μs of momentary reverse voltage on the anode to regain their forward-blocking ability.

Other semiconductor diode topologies are also used for power control. A generic term for these power-control devices is the thyristor (**Fig. 5**).

The silicon controlled rectifier operates only in the first quadrant (forward current and voltage), that is, on either the positive or negative half-cycle of

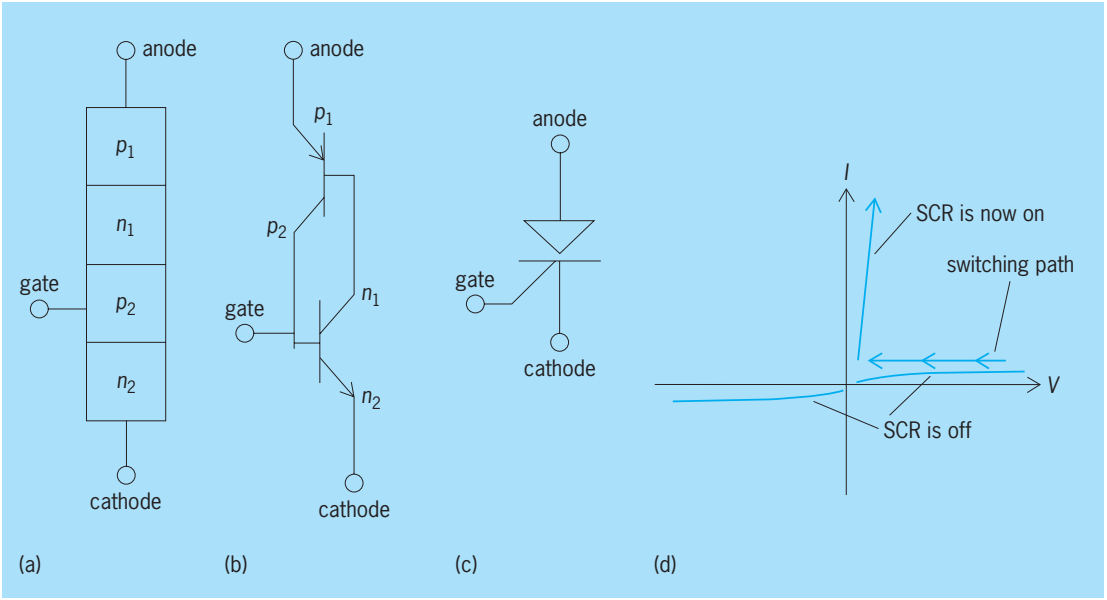


Fig. 4. Silicon controlled rectifier (SCR). (*a*) Four-layer structure. (*b*) Interconnected *npn-pnp* transistor model. (c) Preferred schematic symbol. (*d*) Current-voltage (*I-V*) characteristic. The switching path is followed when a voltage pulse is applied to the gate electrode of the *npn* transistor, switching the SCR to on.

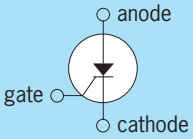
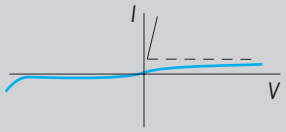
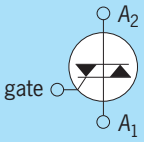
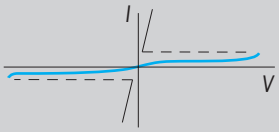
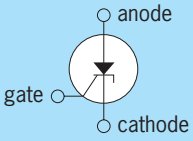
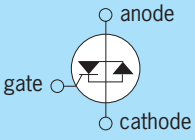
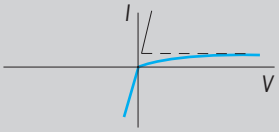
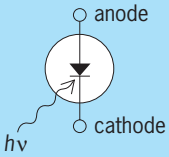
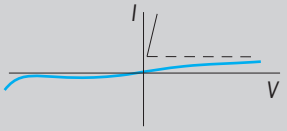
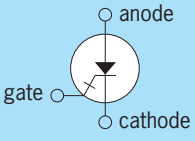
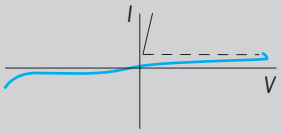
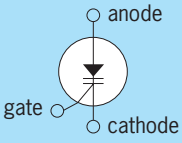
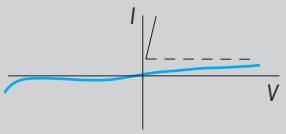
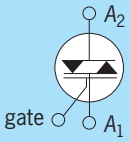
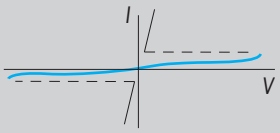
Type of thyristor	Common circuit symbol	Current–voltage (I–V) characteristics
reverse-blocking triode thyristor (SCR)		
bidirectional triode thyristor (TRIAC)		
asymmetrical triode thyristor (ASCR)		
reverse-conducting triode thyristor (RCT)		
light-activated SCR (LASCR)		
commutable reverse-blocking triode thyristor (GTO)		
MOS controlled SCR (MOS SCR)		
MOS controlled thyristor, TRIAC (MCT)		

Fig. 5. Family of thyristor devices.

the ac voltage. For specialized ac-switching power control such as lamp dimmers and variable heating where bidirectional control is required, the bidirectional triode thyristor, called the TRIAC, is widely used. Its current-voltage characteristics (Fig. 5) are the same in the first and third quadrants as the first-quadrant silicon controlled rectifier characteristics. Functionally, the TRIAC behaves like two oppositely oriented silicon controlled rectifiers connected in parallel. The asymmetrical triode thyristor (ASCR), reverse conducting triode thyristor (RCT), and commutable reverse-blocking triode thyristor (GTO) are used in power circuits requiring operation from a dc source.

The asymmetrical triode thyristor has a much lower breakdown voltage (Fig. 5), which presents no problem in dc circuits because there is no need to block a large reverse voltage. The asymmetrical triode thyristor has a faster turn-off time, 5–25 μs, thus requiring a less costly auxiliary circuit to effect turn-off. The reverse conducting triode thyristor is the integrated equivalent of a discrete asymmetrical triode thyristor and includes a parallel discrete diode rectifier (Fig. 5). This provides a lower breakdown voltage. The reverse conducting triode thyristor also has a shorter turn-off time than the silicon-controlled rectifier. In general, both the asymmetrical triode thyristor and the reverse conducting triode thyristor have somewhat lower forward-voltage drops for comparable forward-voltage blocking ratings and silicon die area, thus decreasing the device power dissipation and increasing efficiency.

The gate trigger signal can also be generated optically. If a light pulse is allowed to shine on the second *pn* junction (p_2-n_2; Fig. 4*a*) of a light-activated silicon-controlled rectifier (LASCR), the device will be switched to the on state (Fig. 5). This type of device is used where electrical isolation between the gate-control circuit and output circuit is required as well as to detect the presence of opaque objects in a process assembly line. Its characteristics are similar to a standard silicon-controlled rectifier.

The gate turnoff device (GTO) is a symmetrical reverse-blocking triode thyristor which can block reverse voltages (Fig. 5). In addition, it can turn off anode current when a negative voltage is applied to the gate. Thus it does not require an auxiliary circuit to commutate it off as do the silicon controlled rectifier, asymmetrical triode thyristor, and reverse conducting triode thyristor devices.

All of the thyristors discussed so far are based upon switching a *pn* junction. The trigger circuit must provide adequate current to forward-bias the *pn* junction. A thyristor specification will include the gate turn-on voltage (V_{GT}) as well as the gate turn-on current (I_{GT}). In order to switch a thyristor from metal-oxide-semiconductor field-effect transistor (MOSFET) logic circuits or from low-power bipolar circuits, a circuit must be included in the gate trigger that can interface between the low-current drive capability of the switch-control circuitry and the higher-current requirements for switching on the thyristor. By including negative MOS (NMOS) technology in the thyristor fabrication (**Fig. 6**), the silicon-controlled rectifier can be switched from a voltage source that does not need to provide any significant current. The MOS controlled silicon controlled rectifier (MOS SCR) has high input impedance and requires a very low gate drive current (Fig. 5). Another type of MOS controlled thyristor (MCT) has a TRIAC-type characteristic (Fig. 5). MOS thyristors can be switched with a gate turn-on voltage of 15 V and control current densities of 1000 A/cm^2.

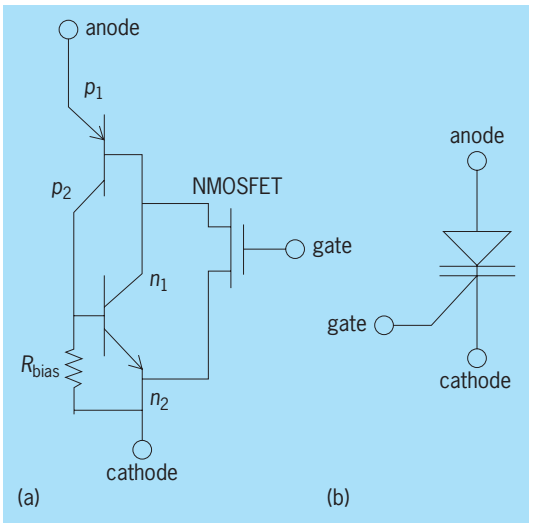


Fig. 6. MOS-controlled thyristor. (*a*) Model as interconnected *pnp* and *npn* transistors gated by an NMOSFET (negative metal-oxide-semiconductor field-effect transistor). (*b*) Schematic symbol.

Forward current ratings for a single thyristor range from 1 to 5000 A, and peak reverse voltages may exceed 4000 V. Thyristor circuit topologies are found within many types of voltage-regulator integrated circuits and are used for short-circuit protection.

Thyristor circuits and applications. Thyristor applications fall into two general categories. The devices can be used from an ac supply, much like silicon rectifier diodes. However, unlike the rectifier diode, which conducts load current as soon as the anode voltage exceeds about 0.7 V, the thyristor will not conduct load current until it is triggered into conduction.

Therefore, the power delivered to the load can be controlled. This mode of operation is called ac phase control. It is extensively used in applications requiring conversion from ac to variable-voltage dc output, such as adjustable-speed dc motor drives, and in lighting and heating control. *See* DIRECT-CURRENT MOTOR.

The other category of applications is operation in dc circuits. This allows power conversion from a battery or rectified ac line to a load requiring either an alternating supply (dc-to-ac conversion) or a variable-voltage dc supply (dc-to-dc conversion). Since the rate of switching the thyristors in dc circuits can be varied by the control circuit, a thyristor inverter circuit can supply an ac load with a variable frequency. The fundamental approach in both cases is to

convert a dc voltage to a chopped voltage of controllable duty cycle. Changing the duty cycle either at a variable rate (frequency power modulation) or by varying the pulse width at a fixed frequency (pulse-width power modulation) effectively controls the power delivered to the load. *See* CHOPPING; PULSE MODULATION.

The major distinguishing feature of dc operation is the absence of any reversal of supply voltage polarity for turning off the thyristor, which would allow the gate electrode to regain control of the device. Either a gate turnoff device must be used, or an auxiliary circuit must be employed to effect turnoff of a silicon-controlled rectifier. One common technique to accomplish this is to switch a previously charged capacitor across the load-carrying silicon-controlled rectifier in such a manner that the voltage on the capacitor reverse-biases the silicon-controlled rectifier sufficiently to reduce the load current through it to zero, and then to allow the device a short time (about 10–50 μs) before reapplying the forward blocking voltage to the anode.

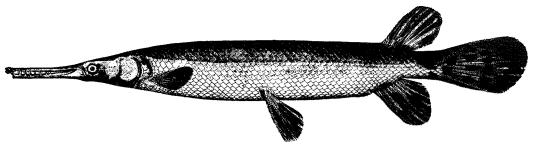
Important applications for dc-to-dc conversion, dc-to-ac power conversion at variable frequency, and dc-to-ac power conversion at fixed frequency are, respectively, control of battery-powered industrial vehicles such as forklift trucks and mining locomotives, adjustable-speed operation of ac synchronous and induction motors in industrial processing, and power transmission conversion. *See* ALTERNATING-CURRENT MOTOR; CONVERTER; INDUSTRIAL TRUCKS.

Stanley G. Burns

Bibliography. B. J. Baliga, *Power Semiconductor Devices*, 1996; *Electronic Engineer's Master Catalog (EEM)*, 4 vols., annually; J. T. Humphries and L. P. Sheets, *Industrial Electronics*, 4th ed., 1993; B. G. Streetman, *Solid State Electronic Devices*, 5th ed., 1999; A. Vladimirescu, *The SPICE Book*, 1993.

Semionotiformes

An order of actinopterygian fishes which appeared first in the upper Permian, reached maximum development in the Triassic and Jurassic, and persists in the Recent fauna as the gars (*Lepisosteus*). The gars are ranked by some as a separate order, called variously Lepisosteiformes, Lepisostei, and Ginglymodi. In the Semionotiformes the body is encased in a heavy armor of interlocking ganoid scales, which are thick, are more or less rhomboidal, and have an enamellike surface. The body form is diverse, and such Mesozoic forms as *Dapedius*, *Lepidotus*, and *Semionotus* were deep-bodied or stocky fishes with a blunt or moderate snout. Modern forms, however, are elongate and have bony jaws provided with enlarged conical teeth. The opposed dorsal and anal fins are short and placed far back on the body, and the caudal fin is abbreviated heterocercal. The vertebrae are distinctive in being opisthocoelous. There are paired prevomers and no gular plates, and the swim bladder is highly vascularized, thus permitting aerial respiration. The single Recent family, Lepisosteidae,



Spotted gar (*Lepisosteus oculatus*). (*After G. B. Goode, Fishery Industries of the United States, 1884*)

contains one genus, *Lepisosteus*, with seven species (see **illus.**). Although fossils are known from all continents, *Lepisosteus* is now restricted to lowland fresh and brackish waters of North and Central America.

Gars are despised by fishers because of their predation on valuable fishes. They are themselves caught with difficulty and are little esteemed as food. The roe is highly toxic when eaten by warm-blooded animals. *See* ACTINOPTERYGII; SWIM BLADDER.

Reeve M. Bailey

Bibliography. H. B. Bigelow (ed.), *Fishes of the Western North Atlantic*, Sears Found. Mar. Res. Memo. 1, pt. 3, 1968.

Sendai virus

A member of the viruses in the type species Parainfluenza 1, genus *Paramyxovirus*, family Paramyxoviridae; it is also called hemagglutinating virus of Japan (HVJ). Sendai virus was originally recovered in Sendai, Japan, from mice inoculated with autopsy specimens from newborns who died of fatal pneumonitis in an epidemic in 1952. Subsequent attempts to isolate this virus from humans were, however, mostly unsuccessful, although mice are commonly infected with Sendai virus along with rats, guinea pigs, hamsters, and pigs. It is believed that the natural host of Sendai virus is the mouse and that the virus is usually nonpathogenic for humans.

A virus serologically related to Sendai virus, named HA-2 virus, was isolated from children with respiratory illness in 1958. This virus is classified in the same type species as Sendai virus, and has been established as a common causative agent of croup in children.

Sendai virus has a remarkable ability to fuse cells. Cell fusion induced by ultraviolet-ray-inactivated Sendai virus is now widely used by investigators in many areas of cell biology, somatic cell genetics, virology, immunology, and oncology. It enables researchers to produce heterokaryons, hybrid cells, or macromolecule-introduced cells whose novel properties provide valuable insight into numerous aspects of cellular organization. *See* SOMATIC CELL GENETICS.

Under electron-microscopic observation, the shape of the Sendai virus particle is roughly spherical, but the diameter varies from 160 to 600 nanometers. The particle has an envelope of lipid bilayer, covered with several hundred spikes 12 nm in length, in which a helical ribonucleoprotein nucleocapsid of 18 nm diameter with a regular periodicity of 5 nm is enclosed.

The genome is a single-stranded ribonucleic acid

(RNA) molecule with a molecular weight of 50,000. The RNA is a negative strand; that is, it is not itself a messenger for viral protein synthesis, but functions as a template for the synthesis of six messenger RNA (mRNA) species in the cytoplasm of infected cells. Six viral structural proteins, as well as one nonstructural protein, are translated from the six mRNAs. *See* ANIMAL VIRUS.

Nakao Ishida

Bibliography. N. Ishida and M. Homma, Sendai virus, *Adv. Virus Res.*, 23:349–383, 1978; T. Shioda et al., Sequence of 3,687 nucleotides from the 3′ end of Sendai virus genome RNA and the predicted amino acid sequences of viral NP, P and C proteins, *Nucl. Acids Res.*, 11:7317–7330, 1983; T. B. Shows and A. Y. Sakaguchi, Gene transfer and gene mapping in mammalian cells in culture, *In Vitro*, 16:55–76, 1980; M. Yamaizumi et al., Rapid transfer of nonhistone chromosomal proteins to the nucleus of living cells, *Nature*, 273:782–784, 1978.

Sensation

The obtaining of information from the environment through the agency of the sense organs. Sensation commonly includes the interpretation of that information as a particular attribute of the world or self.

Sensation is the intermediary between reception and perception, but the boundaries are vague. Reception refers to the absorption of some form of energy (a stimulus) from the external or internal environment through a process called transduction. Perception includes the interpretation of that stimulus to generate a representation of the world. Within the nervous system, the interpretation is generated by multiple feedback loops, so there is no distinct point at which sensation ends and perception begins.

Sometimes there is sensation without perception. Brain damage can lead to a syndrome called blindsight, in which a person who is apparently blind can nevertheless indicate the general direction in which an object is moving. In one case, a patient who could not tell the position or orientation of a slot could nevertheless "post" a card into the slot with nearly normal accuracy. To some extent, normal individuals demonstrate sensory awareness without perception, as when making an abrupt catch. *See* BRAIN; COGNITION; PERCEPTION.

The senses. It is common to speak of the five senses (sight, hearing, taste, smell, and touch), but the senses also include hot and cold, pain, limb position, acceleration, and internal states such as hunger, thirst, and fatigue. Some senses have dedicated organs, such as the eyes, ears, tongue, or nasal mucosa, while others are more diffuse, with receptors throughout the skin or within other organs. Each sense has a specific pathway to the brain and primary areas within the brain where the signals are analyzed.

Sense receptors. Sensation is initiated by the stimulation of receptors. A receptor is a specialized cell that can absorb energy from its environment and signal its presence to other neurons. This is accomplished by a change in electrical potential of the receptor, resulting in a change in the amount of chemical transmitter released to secondary neurons. *See* NEUROBIOLOGY; NEURON.

The process of converting an external stimulus to a neural signal varies across receptor types, but each receptor has a preferred type of stimulus. The receptors in the eye are sensitive to light; they contain a chemical that absorbs photons, initiating a chain reaction that changes the electrical potential by closing the channels that allow ions to cross the cell membrane. Receptors in the ear respond to mechanical vibrations by bending their protruding cilia, opening channels that allow ions to cross the cell membrane. Cells in the tongue and nasal mucosa have receptor sites that bind specific molecules dissolved in the fluids surrounding them, leading to chain reactions that affect their electrical potential. Receptors in the skin and joints respond to mechanical stretch or pressure, or to hot or cold. Many of the channels that allow ions to enter and change the cell potential are members of a superfamily called transient receptor potential (TRP) channels. *See* BIOPOTENTIALS AND IONIC CURRENTS; SYNAPTIC TRANSMISSION.

Since each receptor type is tuned to a specific type of energy, the brain interprets a signal from any given receptor as arising from that form of energy. If a receptor is stimulated by an inappropriate stimulus, it will be misinterpreted. That is why a blow to the eye is seen as a flash of light, called a phosphene. Similarly, punctate pressure on the skin of the back of the hand at some positions causes a sensation of cold; this is where a cold receptor happens to be located.

Sensory receptors that transmit information about the body are found dispersed throughout the skin, muscles, and viscera. Receptors that sense the internal states of the body are generally in the viscera and are called interoceptors. These provide a general feeling of such states as hunger, nausea, distention, chill, or heat. Receptors that sense the position and motion of the body are called proprioceptors. They are generally found in the skin, muscles, and tendons. The receptors that transmit information about the external environment are called exteroceptors. Among these are the receptors for the sense of touch, known as proximoceptors; they provide information about what comes in contact with the body. Other exteroceptors transmit information about the world distant from the body, and are called teleceptors. *See* PROPRIOCEPTION.

Sense organs. Telereceptors are typically concentrated in special organs that channel the appropriate energy to the receptors. Often, the way these organs present the stimuli is a first essential step in the sensory process.

The organ of vision is the eye. The principal function of the cornea and lens of the eye is to focus an image of the visual scene upon the retina, the neural tissue at the back of the eye that includes the visual receptors. The retina signals the pattern of stimulation upon the receptor array. But the eyes also control what part of the scene is analyzed, as they are aimed at visual targets. The contractile iris of the eye also exerts some control over the amount of light presented to the retina, and the lids can shut out the image entirely. *See* EYE; VISION.

The organ of hearing is the ear. The ear also serves a focusing function; the mechanical properties of the inner ear cause different receptor cells to receive different frequencies of sound, as part of the process of analyzing tones. In many animals, the outer ear can be aimed at specific sound sources; humans turn their heads and cup a hand behind the ear to get this effect. *See* EAR; HEARING.

The sense of balance depends on the vestibular organ, an extension of the inner ear. The orientation of the semicircular canals of the labyrinth causes the fluids in them to circulate for particular rotational accelerations, allowing the receptors in each to be specific for those directions of motion. *See* POSTURAL EQUILIBRIUM; VESTIBULAR SYSTEM.

Common properties of the senses. Despite the wide variety of receptor types serving the various senses, and the very different functions each sense serves, there are common features that characterize sensation and sensory processing.

Multiple receptor types. Each sense is served by several different receptor types, often tuned to different aspects of the stimulus. In some cases, this allows coverage of a wider range of stimuli; in many cases, the differences in tuning are essential to be able to discriminate among stimuli of a particular type.

For example, in the vertebrate visual system, there are two receptor types: rods and cones. Rods are more sensitive than cones, being able to be stimulated by a single photon; cones cannot detect such dim light, but extend the range to far greater energies than the rods could handle. With the luxury of more light, cones can afford to be faster and more spatially selective, permitting faster and more accurate vision in brighter lights. In addition, there are three different types of cones in the normal human retina. Each type is most sensitive to a somewhat different range of the visible spectrum. Comparisons among the signals from these cone types is the basis for color vision.

The auditory system has two different types of receptor, the inner and outer hair cells. Like the rods and cones of vision, outer hair cells are more sensitive to weak stimuli than are the inner hair cells.

The olfactory (smell) and gustatory (taste) receptors come in a variety of types, with sensitivity to different chemicals. In the olfactory system, there are apparently thousands of different receptor sites with different distributions on individual receptor cells. *See* OLFACTION; TASTE.

Within the skin, there are nerve endings specific to sudden touch, and others that are sheathed in a fatty capsule so that they respond more slowly to sustained pressure. Pain receptors respond to the chemical products of cell damage, but interact with other receptors to indicate the kind of damage (for example, heat, pressure). Within the muscles, there are two different types of nerve endings on the receptive organ (a special kind of muscle fiber) that are specialized to signal either rapid or sustained lengthening of the muscle; another sensor in the tendons signals potentially damaging strong tension. *See* CUTANEOUS SENSATION; MECHANORECEPTORS; PAIN; PROPRIOCEPTION.

Adaptation. Most sensory systems (certainly all the exteroceptor systems) adapt to the general level of stimulation. By doing so, they concentrate on what is different from the general environment. The process of adaptation is slower than most responses, so one does not adapt to, and thus ignore, the stimuli of interest.

As an example, consider visual light and dark adaptation. If one steps into a darkened theater on a bright afternoon, it is impossible to see the way to an empty seat. After a few minutes, however, fellow theatergoers may be seen with ease. In fact, full dark adaptation in total darkness is not complete for nearly an hour. At the end of that time, sensitivity is such that one can see a single candle at 4 mi (6.4 km). The reverse occurs when one leaves the theater. The sunlight is blinding for a few moments, but then can be tolerated.

The other senses also adapt to the presence or absence of strong stimuli. A strong odor (if not too noxious) fades with time; you do not notice your own perfume or cologne. Gentle pressure is noticed only when first applied; you are not aware of the pressure of your clothing. The shock of jumping into cold water soon dissipates.

Processing to recognize changes. Much of early sensory processing involves comparisons. Localization of a stimulus depends on determining if there is more energy at a particular location than at others. Other aspects may be determined by a similar comparison, which may be thought of as localization in a dimension other than space.

In the visual system, each cell has a receptive field, an area within which light causes a response from that cell. Most receptive fields in early processing (within the retina and beyond) exhibit spatial antagonism; that is, light in a small area centered in the field has one effect, while light in the annular region surrounding that central region has the opposite effect. Thus, the greatest response occurs when the light is concentrated within the center; broader illumination decreases the effectiveness of the light in the center. In this way, the cells are tuned to the contrast of a stimulus—local light compared to the general illumination—and the greatest responses occur where there is the most contrast. The cells are thereby tuned to respond to edges and borders.

Another comparison is made between cones with different spectral properties. Cones responsive to long-wave light (red or orange) are somewhat responsive to all wavelengths, and so would not be able to signal whether a light was dim and red or bright and green (medium-wave light). But by differencing the responses from cones that prefer the green light (middle wavelengths), a positive response can be generated only when the long-wave-sensitive cones have a larger response than the middle-wave-sensitive cones, that is, for a long wavelength (red or orange) light. This processing narrows the response range of subsequent cells, making them more selective. Sensory processing further along in the visual pathway can then use these refined signals to distinguish colors. *See* COLOR VISION.

A similar comparison is made in the cochlear nucleus between cells responding to different locations in the inner ear. Since the different frequencies of sounds are maximally effective at different locations, this sharpens the tuning for specific frequencies.

Psychophysics. Psychophysics is the study of relations existing between energy in the environment and attributes of sensation. Changes of a particular type of sensation are often associated with changes of a particular aspect of the stimulus. For example, changing the spectral characteristics of a light affects the perceived color, while changing the overall energy is associated with a change of brightness. Similarly, changing the frequency of a sound changes the pitch, while changing the amplitude results in a change of loudness. However, these attributes are not independent. Changing the wavelength of a light also affects the brightness; changing the frequency of a sound also affects the loudness. Similarly, changing the energy in a light also affects the color; a more intense red light looks orange, while more intense green lights take a yellowish tinge (this is known as the Bezold-Brücke hue shift). Increasing the amplitude of a sound also affects the pitch, which is why stereo systems boost the bass for lower volumes. In addition, any extremely strong stimulus is interpreted as painful. *See* PSYCHOPHYSICAL METHODS.

Sensitivity and threshold. A primary task of psychophysics is to determine the minimum energy of a stimulus that can be detected, a level called the absolute threshold. Thresholds with different stimulus attributes can indicate the underlying mechanisms of a sensory system. Electrophysiologists who record the activity of neurons in a sensory pathway generally cannot define an absolute threshold, but prefer to find the energy level that produces a particular level of response. The inverse of that energy level is defined as sensitivity. Psychophysicists generally concur that absolute threshold is an arbitrary measure that depends on the subject's criterion; it is now common to determine instead detectability using signal detection theory. *See* SIGNAL DETECTION THEORY.

One application of determining sensitivity is to find different combinations of stimulus attributes that produce the same responses. For example, if a brief flash of light persists longer, it requires less intensity to have the same effect, up to some maximum duration (Bloch's law). That maximum time indicates the temporal summation properties of the system. If a stimulus spot is made larger, less intensity is required for the same effect, up to some maximum diameter. That maximum size indicates the spatial summation properties of the system (Ricco's law). These laws represent the static invariances.

Detection of differences. A revealing determination is the magnitude of the change in stimulus needed for a subject to detect that there is a difference. The requisite change generally is in direct proportion to the magnitude of the stimulus being compared. This is Weber's law, usually stated as $(\Delta I)/I_0 = k$, where ΔI is the change in stimulus energy, I_0 is the energy of the comparison, and k is Weber's fraction, which depends upon the particular sensory system and other attributes of the stimulus.

Scaling. Scaling asks a subject "how big?" is a stimulus. In general, the apparent strength of a stimulus does not grow in direct proportion to the physical energy in the stimulus. Rather, the sensation grows less rapidly for stronger stimuli, as one would expect from Weber's law. What this means is that all changes are relative.

Sensory pathways. Most sensory systems have complex pathways, involving multiple brain areas. Understanding how these systems work means determining the roles of these various brain areas.

Electrophysiologists examine the firing of cells in various brain areas when specific stimuli are presented. This work is generally done in animal models, on the assumption that the basic principles also apply to humans. These studies indicate the specific kinds of responses and cell types that participate in sensory processing.

Brain imaging has now made it possible to see which areas of the human brain are active during specific sensory tasks. There are two major kinds of imaging in use: positron emission tomography (PET) scans and functional magnetic resonance imaging (fMRI) scans. PET scans involve the injection of radioactive compounds and observation of the areas from which radioactivity is emitted to determine the specific chemicals involved in sensory processing pathways. MRI is a technique in which a strong magnetic field causes atoms to align. Subsequent stimulation with a high-frequency pulse temporarily disrupts the alignment, and the radio signals emitted by atomic realignment can then be detected and translated to produce anatomical images. In fMRI, the specific signals sought are from oxygenated hemoglobin. This blood oxygen level detection (BOLD) signal indicates the brain areas most active while the subject performs particular tasks. *See* MEDICAL IMAGING.

Michael W. Levine

Bibliography. R. Axel, The molecular logic of smell, *Sci. Amer.*, pp. 154–159, October 1995; H. B. Barlow and J. D. Mollon, *The Senses*, Cambridge University Press, 1982; R. Blake and R. Sekuler, *Perception*, McGraw-Hill, in press, 2006; D. Dobbs, Fact or phrenology?, *Sci. Amer. Mind*, pp. 24–31, April 2005; M. W. Levine, *Sensation and Perception*, 3d ed., Oxford University Press, 2000; V. S. Ramachandran and E. M. Hubbard, Hearing colors, tasting shapes, *Sci. Amer.*, pp. 53–59, May 2003; R. Shapley and V. H. Perry, *Trends Neurosci.*, pp. 229–235, May 1986; D. V. Smith and R. F. Margolskee, Making sense of taste, *Sci. Amer.*, pp. 32–39, March 2001; H. Wang and C. J. Woolf, *Neuron*, 46:9–12, 2005.

Sense amplifier

An electronic amplifier circuit used to sense and refresh the value of a bit stored in a memory cell of a dynamic random access memory (DRAM) integrated circuit.

In DRAM, bits are represented in memory by the presence or absence of an electric charge stored on tiny capacitors. The charged capacitor has a voltage equal to the charge stored on the capacitor measured in coulombs divided by the cell capacitance measured in farads. A binary one is represented by the presence of charge on the capacitor, while zero charge represents a binary zero. A DRAM chip may store 4 megabits or more on a small piece of silicon. Cell capacitance is directly proportional to cell area, and since the chip area devoted to each bit-cell must decrease with an increase in the memory capacity of a given size of chip, the resulting cell capacitance is tiny, typically on the order of femtofarads (1 femtofarad $= 10^{-15}$ farad). Since the voltage used to charge these capacitors is usually limited to 5 V or less, the charge stored in each memory cell also is quite small, on the order of femtocoulombs. *See* CAPACITANCE; CAPACITOR.

In order to read out the value of a given bit of a word in this type of memory, the bit-cell voltage, or equivalently the magnitude of its charge, needs to be sensed, and the results of this sense operation must be delivered to the rest of the circuit. Both the sense amplifier used to read the bit-cell voltage and the interconnect wiring (the bit line) between the bit-cell and the amplifier have inherent parasitic capacitances which, in total, are much larger than the bit-cell capacitance. If the sense amplifier is simply switched across the bit-cell, the charge on the bit-cell capacitor will redistribute across the parasitic capacitor and the bit-cell capacitor in accordance with Kirchhoff's law, which requires that the two capacitors have the same voltage, where the total charge must come from the original charge stored on the memory-cell capacitor. Thus the voltage seen by the amplifier is equal to the original bit-cell voltage multiplied by the ratio of the bit-cell capacitance to the sum of the two capacitances, and the charge remaining on the memory cell capacitor after sensing would be reduced from the original charge in the same proportion. If the parasitic capacitance is much larger than the cell capacitance, it follows that the voltage available at the sense amplifier input will be small and the voltage and charge remaining on the memory cell will be considerably reduced. More sophisticated approaches need to be taken to sense this tiny voltage and to refresh the memory cell during sense, so that the memory cell capacitor is left fully charged (or fully discharged) when sensing is complete. *See* KIRCHHOFF'S LAWS OF ELECTRIC CIRCUITS.

In practice, all bit-cells in the memory need to be sensed and refreshed periodically, whether they are read or not, since charge leaks from the bit-cell capacitor, even when it is undisturbed by external events. Consequently, a sense amplifier must operate reliably with charge and voltage levels considerably less than the maximum values mentioned above. Since refreshing the contents of the bit-cells takes time from the normal operation of the computer using this memory, maximizing the time between required refresh operations is highly desirable. The amplifier must be designed to function reliably with the smallest possible detectable charge on the bit-cell. Much effort has been expended in developing designs which will meet these stringent criteria.

The sense amplifier action is based on metal-oxide-semiconductor field-effect transistors (MOSFETs). These can be thought of as simple switches that are opened and closed in a predetermined sequence to carry out the read/refresh memory cycle. To assist the sense, a memory reference cell, containing a capacitor of approximately one-half the capacitance of the bit-cell capacitor, is added to the circuit. *See* SEMICONDUCTOR MEMORIES; TRANSISTOR.

Philip V. Lopresti

Bibliography. B. Prince, *Semiconductor Memories: A Handbook of Design, Manufacture, and Application*, 2d ed., 1996.

Sense organ

A structure which is a receptor for external or internal stimulation. A sense organ is often referred to as a receptor organ. External stimuli affect the sensory structures which make up the general cutaneous surface of the body, the exteroceptive area, and the tissues of the body wall or the proprioceptive area. These somatic area receptors are known under the general term of exteroceptors. Internal stimuli which originate in various visceral organs such as the intestinal tract or heart affect the visceral sense organs or interoceptors. A receptor structure is not necessarily an organ; in many unicellular animals it is a specialized structure within the organism. Receptors are named on the basis of the stimulus which affects them, permitting the organism to be sensitive to changes in its environment.

Photoreceptors. Those structures which are sensitive to light and in some instances are also capable of perceiving form, that is, of forming images, are called photoreceptors. Light-sensitive structures include the stigma of phytomonads, photoreceptor cells of some annelids, pigment cup ocelli and retinal cells in certain asteroids, the eyespot in many turbellarians, and the ocelli of arthropods. The compound eye of arthropods, mollusks, and chordates is capable of image formation and is also photosensitive. Dermal light reactions which may be observed in aquatic forms appear to be related to unidentified photosensitive materials which occur in epithelial cells. *See* PHOTORECEPTION.

Phonoreceptors. Structures which are capable of detecting vibratory motion or sound waves in the environment are phonoreceptors. The most common phonoreceptor is the ear, which in the vertebrates has other functions in addition to sound perception. Among the fishes it has been demonstrated that the air bladder and lateral line organs, which in certain species have receptors (neuromasts) in their canals, serve as sound receptors. Sound perception in snakes, which lack a middle ear and are sensitive to airborne sounds, is by bone conduction

which makes them sensitive to ground vibrations. Sensory hairs or hair sensilla and tympanal organs occur in insects. Other organs which may function as sound receptors are the statocysts of certain crustaceans, although they are primarily statoreceptors. *See* PHONORECEPTION.

Statoreceptors. Structures concerned primarily with equilibration, such as the statocysts found throughout the various phyla of invertebrates and the inner ear or membranous labyrinth filled with fluid, are statoreceptors. Halteres are unique structures found in the Diptera which aid in the orientation of these insects. *See* POSTURAL EQUILIBRIUM.

Olfactoreceptors. The sense of smell is dependent upon the presence of olfactory neurons in the olfactory epithelium of the nasal passages among the vertebrates. Olfactory hairs extend from each olfactoreceptor cell, giving it a brushlike appearance. Jacobson's organ in amphibians may play a minor role in olfaction. Little is known about the sensation of smell among the invertebrates. The auricle of planarians has olfactory receptors, and the rhinophores of certain land mollusks are regarded as smell receptors. Odor sensitivity has been observed among many of the invertebrates, but no specific structure has been implicated. The most widely studied invertebrate group is the insects, in which three types of structures have been described: the sensilla placodea, sensilla basiconica, and the sensilla coeloconica. *See* OLFACTION.

Gustatoreceptors. The sense of taste is mediated by the taste buds. In most vertebrates these taste buds occur in the oral cavity, on the tongue, pharynx, and lining of the mouth; however, among certain species of fish, the body surface is supplied with taste buds as are the barbels of the catfish. Again, among the invertebrates, the most studied group has been the insects, and most of the experimental work has been based on behavioral studies. However, because sensilla are known to occur on the antennae, mouthparts, tarsus, and tibia of many species, and on the ovipositor of a few forms, it has been postulated that these structures are involved. Contact stimulation appears to activate the receptors. *See* TASTE.

Cutaneous receptors. The surface skin of vertebrates contains numerous varied receptors associated with sensations of touch, pain, heat, and cold. Tangoreceptors are associated with the phenomena of touch and pressure. Because sensory endings of both tangoreceptors and algesiroreceptors (pain receptors) terminate in the skin, it is often difficult to distinguish between the two. Among the vertebrates, cutaneous receptors are Grandry's corpuscle, Herbst corpuscle, the bulb of Krause, Merkel's corpuscle, Meissner's corpuscle, the Pacinian corpuscle, and the end organ of Ruffini. The end bulb of Krause is a thermoreceptor and is stimulated by cold (frigidoreceptor), whereas the end organs of Ruffini are believed to be influenced by heat (caloreceptor). Sensory and labial pits of certain reptiles are temperature-sensitive, especially during developmental life cycles. Thermal perception among insects has been attributed to the antennae and mouthparts as well as the cerci and tarsi. The ampullae of Lorenzini found in elasmobranchs may have a thermosensory function according to recent studies. *See* CHEMICAL SENSES; PAIN; SENSATION. Charles B. Curtin

Sensitivity (engineering)

A property of a system, or part of a system, that indicates how the system reacts to stimuli. The stimuli can be external (that is, an input signal) or a change in an element in the system. Thus, sensitivity can be interpreted as a measure of the variation in some behavior characteristic of the system that is caused by some change in the original value of one or more of the elements of the system.

There exist many situations where the sensitivity measure indicates the ability of a system to meet certain design specifications. For example, in an electronic system the sensitivity of the output current with respect to the variation of the power-supply voltage can be very critical. In that case, a system with a minimum sensitivity of the output current with respect to the power-supply voltage must be designed. Another example is a high-fidelity audio amplifier whose sensitivity can be interpreted as the capacity of the amplifier to detect the minimum amplifiable signal.

A limiting factor in using the sensitivity of a system to characterize performance at low signal levels is the noise. Noise is a statistical description of a random process inherent in all elements in a physical system. The noise is related to the minimum signal that can be processed in a system as a function of physical variables such as pressure, visual brightness, audible tones, and temperature. *See* ELECTRICAL NOISE.

Sensitivity is commonly used as a figure of merit for characterizing system performance. As a figure of merit, the sensitivity is a numerical indicator of system performance that is useful for predicting system performance in the presence of elemental variations or comparing the relative performance of two or more systems that ideally have the same performance. In the latter case, the performance of the systems relative to some parameter of interest is rank-ordered by the numerical value of the corresponding sensitivity functions. If T is the performance characteristic and X is the element or a specified input level, then mathematically sensitivity is expressed as a normalized derivative of T with respect to X.

A widely used definition of sensitivity in systems is H. W. Bode's logarithmic sensitivity function. It is defined by Eq. (1), where the symbols ∂ and ln stand

$$S_X^T = \frac{\partial T}{\partial X}\frac{X}{T} = \frac{\partial(\ln T)}{\partial(\ln X)} \tag{1}$$

for partial derivative and natural logarithm, respectively.

When sensitivity is used as an indicator of system performance, it follows from Eq. (1) that the percent change in the performance characteristic of interest, $\Delta T/T$, due to a given percent change in the parameter X, $\Delta X/X$, can be expressed in terms of the sensitivity

function by Eq. (2). As an example, if the resonant frequency of a band-pass filter is given by Eq. (3),

$$\frac{\Delta T}{T} \simeq S_X^T \cdot \left[\frac{\Delta X}{X}\right] \tag{2}$$

$$\omega_0 = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}} \tag{3}$$

then it follows from Eq. (1) that the sensitivity of ω_0 with respect to the resistor R_1 is $-1/2$ and from Eq. (2) that the percent change in ω_0 due to a 5% increase in R_1 is given by Eq. (4).

$$\frac{\Delta\omega_0}{\omega_0} \simeq S_{R1}^{\omega_0} \cdot \left[\frac{\Delta R_1}{R_1}\right] = (-1/2)(5\%) = -2.5\% \tag{4}$$

On the one hand, a relatively high sensitivity is desirable in a system whose performance characteristic must vary over a large range. An example could be a resonance filter with variable resonance frequency. More common, however, are situations where a minimum sensitivity is required; in that case, any variation in the component values should render small (or negligible) variation in the performance characteristics. An example could be a carburetor for a gasoline engine in which it is desired to maintain good performance (for example, fuel efficiency or horsepower) in spite of rather large manufacturing tolerances in the small mechanical components used in the construction of the carburetor and the wear of these components.

Edgar Sánchez-Sinencio

Bibliography. H. W. Bode, *Network Analysis and Feedback Amplifier Design*, 1945; J. L. Bordewijk, Inter-reciprocity applied to electrical networks, *Appl. Sci. Res.*, 6:1–74, 1956; M. Eslami and R. S. Marleau, Theory of sensitivity of network: A tutorial, *IEEE Trans. Educ.*, 32(3):319–324, 1989; L. P. Huelsman, *Active and Passive Analog Filter Design: An Introduction*, 1993; J. D. Schoeffler, The synthesis of minimum sensitivity networks, *IEEE Trans. Circuit Theory*, CT-11(2):271–276, 1964; G. C. Temes and S. K. Mitra, *Modern Filter Theory and Design*, 1973.

Sensory cell regeneration

The replacement of receptor cells with sensory end organs (retina, cochlea, taste buds, olfactory epithelium), most commonly via addition of newly differentiated cells to the systems.

Humans and other organisms use a variety of sensory receptors to detect and interpret information from the surrounding environment. These receptors may be damaged by environmental toxins, injury, or overstimulation, thereby reducing the sensory input received by the organism and the animal's ability to respond appropriately to stimuli. In some systems, such as taste, new sensory receptors are produced on a regular basis. In others, including hearing in humans, the loss of sensory cells is permanent. This article briefly describes the major classes of sensory receptors in vertebrates and some mechanisms of receptor cell regeneration.

Receptor cells. Most sensory stimuli are grouped into three main categories: chemical, visual, and mechanical. Taste receptor cells in the mouth and olfactory receptor neurons in the nose detect chemical stimuli and provoke the senses of taste and smell. Photoreceptor cells in the eye detect visual stimuli. Hair cells in the inner ear detect mechanical (vibrational) stimuli as either sound or balance depending on the frequency of the vibration.

Sensory information from each type of peripheral receptor is transmitted to the brain along pathways that are distinct from one another, except within regions of the brain where sensory information is compared or integrated. This separation of sensory systems is fortunate, since damage to one will not significantly impair the function of another. On the other hand, even minor damage to a given pathway can result in substantial sensory deficits. The most peripheral components, the sensory receptor cells, are often the most vulnerable. *See* BRAIN; NERVOUS SYSTEM (VERTEBRATE).

This vulnerability of sensory receptor cells stems from their position within the organism and from the extreme sensitivity of the receptors to their respective stimuli. Receptor cells such as olfactory receptor neurons and taste receptors must come into direct contact with a chemical stimulus in order to generate a response. Direct exposure makes these cells susceptible to damage from environmental toxins such as harsh chemicals. Photoreceptors in the eye and sensory hair cells in the ear do not make direct contact with the external environment but these cells are easily damaged by intense light or sound. Photoreceptors are exquisitely sensitive, capable of responding to as little as one photon of light. Hair cells are sensitive to mechanical stimuli as minute as 1 nm, making them extremely sensitive to low-intensity sound and balance cues. Since the loss of sensory receptors makes an organism vulnerable to dangers in its environment, numerous mechanisms have evolved to protect sensory systems from debilitating permanent damage. In many sensory systems, one such protective mechanism is the ability to regenerate receptor cells. Researchers struggle to understand these mechanisms with the hope of one day applying their findings to human sensory systems such as the eye and ear. *See* CHEMICAL SENSES; HEARING (HUMAN); OLFACTION; SENSATION; TASTE; VISION.

Innate regenerative capabilities. Regeneration of sensory receptor cells occurs in three manners: repair of damaged receptors, direct conversion (transdifferentiation) of a local nonreceptor cell into a receptor, or proliferation of new receptor cells from a pool of progenitor cells. While all three regenerative processes may occur in sensory systems, proliferation of new cells is arguably the most important and will be discussed briefly here.

In most adult vertebrates, chemical receptor cells possess the greatest regenerative potential. Aging taste receptor cells and olfactory receptor neurons undergo programmed cell death (apoptosis) and are replaced by newly proliferated receptor cells that arise within the sensory epithelia. In the mouse, the average taste receptor cell lives for only nine days, demonstrating that continuous receptor turnover is

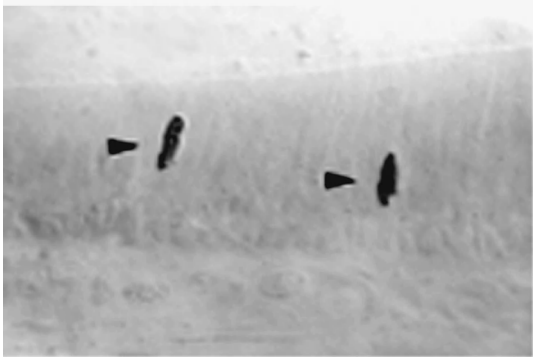


Fig. 1. Light micrograph of the inner ear of the goldfish, *Carassius auratus*. The inner ears of goldfish and other teleost fish continue to add sensory hair cells throughout the life of the animal. The nuclei of two progenitor cells are labeled with BrdU, a substance that marks dividing cells (arrowheads). These cells will go through mitosis and may eventually differentiate as hair cells in the sensory end organ.

normal in this system. There is a close link between apoptosis and proliferation in gustatory and olfactory systems such that a constant number of mature receptor cells exist in the system at any given time. Taste buds and olfactory epithelia can also increase production of receptor cells following end organ damage, leading to the restoration of epithelial structure and function. *See* APOPTOSIS; EPITHELIUM.

In contrast, regeneration of photoreceptors and sensory hair cells is dependent upon the species and age of the animal. In teleost fish and amphibians, the retina and inner ear epithelia grow throughout the life of the organism and new receptor cells are added to the growing sensory structure. Fishes and amphibians also possess increased regenerative capabilities in the eye and ear following sensory overstimulation or drug-induced damage to sensory receptors. This regeneration occurs mostly through an increase in proliferation of new receptor cells similar to that seen in olfactory and gustatory end organs (**Fig. 1** and **2**).

The situation in amniotic vertebrates (reptiles, birds, and mammals) is very different. It was originally thought that these vertebrates lost all regenerative capacity in the eye and ear. However, landmark studies in the past few decades show that birds and perhaps mammals do retain limited regenerative potential. In adult chickens, Müller glia cells in the retina respond to retinal damage by increasing proliferation and producing new retinal cells. Similarly, studies using a radioactive marker to identify proliferating cells in the ear demonstrate that sound or drug-induced damage to the auditory end organ in the bird ear causes dormant progenitor cells to proliferate, producing new sensory hair cells. The vestibular (balance) portions of the mammalian inner ear also possess the ability to produce small numbers of new hair cells. However, innate proliferative regeneration does not occur in the mammalian retina or cochlea. *See* CELL DIFFERENTIATION; EAR (VERTEBRATE); EYE (VERTEBRATE).

Therapeutic regenerative strategies. External intervention may be necessary to achieve regeneration in the mammalian retina and cochlea. Stem cell therapy provides one possible strategy. In vitro experiments with adult mammalian retinal cells show that proliferation is possible given the proper culture conditions. In some cases these newly proliferated cells do differentiate into specific retinal cell types. Transplantation of embryonic stem cells into damaged mammalian retina also shows promise, as some transplanted cells are incorporated into retinal cell layers. Introduction of neural stem cells into the mammalian cochlea has also met with some success, leading to differentiation of new cells that appear structurally similar to hair cells. However, functional recovery following stem cell transplantation has not yet been demonstrated in either sensory system. *See* STEM CELLS.

An alternative strategy is to harness the dormant proliferative potential of the mammalian eye and inner ear. Research on the molecular pathways involved in normal embryonic development of these sensory structures may reveal suites of molecular factors necessary for sensory receptor proliferation and differentiation. These factors could then be exogenously introduced in the damaged retina or cochlea to stimulate regeneration in the mature animal. The Notch signaling pathway, a genetic cellular patterning program, operates in normal retinal and cochlear development and may be involved in regeneration in nonmammalian vertebrates. *See* EMBRYONIC DIFFERENTIATION; EMBRYONIC INDUCTION; PATTERN FORMATION (BIOLOGY).

Notch signaling pathway. The Notch signaling pathway functions as a switching mechanism by

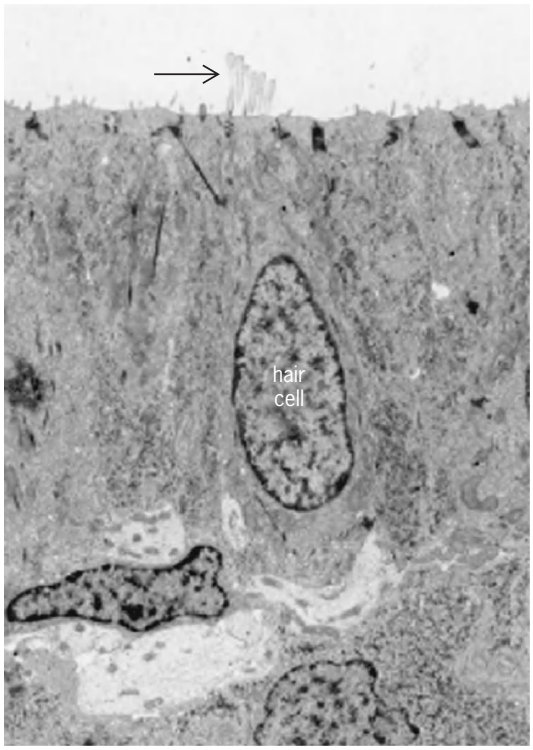


Fig. 2. Transmission electron micrograph of the goldfish inner ear. The inner ears of all vertebrates contain sensory hair cells, from which protrude specialized hairlike structures (arrow) to detect vibrations associated with the senses of hearing and balance.

which progenitor cells may be directed to exit the cell cycle and differentiate as one type of cell versus another, or to continue to divide. The pathway is activated when a progenitor cell bearing the receptor (Notch) comes into physical contact with a progenitor cell that bears one of the receptor ligands: Delta, Serrate, or Jagged1 (**Figs. 3** and **4**). Binding of Notch to the ligand results in an increase in the transcription of genes that encode inhibitory factors (*Enhancer of Split, HES*) [Fig. 4]. These inhibitory factors reduce the activity of cell-specific genes (for example, hair cell or retinal neuron-specific genes) that will enable the progenitor to differentiate as a particular cell type. Consequently, activation of the pathway in a given progenitor cell prevents that cell from differentiating and may result in further rounds of cell division. In this way, the Notch pathway acts as a developmental switching mechanism in which a single pool of progenitor cells may be divided into multiple mature cell types. The Notch signaling pathway plays an important role in retinal cell differentiation; in zebrafish, inhibition of this pathway has been shown to disrupt the spatial organization of retinal cell types during development. In the mammalian cochlea, activation of the Notch pathway apparently results in the presence of two basic cell types in the end organ: hair cells, which differentiate first, and nonsensory supporting cells, which differentiate after subsequent rounds of division. *See* CELL CYCLE.

Interestingly, the Notch pathway also appears to be involved in regeneration and postembryonic hair cell addition in nonmammalian vertebrates. For example, regions of the bird ear that produce large numbers of new hair cells as part of regeneration also express increased levels of Notch and other components of the Notch pathway. Notch expression continues in adult teleost fish, animals which generate postembryonic hair cells. Combined, these results indicate that Notch appears to play an important role in

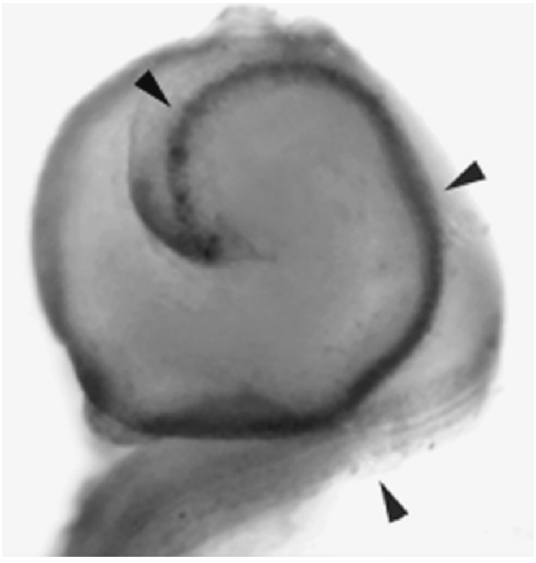
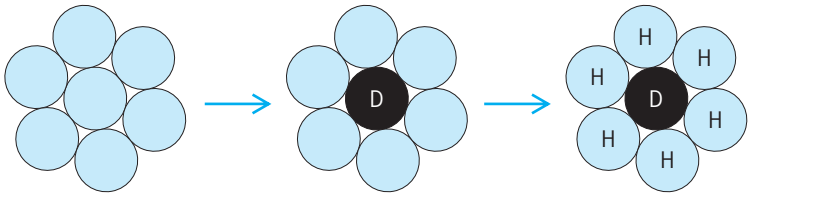


Fig. 3. Notch activity in the developing mouse cochlea. Gene expression in the cochlea (arrowheads) demonstrates the region of the cochlea in which one of the Notch ligands, Jagged1, is expressed.



A cluster of proneural cells before differentiation. All cells at this point are capable of differentiating as neural or nonneural cells

The black cell upregulates a Notch ligand (for example, Deltal "D"). This cell then assumes a neural fate (hair cell or retinal neuron).

Delta (or another Notch ligand) in the black cell binds Notch in the color cells. Binding activates downstream targets such as HES1 (H) and the cell differentiates as a supporting cell (ear) or glial cell (eye).

Fig. 4. Simplified model of Notch pathway activation during sensory system development. Differential activation of Notch and its ligands leads to patterning of neural and nonneural cells in the vertebrate inner ear and eye. Details are given under each panel. (*Adapted from J. E. Rooke and T. Xu, Positive and negative signals between interacting cells for astablishing neuronal fate, Bioessays, 20:209–214, 1998, and N. E. Baker, Notch signalling in the nervous system. Pieces still missing from the puzzle, BioEssays, 22:264–273, 2000*)

the postembryonic growth and regeneration of the inner ear in nonmammalian vertebrates. However, since Notch acts primarily as a molecular switch, the expression of Notch alone is not likely to produce new retinal neurons or hair cells. A second set of molecules, encoded by proneural genes, is required as well.

Proneural genes. Proneural genes produce a set of molecular factors that direct developing progenitor cells toward a neural fate. These genes are present in both vertebrates and invertebrates and promote differentiation of a variety of neural cell types depending upon the region of the nervous system in which they are expressed. Recent studies in zebrafish and mice have demonstrated that photoreceptor and hair cell progenitors express specific proneural genes, some of which are related to *atonal*, a gene first identified in the fly.

The mammalian retina expresses multiple proneural genes, including the atonal homologs *Math3* and *Math5*. Mutant mice lacking *Math3* and two other proneural genes, *Mash1* and *NeuroD*, show a severe decrease in the number of photoreceptors but an increase in ganglion cells (another retinal cell type), suggesting that multiple proneural genes act in a complementary fashion in the mouse retina to produce photoreceptor cells. Disruption of *Math5* expression, however, decreases retinal ganglion cells numbers and increases expression of other retinal proneural genes. Taken together, these results indicate that different proneural genes specify different subsets of retinal cell types and that loss of one or more cell types leads to overproduction of others.

In the mammalian inner ear, *Math1*, an atonal homolog, is both necessary and sufficient for hair cell development. In mutant mice that lack functional *Math1*, the inner ear develops normally but completely lacks hair cells. Conversely, introduction of *Math1* into the ears of postnatal guinea pigs leads to production of additional hair cells in regions of the ear that do not normally produce them. Finally, it has been shown that *Math1* is one of the cell-specific genes that is targeted by inhibitory factors of the Notch pathway and that decreased levels of Notch activity result in increased numbers of hair cells in

both mice and zebrafish. Collectively, these results indicate that progenitor cells in the inner ear are directed toward specific cell fates through a combination of proneural gene expression and Notch pathway activation.

In summary, many vertebrate sensory systems regenerate throughout the life of the animal, allowing continued sensitivity to sensory information in the environment. Regenerative capability is lost, however, in the postembryonic mammalian retina and cochlea. Research in nonmammalian model organisms such as birds and fishes that retain regenerative abilities as adults may lead to novel clinical therapies for human retinal degeneration and hearing impairment. *See* DEVELOPMENTAL GENETICS.

Pamela Lanford; Allison Coffin

Bibliography. T. Akagi et al., Requirement of multiple basic helix-loop-helix genes for retinal neuronal subtype specification, *J. Biol. Chem.*, 279:28592–28498, 2004; N. A. Bermingham et al., *Math1*: An essential gene for the generation of inner ear hair cells, *Science*, 284:1837–1842, 1999; R. L. Bernardos et al., Notch-Delta signaling is required for spatial patterning and Müller glia differentiation in the zebrafish retina, *Develop. Biol.*, 278:381–395, 2005; J. T. Corwin and D. A. Cotanche, Regeneration of sensory hair cells after acoustic trauma, *Science*, 240:1772–1774, 1988; C. Haddon et al., Delta-Notch signaling and the patterning of sensory cell differentiation in the zebrafish ear: Evidence from the *mind bomb* mutant, *Development*, 125:4637–4644, 1998; P. Hitchcock et al., Persistent and injury-induced neurogenesis in the vertebrate retina, *Prog. Ret. Eye Res.*, 23:183–194, 2004; R. C. Murray and A. L. Calof, Neuronal regeneration: Lessons from the olfactory system, *Cell Develop. Biol.*, 10:421–431, 1999; M. A. Parker and D. A. Cotanche, The potential use of stem cells for cochlear repair, *Audiol. Neurotol.*, 9:72–80, 2004; P. A. Tsonis and K. Del Rio-Tsonis, Lens and retina regeneration: Transdifferentiation, stem cells and clinical applications, *Exp. Eye Res.*, 78:161–172, 2004.

Sepioidea

An order of the class Cephalopoda (subclass Coleoidea) including the cuttlefishes (*Sepia*), the bobtail squids (*Sepiola*), and the ram's-horn squid (*Spirula*). The group is characterized by an internal shell that is calcareous and broad with closely packed laminate chambers (the cuttlebones of cuttlefishes); calcareous, cylindrical, and coiled with about 40 subspherical chambers (the ram's-horn shell of *Spirula*); chitinous and rod- or bladelike (the gladius of bobtail squids); or absent. The mouth is surrounded by ten appendages (eight arms and two longer tentacles) that bear suckers with chitinous rings. The tentacles are contractile and retractile into pockets at their bases (see **illus.**).

Cuttlefishes are common benthic or epibenthic (living on or just above the bottom, respectively) animals that occur in the warm and temperate waters of the nearshore and continental shelf zones of the

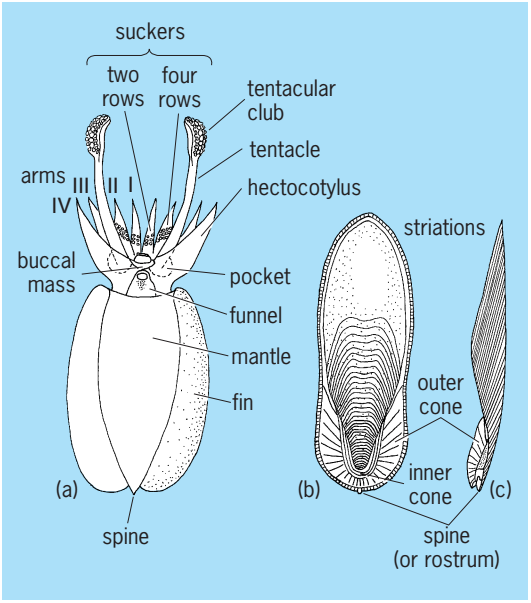


Diagram of some features of (*a*) a cuttlefish and of (*b*) a cuttlebone shown in ventral view and (*c*) in cross section.

Old World, but they are excluded from the Western Hemisphere (North and South America). They prey on shrimps, crabs, small fishes, and other cuttlefishes. The sexes are separate, and during mating, which follows a colorful ritualistic courting behavior, sperm is transferred to the female in cylindrical packets (spermatophores) by a modified arm (the hectocotylus) of the male. The eggs are relatively large, are often stained black with the female's ink, and are laid in grapelike clusters attached to algae, shells, or rocks. Young hatch in a few weeks to 2 months, depending on temperature, and animals live 1–2 years. The cuttlebones are used to control the buoyancy of the cuttlefishes as fluid is pumped into and out of the laminar chambers. Dried cuttlebones are a source of calcium for cage birds and are used for fine jewelry molds, dentifrices, and cosmetics. The cuttlefishes eject an attention-getting blob of brownish-black ink when threatened by predators, then change color, become transparent, and jet-swim away, leaving the predator to attack the false body (pseudomorph) of ink. Artists have used the ink, called sepia, for centuries. Cuttlefishes are important in world fisheries; about 200,000 metric tons are caught each year for human food. Sizes of sepioids range from the smallest known cephalopods of about 0.2–0.7 in. (6–17 mm) body length (*Idiosepius pygmaea*) to cuttlefishes of about 1.6 ft (50 cm) body length and 25 lb (12 kg). *See* CEPHALOPODA; COLEOIDEA.

Clyde F. E. Roper

Bibliography. P. R. Boyle (ed.), *Cephalopod Life Cycles*, vol. 1, 1983, vol. 2, 1987; F. W. Lane, *Kingdom of the Octopus*, 1960; K. N. Nesis, *Cephalopods of the World*, transl. from Russian, 1987; C. F. E. Roper, M. J. Sweeney, and C. E. Nauen, *Cephalopods of the World*, FAO Fisheries Synopsis 125, vol. 3, 1984; K. M. Wilbur et al. (eds.), *The Mollusca*, vol. 11: *Form and Function*, 1988; K. M. Wilbur and M. R. Clarke (eds.), *The Mollusca*, vol. 12: *Paleontology and Neontology of Cephalopods*, 1988.

Sepiolite

A complex hydrated magnesium silicate mineral named for its resemblance to cuttlefish bone, alternately named meerschaum (sea foam). The ideal composition, $Mg_8(H_2O)_4(OH)_4Si_{12}O_{30}$, is modified by some additional water of hydration, but is otherwise quite representative. The ideal density, 2.26 g/cm^3 (1.31 oz/in.^{3}), is never realized in aggregates. Interlaced disoriented fibers aggregate into a massive stone so porous that it floats on water. These stones are easily carved, take a high polish with wax, and harden when warmed. They have been valued from antiquity in the eastern Mediterranean region for ornaments and make attractive smoking pipes.

Two related types of fibers exist; either of these may occur as the massive interlacings, as paperlike mattings, as bedded clayey deposits, or as more pronounced asbestoslike fibers. They make up a group related in crystal structure to amphibole. Palygorskite and attapulgite are used as names when the amphibole chain relationship is in double units, and sepiolite when it is in triple units. *See* AMPHIBOLE.

Extensive bedded fibrous deposits of either type have been found useful as dispersoids in oil-well drilling fluids for operations that penetrate high-salt formation waters. They are also active as absorptive agents for decolorizing oils, as carriers for insecticides, and as fillers. The bedded palygorskite (attapulgite) variety is produced in quantity in Florida and Georgia in the United States, and in the Ukraine and the Caucasus in Russia; much of the sepiolite variety is produced in Spain, North Africa, and Arabia.

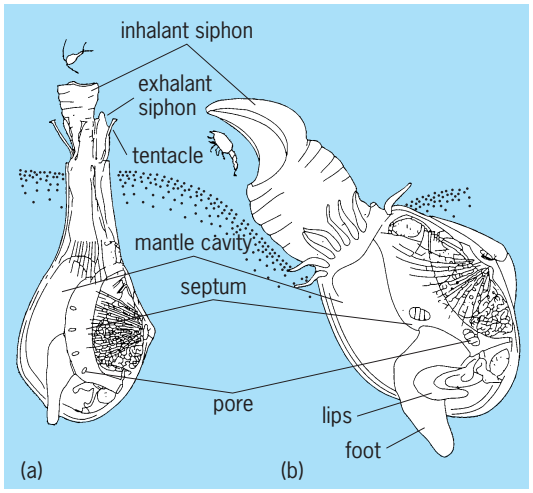
A third fiber, loughlinite, contains up to four sodium ions in place of two magnesium of the sepiolite formula but is changed to asbestoslike fibers of sepiolite by soaking in magnesium chloride. *See* CLAY MINERALS; SILICATE MINERALS.

William F. Bradley

Septibranchia

A subclass of bivalve mollusks (class Bivalvia) that are unique in their possession of a muscular septum instead of a filamentous gill. The Septibranchia equate in great part to the superfamily Poromyacea, which includes the septibranch families Cuspidariidae and Poromyidae, and the Verticordiidae. The Verticordiidae have gills that are greatly reduced in size. Although there are a few cuspidariid species in shallow seas, the great majority of septibranchs are found at lower slope and abyssal depths which are deficient in food for filter-feeding bivalves. They live close to the surface in soft sediments. Most species are less than 20 mm maximum length. *See* BIVALVIA; MOLLUSCA.

The term septibranch remains extremely useful because it describes mollusks having a septum and other morphological specializations that relate to the septibranch's unique carnivorous habits. The septum is a muscular, horizontal partition dividing the mantle cavity (see **illus.**). It is derived from the enormous gill found in the mollusks of the subclass Lamellibranchia: The filaments are reduced in size and



Morphology of two genera of septibranchs. (*a*) *Cuspidaria* (*after J. A. Allen and R. E. Morgan, The functional morphology of Atlantic deep water species of the families Cuspidariidae and Poromyidae (Bivalvia): An analysis of the evolution of the septibranch condition, Trans. Roy. Soc. Lond., 294:413–546, 1981*). (*b*) *Poromya* (*after B. Morton, Prey capture in the carnivorous septibranch Poromya granulata (Bivalvia: Anomalodesmata: Poromacea), Sarsia, 66:241–256, 1981*).

modified. The septum is perforated with cilia-lined pores on either side of the midline. Water circulates into the lower mantle cavity, and then through pores in the septum into the supramantle cavity and exits. This action is brought about by the raising and lowering of the septum and a related cycle of opening and closing the pores, in contrast to other bivalves in which this type of circulation is driven by the cilia of the gills.

The septum, as in all bivalve gills, is used in respiratory and feeding functions. In the cuspidariids the action of rapid raising of the septum draws water and epibenthic and planktonic animals into the mantle cavity. The capture of organisms is precise, the presence of prey being detected first by the siphonal tentacles, and then, via nerve pathways, which trigger a sharp inhalant response from the septum. In contrast, the carnivorous poromyids which have a less muscular septum capture their prey by means of a hood which is an extension of the inhalant siphon. The prey, in this case epibenthic organisms, is also located by sensory tentacles around the inhalant siphon. The closely related verticordiids have adhesive-coated, muscular, branched tentacles surrounding the inhalant aperture, to which planktonic organisms probably adhere. It is thought that the prey is passed into the mantle cavity by the tentacles' bending into the inhalant aperture, the prey being transferred to the mantle cavity possibly with the help of a valve to the inside of the aperture. Food then passes to enlarged manipulative lips of the mouth.

T 'ese prey-capturing movements require rapid redistribution of the blood, and there are various septal and mantle sinuses into which large quantities of blood are shunted in sequence. Other modifications include a muscular stomach for crushing the prey and high proteolytic activity of the gastric juice. The need for rapid muscular and hydraulic movements

can also be related to the nature of the hinge of the shell, the relatively small central ligament, and the development of a ventral part of the ligament (the lithodesma). The lithodesma provides a flexible hinge mechanism that is found only within the anomalodesmatan bivalves. Similarly, the ventral margins of the mantle are fused, assisting in preventing the escape of material from the mantle cavity. *See* FEEDING MECHANISMS (INVERTEBRATE). J. A. Allen

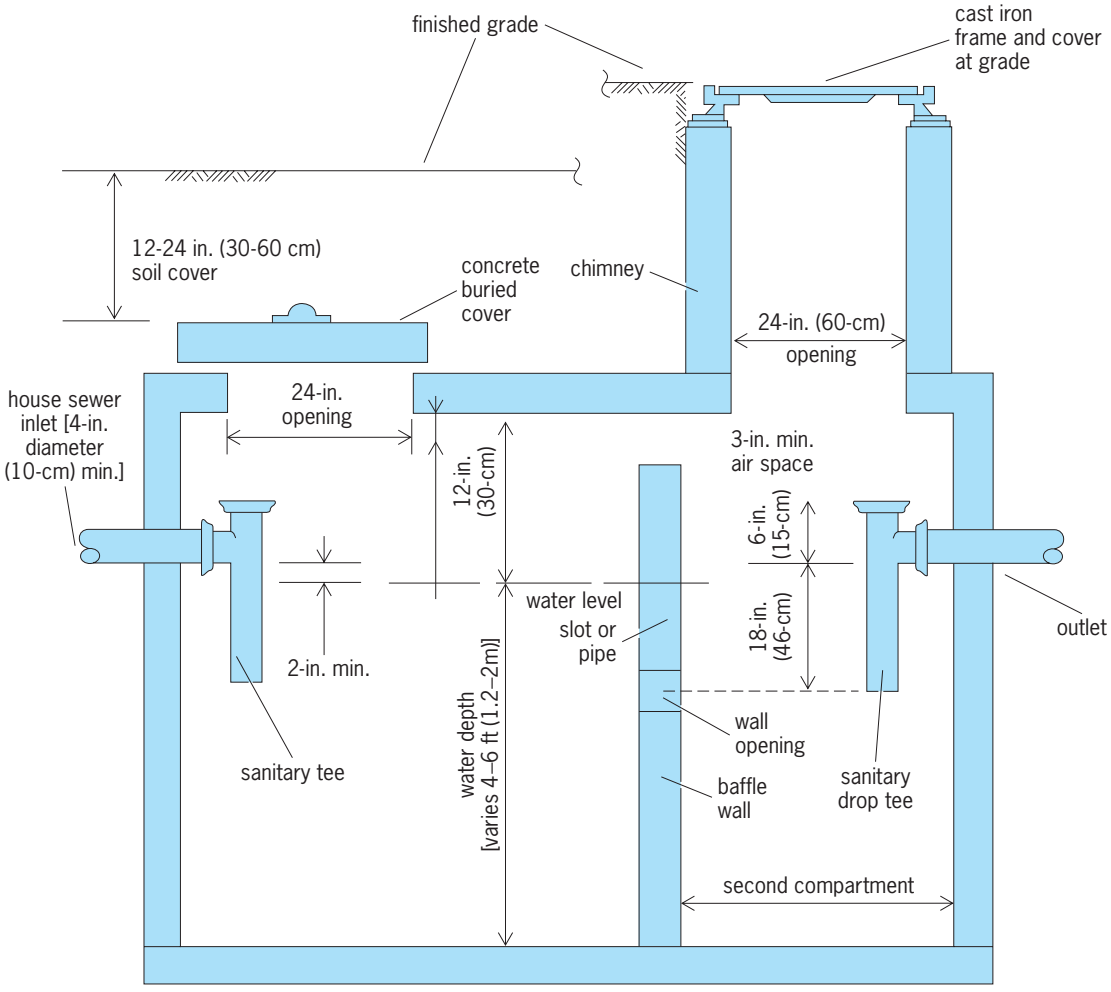
Bibliography. J. A. Allen, On the functional morphology of the family Verticordiidae (Bivalvia) with descriptions of new species from the abyssal Atlantic, *Trans. Roy. Soc. Lond.*, 268:401–536, 1974; J. A. Allen and R. E. Morgan, The functional morphology of Atlantic deep water species of the families Cuspidariidae and Poromyidae (Bivalvia): An analysis of the evolution of the septibranch condition, *Trans. Roy. Soc. Lond.*, 294:413–546, 1981; S. P. Parker (ed.), *Synopsis and Classification of Living Organisms*, 2 vols., 1982.

Septic tank

A single-story, watertight, on-site treatment system for domestic sewage, consisting of one or more compartments, in which the sanitary flow is detained to permit concurrent sedimentation and sludge digestion. The septic tank is constructed of materials not subject to decay, corrosion, or decomposition, such as precast concrete, reinforced concrete, concrete block, or reinforced resin and fiberglass. The tank must be structurally capable of supporting imposed soil and liquid loads. Septic tanks are used primarily for individual residences, isolated institutions, and commercial complexes such as schools, prisons, malls, fairgrounds, summer theaters, parks, or recreational facilities. Septic tanks have limited use in urban areas where sewers and municipal treatment plants exist. *See* CONCRETE; REINFORCED CONCRETE; STRUCTURAL MATERIALS.

Design. Daily flow is based either on a per-capita flow or on gallons per bedroom. Minimum size of tank (rectangular or circular) for a private residence with one or two bedrooms should be 750 gal (3 m^3), while for a seven-bedroom house it should be 2000 gal (7.5 m^3). The tank should provide a minimum of 24-h detention, and up to 72 h is recommended.

The liquid depth, from the invert of the outlet line to the tank floor, should be at least 4 ft (1.2 m) but not greater than 6 ft (2 m). An air space of at least 12 in. (30 cm) should be provided above the stored



Section of a typical septic tank for on-site residential wastewater treatment.

liquid. The tank length, or diameter, should be not less than 5 ft (1.5 m), and for rectangular tanks should be at least three times the tank width. The inlet and outlet pipes should be at least 4 in. (10 cm) in diameter, with the inlet at least 2 in. (5 cm) above the outlet; and each pipe should be provided with vented tees or baffles to maintain effective retention of scum and sludge and to provide adequate venting. For better operation, the tank should be divided into compartments, with the first compartment containing one-half to two-thirds of the total volume. No compartment should be less than 24 in. (60 cm) in length. The compartment wall must have adequate liquid connections and not block the air-vent space. There should be adequate access to each compartment of the tank for inspection and sludge removal (see **illus.**).

Commercial and institutional septic tanks, with flows in excess of 1500 gal (6 m^3) per day, are designed to store the peak flow for a minimum of 12 h, in addition to sludge storage, although 24-h detention is recommended. If the septic tank is to receive ground garbage, the storage capacity should be increased by at least 30%, although 50% is recommended.

Operation. Septic tanks do not treat sewage; they merely remove some solids and condition the sanitary flow so that it can be safely disposed of to a subsurface facility such as a tile field, leaching pools, or buried sand filter. The septic tank, because of its surface area and detention time, permits heavier solids to settle and lighter particles to rise, thus clarifying the liquid to some extent. The organic solids retained in the tank undergo a process of liquefaction and anaerobic decomposition by bacterial organisms. The clarified septic tank effluent is highly odorous, contains finely divided solids, and may contain enteric pathogenic organisms. The detention and conditioning of the sanitary flow alters the raw sewage so that it may be applied to a properly sized, permeable subsurface soil stratum that will receive the flow with minimal plugging or clogging. The small amounts of gases produced by the anaerobic bacterial action are usually vented and dispersed to the atmosphere without noticeable odor or ill effects.

Maintenance. Septic tanks for private dwellings generally require yearly inspection, and removal of accumulated scum and sludge approximately every 3–5 years. If the inspection indicates that the sludge layer is within 12 in. (30 cm) of the bottom of the outlet tee, or the scum layer is either within 3 in. (7.5 cm) of the top of the outlet tee or more than 10 in. (25 cm) thick, the tank should be pumped out by an experienced scavenger and the contents disposed of in a safe and sanitary manner. *See* RURAL SANITATION; SEWAGE; SEWAGE TREATMENT.

Gerald Palevsky

Bibliography. G. Tchobanoglous, F. L. Burton, and H. D. Stensel (eds.), *Wastewater Engineering: Treatment and Reuse*, 4th ed., 2002; J. A. Salvato, *Environmental Engineering and Sanitation*, 4th ed., 1994.

Sequence stratigraphy

The study of stratigraphic sequences, defined as stratigraphic units bounded by unconformities. L. L. Sloss was the first to formally establish, in 1963, that Phanerozoic cratonic successions in North America could be subdivided into six broad sequences that spanned the entire interior of the continent. However, his work was based on widespread knowledge among North American petroleum geologists, going back to E. Blackwelder in 1909, that the stratigraphy of the continent could be subdivided into a limited number of unconformity-bounded sequences. In later work, Sloss and others demonstrated that these sequences could be correlated partially or entirely with similar successions within other continental interiors, in Russia and South America. With improvements in the acquisition and processing of reflection-seismic data by petroleum exploration companies, in the 1970s came the recognition that unconformity-bounded sequences could be recognized in most sedimentary basins. This was the beginning of an important development, seismic stratigraphy, which also included the use of seismic reflection character to make interpretations about large-scale depositional facies and architecture (**Fig. 1**). These breakthroughs were accompanied by the proposal by Peter Vail that the bounding unconformities are generated by erosion resulting from eustatic (worldwide) changes in sea level and they are therefore correlatable worldwide. These developments were rapidly accepted within the petroleum industry and brought about a revolution in stratigraphy that is still under way. *See* SEISMIC STRATIGRAPHY; UNCONFORMITY.

Correlation problem. If Vail's proposal is correct, it provides the basis for the development of a stratigraphic template for worldwide correlation. However, research during the 1980s and 1990s suggested that the global eustasy model is probably simplistic. There are several other important mechanisms that generate widespread unconformities and sequence boundaries, foremost among them tectonism—the uplift and subsidence of basins resulting from plate-tectonic movements and the effects of thermal convection in the mantle (dynamic topography). Such mechanisms operate regionally, from the scale of individual basins up to that of major continental areas; but tectonic mechanisms do not generate vertical motions of basins that are globally correlatable both in timing and in direction. This being the case, the validity of the global chart of stratigraphic cycles first published by Vail and his coworkers in 1977 (with a revised version appearing in 1987) has been questioned. Current methods for the dating of sequences (primarily biostratigraphy) are insufficiently accurate to provide independent confirmation of global correlations. *See* PLATE TECTONICS.

The controversy regarding global sequence correlation should not detract from the fact that techniques for the documentation, correlation, and analysis of stratigraphic successions have been revolutionized by the methods of sequence stratigraphy. This is because sequence concepts have led

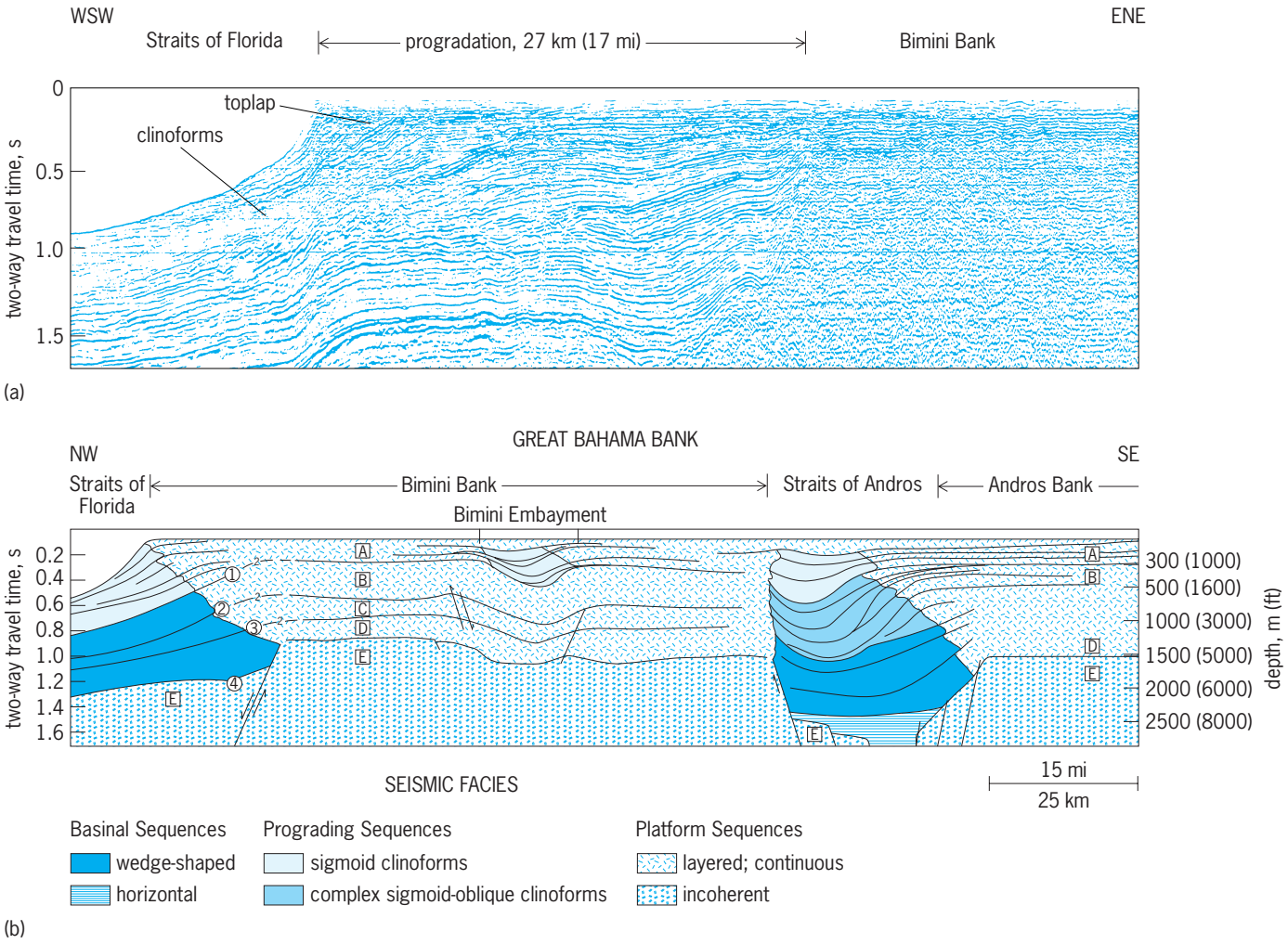


Fig. 1. Seismic reflection characterization. (*a*) Seismic transect across the western margin of Bimini Bank, Bahamas Platform, showing variations in seismic facies. Note the poorly reflective rocks at lower right, representing a sedimentary basement bounded by a normal fault, and well-defined layering in the carbonate-dominated succession above. (*b*) Interpretation of a seismic transect across the Bahamas Bank, showing subdivision of the succession into sequences (indicated by capital letters) bounded by regional unconformities (indicated by numbers). The oldest sequence is mid-Cretaceous. The youngest is still forming. (*After G. P. Eberli and R. N. Ginsburg, Soc. Sedimentary Geol. Spec. Publ., no. 44, 1989*)

to the development of predictive models that have dramatically improved methods for mapping and interpreting stratigraphic units. They are a vast improvement on the older methods of stratigraphic documentation and correlation, which were based on lithostratigraphy—the definition of units (formations, members, groups) based on similarities in facies. Lithostratigraphy embodies no predictive capability.

Sequence model. Underpinning sequence-stratigraphic methods are the following interrelated principles: (1) The volume of sediment accumulating in any part of a sedimentary basin is dependent on the space made available for sediment by changes in sea level or basin-floor elevation. This space is referred to as accommodation. (2) Changes in accommodation tend to be cyclic, and they are accompanied by corresponding changes in sedimentary environment and depositional facies. Thus, a rise in base level typically leads to an increase in accommodation, deepening of the water in the basin, with corresponding changes in facies, and a transgression, with a consequence landward shift in depositional environments and in depositional facies. A fall in base level may lead to exposure and erosion (negative accommodation), with the development of a widespread unconformity. (3) These predictable changes provide the basis for a model of the shape and internal arrangement or architecture of a sequence, including the organization and distribution of sedimentary facies and the internal bedding surfaces that link these facies together. *See* BASIN; DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; FACIES (GEOLOGY).

There are two major kinds of components of sequence models: systems tracts and bounding surfaces (**Fig. 2**). The details of these components vary with depositional and tectonic setting. Thus, the subsidence patterns and regional variations in generation of accommodation are very different in the different tectonic classes of sedimentary basin (rift basins, extensional-margin prisms, foreland basins, and so forth). Sequence architectures also vary

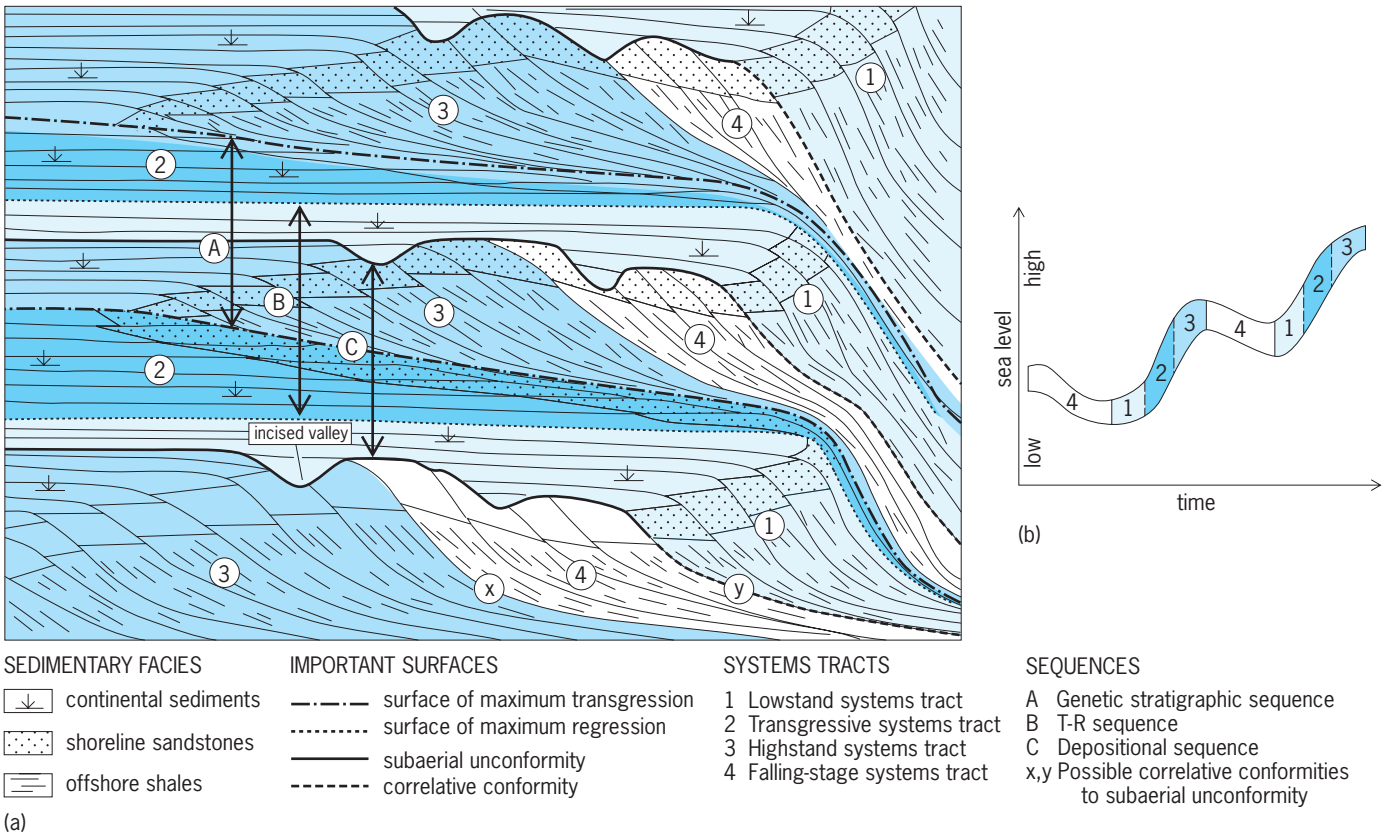


Fig. 2. Clastic-dominated sequence. (*a*) Model for clastic-dominated sequences on an extensional continental margin, showing the key stratigraphic surfaces and the four major systems tracts that develop within each sequence. Double-headed arrows show three ways in which sequences have been defined (see text). (*b*) Diagram showing how the systems tracts relate to changes in relative sea level with time. (*After W. Helland-Hansen and J. G. Gjelberg, Sedimentary Geol., 92:31–52, 1994*)

markedly with differences in the overall sedimentary style. For example, carbonate and clastic depositional systems respond very differently to changes in accommodation. The first sequence model was for clastic sequences on extensional continental margins. *See* CONTINENTAL MARGIN.

Clastic-dominated sequences. As originally defined, sequences are bounded by unconformities (the depositional sequences of Fig. 2). These surfaces (sequence boundaries) are typically well developed within coastal and shelf sediments, where they form as a result of subaerial exposure and erosion during falling sea level. In deeper-water settings, including the continental slope and base of slope, there may be no corresponding sedimentary break; and sequences may be mapped into such settings only if the unconformity can be correlated to the equivalent conformable surface (the correlative conformity). In some instances, the surface of marine transgression, which develops during the initial rise in sea level from a lowstand, forms a distinctive surface that is close in age to the subaerial unconformity and may be used as the sequence boundary. *See* MARINE GEOLOGY; MARINE SEDIMENTS.

The subaerial unconformity is commonly cut by valleys formed by lowstand incision, and may be characterized by erosional relief, making it a difficult surface to correlate and map in the subsurface. For this reason, some workers prefer to map sequences by correlating the deeper-water sediments, typically mudstones, formed at the time of maximum transgression (the top of which is commonly termed the maximum flooding surface). These tend to be widespread. Some workers map sequences using these mudstones as the sequence boundaries, and these sequences have been termed genetic stratigraphic sequences (Fig. 2).

The cycle of rise and fall of sea level may be divided into four segments: lowstand, transgressive, highstand, and falling stage. The deposits that form at each stage are distinctive, and are assigned to systems tracts named for each of these stages.

Falling stage. The sequence-bounding unconformity is formed by subaerial erosion during the falling stage. Incised river valleys form on the continental shelf, and they may preserve remnants of river terraces on the flanks of the valleys. The seaward ends of the valleys may form the heads of submarine canyons, down which sediment is funneled by sediment gravity flows to the base of the continental slope. The submarine-fan and other deep-marine clastic deposits in the deep ocean are generated primarily at this stage of the sea-level cycle because of the abundant sediment supply and connected transport system available to deliver the detritus. *See* SUBMARINE CANYON.

Sea-level lowstand. During the sea-level lowstand, incised valley systems may deepen and widen and

begin to fill with fluvial deposits. Deltas and other coastal depositional systems may build seaward at the margins of the continental shelf. On the coastal plain, mature soils develop on the upland areas between river systems.

Transgression. The transgressive systems tract forms as sea level rises, and accommodation generation is rapid. Typically, coastal regions undergo transgression, and coastal deposits backstep landward (retrogradational stacking of deposits such as shoreline and barrier facies). Incised valleys flood and become estuaries. Sedimentary facies in these valleys change from nonmarine to marine, commonly exhibiting the influence of waves or tides. Later, the transgression may drown the entire coastal plain, blanketing it with mud. Wave erosion may plane off coastal sediments as the shoreline retreats, forming a prominent time-transgressive erosion surface called a ravinement surface. Little sediment may reach deeper-water areas, which may accumulate organic and other detritus in a condensed section. Submarine fans are starved of sediment (except where they are fed by deep canyons at the mouths of some of the largest rivers, such as the Amazon) and are blanketed with a thin mud layer.

Highstand phase. The rate of sea-level rise slows as the cycle nears its end, and enters the highstand phase. The rate of sediment supply typically begins to outpace the rate at which new accommodation is generated, and coastal depositional systems change from retrogradational to aggradational (vertical stacking) and then typically to progradational (seaward-stepping), forming seaward-dipping deltaic and continental-slope depositional surfaces, known as clinoforms. These prograding deposits build back out over the mudstones and condensed sediments of the transgressive systems tract, generating a downlap architecture. During the main highstand phase, sediment supply outpaces the rate of generation of new accommodation, and much sediment bypasses the coastal plain and continental shelf. Slope clinoforms thin upward into these zones of bypass and nondeposition, a condition known as toplap (Fig. 1).

Seaward regression. In the early phases of the next fall in sea level, coastal environments regress seaward, and the shelf may be eroded by wave action, generating a regressive surface of marine erosion. Shoreline deposits may form thin sheets or isolated pods resting with sharp bases on this surface, but all these features may be removed by subaerial erosion as sea level continues to fall.

Carbonate-dominated sequences. Carbonate sedimentation is most active in warm, clear, shallow, shelf seas. During the sea-level cycle, these conditions tend to be met during the highstand phase. Sediment production may be so active, including that of reef development at the platform margin, that it outpaces accommodation generation, leading to deposition on the continental slope. Oversteepened sediment slopes there may be remobilized, triggering sediment gravity flows and transportation into the deep ocean. This process is called highstand shedding. During lowstand, continental platforms may be exposed, leading to erosion and karst development.

Sequences containing evaporites. Evaporites, such as gypsum, anhydrite, or halite, typically form during one of two phases of the sea-level cycle. During lowstand, some marine basins become partially or entirely cut off from marine circulation. Given a hot, arid climate, evaporation may outpace fluvial influx, and evaporite layers may form in environments ranging from shallow to deep. Evaporites may also form in shallow basins on the continental platform during transgression.

Scales and mechanisms of sequence development. The sequence cycle presented here has been described without reference to a time scale. This is because there are several processes of sequence generation that range from a few tens of thousands of years to hundreds of millions of years for the completion of a cycle of rise and fall of sea level. More than one such process may be in progress at any one time within a basin, with the production of a range of sequence styles nested within or overlayering each other.

High-frequency sequence generation is driven by orbital forcing of climate (the so-called Milankovitch effects), of which glacial eustasy is the best-known outcome (although subtle effects on oceanic temperatures and oxygenation levels can also generate significant sequence-stratigraphic effects). The effects of glacioeustasy have dominated continental-margin sedimentation since the freeze-up of Antarctica in the Oligocene. Major glacial episodes also occurred during the late Paleozoic, at the end of the Ordovician, and at several times during the Precambrian. Sequences with durations between about 20,000 and 100,000 years characterize all these periods of Earth history. *See* CLIMATE HISTORY; GEOLOGIC TIME SCALE; GLACIOLOGY; PALEOCLIMATOLOGY.

Regional tectonism—such as the process of thermal subsidence following rifting, and flexural loading in convergent plate settings—develops changes in basement elevation that drive changes in relative sea level. These cycles have durations of a few millions to a few tens of millions of years, and they are confined to individual basins or the flanks of major orogens or plate boundaries. Movement on individual thrust sheets or nappes may generate local cycles of loading and erosional unloading with durations of thousands to tens of thousands of years within limited areas of flanking foreland basins. These tectonically driven sequences commonly dominate the regional stratigraphy; but they are strictly local to regional in extent, and cannot be used in the construction of any global cycle chart.

Low-frequency cycles of eustatic sea-level change are driven by changes in the average rate of sea-floor spreading. Fast rates generate thermally elevated seafloor spreading centers that displace the ocean waters up onto the land (highstand phase); whereas reorganization of spreading centers tends to lead to thermal subsidence, deepening of the ocean basins, and eustatic lowstands. Regional changes in plate

motions occur every few millions to tens of millions of years, with corresponding consequences for global sea level. These changes tend to be additive during the gradual formation and breakup of supercontinents over periods of a few hundreds of millions of years. The highest global sea levels occur during supercontinental break-up, as during the fragmentation of Pangea and the formation of the Atlantic, Indian, and Southern oceans. During the Cretaceous, when this process was at its peak, global sea levels were 100–300 m (300–1000 ft) higher than at the present. Exceptionally low sea levels follow the assembly of a supercontinent, when spreading centers undergo reorganization. *See* PALEOGEOGRAPHY.

Applications. Sequence concepts enable petroleum exploration and development geologists to construct predictive sequence models for stratigraphic units of interest from the limited information typically available from basins undergoing petroleum exploration. These models can guide regional exploration, and can also assist in the construction of production models that reflect the expected partitioning of reservoir-quality facies within individual stratigraphic units.

Documentation of the sequence stratigraphy of sedimentary basins is an essential component of regional geological analysis, and may yield invaluable insights into the tectonic and climatic history of the area. *See* GEOPHYSICAL EXPLORATION; STRATIGRAPHY; WELL LOGGING. Andrew D. Miall

Bibliography. D. Emery and K. Myers (eds.), *Sequence Stratigraphy*, Blackwell Science, Oxford, 1996; A. D. Miall, *The Geology of Stratigraphic Sequences*, Springer-Verlag, Berlin, 1997; P. R. Vail et al., Seismic stratigraphy and global changes of sea-level, in C. E. Payton (ed.), Seismic Stratigraphy: Applications to Hydrocarbon Exploration, *Amer. Ass. Petrol. Geol. Mem.*, 26:49–212, 1977; J. C. Van Wagoner et al., *Siliciclastic Sequence Stratigraphy in Well Logs, Cores, and Outcrops*, American Association of Petroleum Geologists Methods in Exploration Series, no. 7, 1990.

Sequoia

The giant sequoia or big tree (*Sequoia gigantea*) occupies a limited area in California and is said to be the most massive of all living things (**illus.** *a*). The leaves are evergreen, scalelike, and overlapping on the branches (illus. *b*). In height sequoia is a close second to the redwood (300–330 ft or 90–100 m) but the trunk is more massive. Sequoia trees may be 27–30 ft (8–9 m) in diameter 10 ft (3 m) from the ground. The stump of one tree showed 3400 annual rings. The red-brown bark is 1–2 ft (0.3–0.6 m) thick and spongy. Vertical grooves in the trunk give it a fluted appearance. The heartwood is dull purplish-brown and lighter and more brittle than that of the redwood. Because the loss in felling the trees is so great, the logs so difficult to handle, and the wood so brittle, sequoia is little marketed. The wood and bark contain much tannin, which is probably the cause of the great resistance to insect and fungus attack. The most magnificent trees are within the Kings Canyon and Sequoia National Parks, where there are many individual trees 25 ft (7.5 m) in diameter, containing 500,000 board ft (1180 m^3) and up, and probably 3000 years old. The largest specimen weighs about 6000 tons (5400 metric tons). Giant sequoia is sometimes grown in the eastern United States and responds well to cultivation in the British Isles and in central Europe. *See* FOREST AND FORESTRY; PINALES; REDWOOD; TREE. Arthur H. Graves; Kenneth P. Davis

Giant sequoia (*Sequoia gigantea*). (*a*) Tree, showing relative size of man at base. (*b*) Branch and cone.

Series

The indicated sum of a succession of numbers or terms. Series are used to obtain approximate values of infinite repeating decimals, to solve transcendental equations, to obtain values of logarithms or trigonometric functions, to evaluate integrals, and to solve boundary value problems.

For a finite series, with only a limited number of terms, the sum is found by addition. For an infinite series, with an unlimited number of terms, a sum or value can be assigned only by some limiting process. When the simplest such process yields a value, the infinite series is convergent. Many tests for convergence enable one to learn whether a sum can be found without actually finding it.

If each term of an infinite series involves a variable x and the series converges for each value of x in a

certain range, the sum will be a function of x. Often the sum is a given function of x, $f(x)$, for which a series having terms of some given form is desired. Thus the Taylor's series expansion

$$f(x) = \sum_{n=0}^{i} f^{(n)}(a)\frac{(x-a)^n}{n!}$$

can be found for a large class of functions, the analytic functions, and represents such functions for sufficiently small values of $|x - a|$. For a much less restricted type of function on the interval $-\pi < x < \pi$, a Fourier series expansion of the form

$$^1/_2A_0 + \sum_{n=1}^{i}(A_n \cos nx + B_n \sin nx)$$

can be found.

Finite series. Here the problem of interest is to determine the sum of the first n terms,

$$S_n = u_0 + u_1 + u_2 + \cdots + u_{n-1}$$

when u_n is a given function of n. Examples are the arithmetic series, with $u_n = a + nd$ and $S_n = (n/2)[2a + (n-1)d]$, and the geometric series, with $u_n = ar^n$ and $S_n = a(1 - r^n)/(1 - r)$. *See* PROGRESSION (MATHEMATICS).

If v_n is any function such that $v_{n+1} - v_n = u_n$, then $S_n = v_n - v_0$. For methods of solving this difference equation $\Delta v_n = u_n$ *see* INTERPOLATION.

For example, if u_n is any polynomial of the nth degree, v_n will be a polynomial of the $(n + 1)$st degree. In particular,

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$$

As an example with u_n a rational function of n, from

$$\frac{-1}{n+2} - \frac{-1}{n+1} = \frac{1}{(n+2)(n+1)}$$

it may be concluded that if

$$u_n = \frac{1}{(n+1)(n+2)}$$

then

$$S_n = 1 - \frac{1}{n+1} = \frac{n}{n+1}$$

Convergence and divergence. An infinite series is the indicated sum of an unlimited number of terms

$$u_0 + u_1 + u_2 + \cdots + u_n + \cdots$$

or more briefly

$$\sum_{n=0}^{\infty} u_n$$

or simply Σu_n, read "sigma of u_n." The sum S_n of the first n terms is known as the nth partial sum. Thus S_n is the finite sum

$$\sum_{k=0}^{n-1} u_k$$

If, as n increases indefinitely or becomes infinite, the partial sum S_n approaches a limit S, then the infinite series Σu_n is convergent. S denotes the sum or value of the series. For example, if $|r| < 1$,

$$S = \Sigma ar^n = \frac{a}{1-r} \quad \text{since} \quad S_n = a\frac{1-r^n}{1-r}$$

and

$$S = \Sigma \frac{1}{(n+1)(n+2)} = 1$$

since

$$S_n = 1 - \frac{1}{n+1}$$

If, as n becomes infinite, the partial sum S_n does not approach a finite limit, then the infinite series Σu_n is divergent. For example, $\Sigma 1$ diverges, since here $S_n = n$ becomes infinite with n. Also $\Sigma(-1)^n$ diverges, since here $S_n = {}^1/_2[1 - (-1)^{n+1}]$ which is alternately one and zero.

It follows from the definition of S_n that $S_{n+1} - S_n = u_n$. For a convergent series, one may take limits in this equality and so deduce that

$$\lim_{n\to\infty} u_n = S - S = 0$$

Thus, if as n becomes infinite, u_n either approaches a nonzero limit or fails to approach a limit, the series must diverge. This checks the earlier conclusion about $\Sigma 1$, with $\lim u_n = 1$, and also that about $\Sigma(-1)^n$, where $u_n = (-)^n$ does not approach a limit. A convergent series remains convergent if a finite number of terms is added, removed, or changed either at the beginning or distributed throughout the series. This changes all S_n after a certain point by the same finite constant. Thus the limit S is changed by this same constant, but there is no change in the fact of approach to a limit.

Positive series. These are series each of whose terms is a positive number or zero. For such series, the partial sum S_n increases as n increases. If for some fixed number A, no sum S_n ever exceeds A, the sums are bounded and admit A as an upper bound. In this case, S_n must approach a limit, and the series is convergent. If every fixed number is exceeded by some S_n, the sums are unbounded. In this case, S_n must become positively infinite and the series is divergent. The tests for convergence of positive series are tests for boundedness, and this is shown by a comparison of S_n with the partial sums of another series or with an integral.

Integral test. Let the function $f(x)$ be positive and always decrease as x increases for x greater than m, some fixed positive integer. Then the series $-f(n)$, with $u_n = f(n)$, converges if the integral

$$\int_m^{\infty} f(x)\,dx$$

converges. The series diverges if the integral diverges. The principal application of this test is that with $f(x) = 1/x^p$, where p is any positive constant. The result is that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges if p is greater than 1 and diverges if p is less than or equal to 1. Such divergent series as

$$\sum_{n=1}^{\infty} \frac{1}{n} \quad \text{or} \quad \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

illustrate that series may diverge and still have $\lim u_n = 0$. The slow divergence of

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

may be seen from the fact that over 10^{433} terms must be taken to make the partial sums exceed 1000.

Comparison tests. Let k be any positive constant, Σu_n a positive series to be tested, and Σc_n a positive series known to be convergent. Then if $u_n \leqq kc_n$ for all n greater than some fixed integer m, the series Σu_n converges.

If Σd_n is a positive series known to be divergent, and $u_n \geqq kd_n$ for all n greater than some fixed integer m, then the series Σu_n is divergent.

As corresponding tests involving limits, if

$$\lim_{n\to\infty} \frac{u_n}{c_n} = L$$

where L is a finite limit, then Σu_n converges. But if

$$\lim_{n\to\infty} \frac{u_n}{d_n} = L \quad (L > 0,\ L = +\infty)$$

then the series Σu_n is divergent.

Ratio test. For positive series, the simple ratio test is based on a consideration of

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = t$$

If t is less than unity, the series converges. If t is greater than unity, or if the ratio becomes positively infinite, the series diverges.

If $t = 1$, no conclusions can be drawn directly. But in many such cases

$$\frac{u_{n+1}}{u_n} = 1 - \frac{b}{n} + \frac{c}{n^2} + \cdots$$

may be written. In this case the series converges if $b > 1$, and diverges if $b \leqq 1$.

Cauchy's test, which is related to the ratio test but depends on a single term, is as follows. If for a positive series

$$\sqrt[n]{u_n} \leq r < 1 \quad \text{for all } n \text{ greater than } m$$

the series converges. If $\sqrt{u_n} \geqq 1$ for an infinite number of values of n, the series diverges.

Alternating series. These are series whose terms are alternately positive and negative. For such a series, if each term is numerically less than the preceding term, and

$$\lim_{n\to\infty} u_n = 0$$

the series converges. An example is

$$\Sigma(-1)^n \frac{1}{n+1}$$

For such a series, the difference between S_n, the nth partial sum, and S, the sum of the series, is numerically less than the first unused term, u_n.

Absolute convergence. For any series Σu_n which may have both positive and negative terms, the series of absolute values, $\Sigma|u_n|$, is a positive series whose convergence may be proved by one of the tests for positive series. If $\Sigma|u_n|$ converges, then Σu_n necessarily converges and is said to converge absolutely. The sum of an absolutely convergent series is independent of the order of the terms.

Conditional convergence. A series which converges but which does not converge absolutely is said to be conditionally convergent. For such a series, a change in the order of the terms may change the sum or cause divergence. In fact, by a suitable rearrangement, any sum may be obtained. The series $1 - {}^1/_2 + {}^1/_3 - \cdots$ is conditionally convergent with sum $\log_e 2$. The rearrangement $1 + {}^1/_3 - {}^1/_2 + {}^1/_5 + {}^1/_7 - {}^1/_4 + \cdots$ obtained by taking blocks of two positive terms and then one negative term is conditionally convergent with sum ${}^3/_2 \log_e 2$.

Operations on series. Two convergent series, if added termwise, give a convergent series. If $\Sigma u_n = S$ and $\Sigma v_n = T$, then $\Sigma(u_n + v_n) = S + T$. If both series are absolutely convergent, then the double series $\Sigma\Sigma u_m v_n$, where m and n each run from 1 to infinity, converges to the product ST absolutely, and so does any rearrangement. In particular, this is true for the Cauchy product, with $u_0v_n + u_1v_{n-1} + \cdots + u_{u-1}v_1 + u_nv_0$ as its $(n+1)$st term.

If Σu_n and Σv_n each converge, and if the Cauchy product series converges, its sum is ST. This will necessarily be the case if at least one of the series converges absolutely.

In any convergent series, parentheses may be inserted to form a new convergent series, with the same sum. But the removal of parentheses may convert a convergent series to a divergent one; for example, $\Sigma(1 - 1) = 0$ becomes $\Sigma(-1)^n$, which diverges.

Power series. These are series with $u_n = a_n x^n$. For such a series, it may happen that

$$\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = A$$

If $A = 0$, the series converges for all values of x. If $A \neq 0$, the series converges for all x of the interval $-1/A < x < 1/A$. It will diverge for all x with $|x| > 1/A$. For any power series, the interval of convergence is related in this way to a number A, which, however, in the general case has to be given by the superior limit of

$$\sqrt[n]{|a_n|}$$

Similar remarks apply to the series with $u_n = a_n(x - c)^n$. Here the interval of convergence is $|x - c| < 1/A$, where A may be

$$\lim \left| \frac{a_{n+1}}{a_n} \right| \quad \text{or} \quad \lim \sqrt[n]{|a_n|}$$

if these limits happen to exist. In any case, A is equal to the superior limit

$$\overline{\lim} \sqrt[n]{|a_n|}$$

One of the most important power series is the binominal series:

$$1 + mx + \frac{m(m-1)}{1 \cdot 2} x^2 + \cdots + \frac{m(m-1)(m-2)\cdots(m-n+1)}{n!} x^n + \cdots$$

When m is a positive integer, this is a finite sum of $m + 1$ terms which equals $(1 + x)^m$ by the binomial theorem. When m is not a positive integer, the interval of convergence is $-1 < x < 1$, and for x in this interval, the sum of the series is $(1 + x)^m$. *See* BINOMIAL THEOREM.

The sum function of a power series is continuous inside the interval of convergence. If the series converges at either end of the interval, the function is continuous at this end. Inside the interval of convergence, a power series may be integrated termwise. At any point inside the interval, it may be differentiated termwise. For example, from the series

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots \qquad \text{for } -1 < x < 1$$

differentiation gives

$$\frac{1}{(1-x)^2} = 1 + 2x + \cdots + nx^{n-1} + \cdots$$

Integration gives

$$\log_e(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^{n+1}}{n+1} - \cdots$$

Since this converges when $x = 1$, $\log_e 2 = 1 - {}^1/_2 + {}^1/_3 - {}^1/_4 + \cdots$.

Taylor series. Let the power series $\Sigma a_n(x - c)^n$ have the sum function $f(x)$. Then $a_0 = f(c)$, $a_n = f^{(n)}(c)/n!$, and the series is the Taylor series of $f(x)$ at $x = c$. Thus, every power series whose interval of convergence has positive length can be put in the form

$$f(x) = f(c) + f'(c)\frac{x-c}{1!} + \cdots + f^{(n)}(c)\frac{(x-c)^n}{n!} + \cdots$$

where $f(x)$ is the sum function.

The Maclaurin series is the special case of Taylor series with $c = 0$

$$f(x) = f(0) + f'(0)\frac{x}{1!} + \cdots + f^{(n)}(0)\frac{x^n}{n!} + \cdots$$

Remainder term. For any function which is finite together with all of its derivatives for $x = c$, the difference between the function and the first $(n + 1)$ terms of its Taylor series is the remainder R_n. For a suitable value x_1 in the interval $c < x_1 < x$, which is

$$R_n = f^{(n+1)}(x_1)\frac{(x-c)^{n+1}}{(n+1)!}$$

Lagrange's form for R_n. This gives $x_1 = c + \theta h$, where $h = x - c$, and θ is a suitable value between 0 and 1. It is true that frequently, if $h \to 0$, $\theta \to 1/(n+2)$. Cauchy's form for R_n is

$$R_n = f^{(n+1)}(c + \theta h)(1 - \theta)^n \frac{h^{n+1}}{n!}$$

This may be used to prove that the binomial series converges to $(1 + x)^m$, by showing that $R_n \to 0$ as $n \to \infty$. Lagrange's form for R_n may be used to show that

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

are each Maclaurin series valid for all values of x.

Operations on power series. Two power series $\Sigma a_n(x - c)^n$ and $\Sigma A_n(x - c)^n$ may be added or multiplied by the Cauchy product rule. The resultant series will converge in an interval at least as large as the smaller of the intervals of convergence of the two given series. The first series may be divided by the second, provided that the divisor series is not zero at $x = c$, to give a series for the quotient with some nonzero interval of convergence. A Taylor series with constant term c may be substituted in another series about $x = c$. A power series may be inverted; that is, if $y = \Sigma a_n(x - c)^n$, with $a_1 \neq 0$, there is an expansion $x = \Sigma b_n(y - a_0)^n$ where $b_0 = c$, and the other b_n may be found by substitution.

Complex series. The series Σz_n, in which the general term is the complex number $x_n + iy_n$, with $i^2 = -1$, is said to converge to $A + iB$ if $\Sigma x_n = A$ and $\Sigma y_n = B$. Thus the tests for real series may be made on Σx_n and Σy_n, or the positive series $\Sigma |x_n|$ may be tested. If this converges, the given complex series converges absolutely, and necessarily converges.

Complex power series. The series with $u_n = a_n z^n$, where the $a_n = p_n + iq_n$ are now complex numbers, and the complex variable $z = x + iy$, are of great importance in the theory of analytic function of a complex variable. More generally, power series occur in $(z - c)$, with $u_n = a_n(z - c)^n$, where c is any complex number. There is always a radius of convergence $R = 1/A$, where A may be

$$\lim \left| \frac{a_{n+1}}{a_n} \right| \quad \text{or} \quad \lim \sqrt[n]{|a_n|}$$

if these limits exist. In any case, A is equal to the superior limit

$$\overline{\lim} \sqrt[n]{|a_n|}$$

If A is finite and not zero, so is R, and there is a circle of convergence $|z-c| < R$ within which the series converges. If $A = 0$, the series converges for all values of z. If $A = \infty$, $R = 0$, the series converges for no value except $z = c$. For example, $\Sigma n!z^n$ converges for $z = 0$, but for no other values of z; $\Sigma(z-3)^n/2^n$, with $A = 1/2$, $R = 2$, converges [to $2/(5-z)$] for $|z-3| < 2$; $\Sigma\, z^n/n!$, with $A = 0$, $R = \infty$, converges (to e^z) for all values of z.

Uniform convergence. Let each term of a series be a function of z, $u_n = g_n(z)$. Let S_n be the sum of the first n terms, and S the sum to which the series converges for a particular value of z. Then $R_n = S - S_n$ is the remainder after n terms, and for the particular value of z,

$$\lim_{n\to\infty} R_n$$

must equal zero. If, for a given range of z, it is possible to make $R_n(z)$ arbitrarily small for sufficiently large n without specifying which z in the range is under consideration, the series converges uniformly. The Weierstrass comparison test may be used to test for uniformity. If ΣU_n is a convergent series of positive constants, or a uniformly convergent series of positive terms, and $|u_n| \leqq U_n$ in the range considered, then Σu_n converges uniformly in the range. For a uniformly convergent series, if each $u_n(z)$ is continuous, the sum function $S(z)$ is continuous. A uniformly convergent series may be integrated termwise. If the differentiated series converges uniformly, the series may be differentiated termwise.

Analytic functions. A function $f(z)$ is analytic in a two-dimensional region if, at each point of the region, the derivative $f'(z)$ exists. A uniformly convergent series of analytic functions has a sum function which is analytic and, in particular, may be integrated or differentiated termwise. These results all apply to any power series whose radius of convergence R is not zero, since every power series $\Sigma a_n(z-c)^n$ converges uniformly in any circle $|z-c| < R_1$, where $R_1 < R$. Thus the sum function $f(z) = \Sigma a_n(z-c)^n$ is an analytic function of z for any z such that $|z-c| < R$. For such z, the function $f(z)$ possesses derivatives of every order, each expandable in a series obtained by termwise differentiation. Each such series has the same radius of convergence R. The power series is necessarily the Taylor series of the function $f(z)$, so that $a_n = f^{(n)}(c)(z-c)^n/n!$. If a function is single-valued and analytic in a two-dimensional region of the complex plane, and c is any point inside this region, then there is a Taylor expansion $f(z) = \Sigma f^{(n)}(c)(z-c)^n/n!$. Its radius of convergence R is at least as great as the largest circle with center at c, all of whose interior points are interior points of the given region of analyticity.

Fourier series. Let $f(x)$ be a periodic function of period T, so that $f(x+T) = f(x)$. Then the Fourier series for $f(x)$ is

$$A + \sum_{n=1}^{\infty}(A_n \cos n\omega x + B_n \sin n\omega x)$$

where $\omega = 2\pi/T$ and

$$A = \frac{1}{T}\int_a^{a+T} f(x)\,dx \qquad A_n = \frac{2}{T}\int_a^{a+T} f(x)\cos n\omega x\,dx$$

$$B_n = \frac{2}{T}\int_a^{a+T} f(x)\sin n\omega x\,dx$$

That is, A is the average of $f(x)$, A_n is twice the average of $f(x)\cos n\omega x$, and B_n is twice the average of $f(x)\sin n\omega x$ over any interval of length T. Because of the periodicity, it is immaterial which interval, or value of a, is used. If, in each interval of period T, the graph of $f(x)$, which need not be continuous, is made up of arcs which collectively have finite length, then the Fourier series necessarily converges to $f(x)$ at each point where $f(x)$ is continuous. At each point of discontinuity the series converges to $\frac{1}{2}[f(x-) + f(x+)]$, the average of the right- and left-hand limits. If $f(x)$ is continuous at all points, and has a uniformly bounded derivative, the series will converge uniformly for all values of x.

A Fourier series may always be integrated termwise, but it may not be permissible to differentiate such a series termwise unless $f'(x)$ satisfies some condition for development in a Fourier series.

Fourier sine series. On the interval $0 < x < L$, the Fourier sine series

$$f(x) = \sum_{n=1}^{\infty} B_n \sin n\omega x$$

where

$$\omega = \frac{\pi}{L} \qquad B_n = \frac{2}{L}\int_0^L f(x)\sin n\omega x\,dx$$

may be considered as the Fourier series for the function $F(x)$ which is odd, $F(x) = -F(-x)$, of period $2L$, $F(x+2L) = F(x)$, and which equals $f(x)$ on the interval $0 < x < L$. Thus if $f(x)$ was an odd function for $-L < x < L$, the sine series is valid inside this interval.

Fourier cosine series. On the interval $0 \leqq x \leqq L$, the Fourier cosine series

$$f(x) = A + \sum_{n=1}^{\infty} A_n \cos n\omega x$$

where

$$\omega = \frac{\pi}{L} \qquad A = \frac{1}{L}\int_0^L f(x)\,dx$$

$$A_n = \frac{2}{L}\int_0^L f(x)\cos n\omega x\,dx$$

may be considered as the Fourier series for the function $F(x)$ which is even, $F(x) = F(-x)$, of period $2L$, $F(x+2L) = F(x)$, and which equals $f(x)$ on the interval $0 < x < L$. Thus if $f(x)$ was an even function for $-L \leqq x \leqq L$, the cosine series is valid inside this interval. The sum of the sine or cosine series is always $\frac{1}{2}[F(x-) + F(x+)]$ for each value of x, so that the sum of $F(x)$ at each point of continuity.

Cesaro summability. Let S_n be the sum of the first n terms of a series Σu_n, and form the sequence

$$C_1 = \frac{S_1}{1},\ C_2 = {}^1\!/_2(S_1 + S_2), \ldots, C_n = \frac{1}{n}\sum_{k=1}^{n} S_k$$

If

$$\lim_{n\to\infty} C_n = L$$

exists, the series is said to be summable in the sense of Cesaro, or $C(1)$ to L. A convergent series with sum S is necessarily summable $C(1)$ to S, but a divergent oscillating series may be summable $C(1)$. For example, if $u_n = (-1)^n$, the series is $1 - 1 + 1 - \cdots$ which diverges. But $S_k = 1$ and 0 alternately, and $L = {}^1\!/_2$. Thus the series is summable $C(1)$ to ${}^1\!/_2$, a value also made plausible from consideration of the limit as $x \to 1$ of the identity

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots$$

There are other more elaborate methods of summability by which sums can be assigned to certain divergent series.

Fejér theorem. There are continuous functions whose Fourier series do not converge. But, for any continuous periodic function, the theorem of Fejér asserts that the Fourier series is always summable $C(1)$ to the function for every value of x.

Convergence in the mean. Let $f(x)$ be periodic with T, and

$$T_n = a + \sum_{k=1}^{n=1}(a_k \cos k\omega x + b_k \sin k\omega x)$$

be any trigonometric sum of order $n - 1$. Then T_n may be regarded as an approximation to $f(x)$, and the degree of the approximation may be measured by the average of the square of the error,

$$E_n = \frac{1}{T}\int_a^{a+T} [f(x) - T_n(x)]^2\, dx$$

Then

$$E_n = \frac{1}{T}\int_a^{a+T} [f(x)]^2\, dx - A^2 - {}^1\!/_2\sum_{k=1}^{n-1}({A_k}^2 + {B_k}^2) + (A - a)^2 + {}^1\!/_2\sum_{k=1}^{n-1}(A_k - a_k)^2 + (B_k - b_k)^2$$

This shows that, for all trigonometric sums of given order, S_n, the one formed with Fourier coefficients, $a = A$, $a_k = A_k$, $b_k = B_k$ makes the average error least. Moreover, for any function such that

$$\int_a^{a+T} [f(x)]^2\, dx$$

is finite, as n becomes infinite the limit of E_n, the average squared error for S_n, is zero. Thus

$$\lim_{n\to\infty}\int_a^{a+T} [f(x) - S_n(x)]^2\, dx = 0$$

This is the condition for the trigonometric sums S_n to converge in the mean to $f(x)$.

Integration. If the sequence $S_n(x)$ converges in the mean to $f(x)$, and $g(x)$ is any square-integrable fixed function, it is true that

$$\lim \int S_n(x)g(x)\, dx = \int f(x)g(x)\, dx$$

for the same limits on the integrals in both members. That is, the series with partial sums $S_n(x)$ may be integrated termwise, and the same is true of the series with partial sums $g(x)S_n(x)$. This fact may be used to derive the formulas for the Fourier coefficients in terms of integrals which were given above.

Again, if $T_n(x)$ converges in the mean to $g(x)$, then

$$\lim \int S_n(x)T_n(x)\, dx = \int f(x)g(x)\, dx$$

with the same limits on the integrals in both members. From this Parseval's identity may be deduced,

$$\int_a^{a+T} f(x)g(x)\, dx = AA' + {}^1\!/_2\sum_{k=1}^{\infty}(A_kA_k' + B_kB_k')$$

where A', A_k', B_k' are Fourier coefficients for $g(x)$.

Examples of Fourier series. For $-\pi < x < \pi$, there is a series

$$\sin ax = \frac{2\sin a\pi}{\pi}\cdot\left(\frac{\sin x}{1^2 - a^2} - \frac{2\sin 2x}{2^2 - a^2} + \frac{3\sin 3x}{3^2 - a^2} - \cdots\right)$$

The analogous expansion for $\cos ax$, with $x = 0$ and $a = z/\pi$ leads to

$$\csc z = \frac{1}{z} - \frac{2z}{z^2 - \pi^2} + \frac{2z}{z^2 - 2^2\pi^2} - \frac{2z}{z^2 - 3^2\pi^2} + \cdots$$

The expansion for $\cos ax$ is valid for $x = \pi$. With this and $z = z/\pi$, it gives

$$\cot z = \frac{1}{z} + \frac{2z}{z^2 - \pi^2} + \frac{2z}{z^2 - 2^2\pi^2} + \frac{2z}{z^2 - 3^2\pi^2} + \cdots$$

The last two expansions hold for any complex z for which no denominator is zero.

Infinite product for the sine. A series for $\log_e[(\sin z)/z]$ may be found from the expansion of $\cot z - 1/z$ by integration. This leads to

$$\sin z = z\left(1 - \frac{z^2}{\pi^2}\right)\left(1 - \frac{z^2}{2^2\pi^2}\right)\left(1 - \frac{z^2}{3^2\pi^2}\right)\cdots$$

the infinite product for the sine. Putting $z = \pi/2$ leads to

$$\frac{\pi}{2} = \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\cdots}$$

which is Wallis's product. Equating the z^3 term on the right with the term $-z^3/3!$ in $\sin z$ gives

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}$$

This is a special case of

$$\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = \frac{2^{2k-1}\pi^{2k}}{(-1)^{k-1}(2k)!} B_{2k}$$

where the B_{2k} are rational fractions, the Bernouilli numbers. $B_2 = {}^1\!/_6$, $B_4 = -{}^1\!/_{30}$, $B_6 = {}^1\!/_{42}$. For n odd, B_n is zero unless $n = 1$, and $B_1 = -{}^1\!/_2$. These B_n are the coefficients in

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}$$

They occur in such expansions as

$$\tan x = \sum_{k=1}^{\infty} \frac{(2^{2k} - 1)2^{2k}(-1)^{k-1}}{(2k)!} B_{2k} x^{2k-1}$$

$$\cot x = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{(-1)^k 2^{2k}}{(2k)!} B_{2k} x^{2k-1}$$

$$\csc x = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{(2 - 2^{2k}(-1)^k}{(2k)!} B_{2k} x^{2k-1}$$

Stirling's formula for m!. The expansion

$$\log_e(m!) = \log_e \sqrt{2\pi} + (m + {}^1\!/_2)(\log_e m) - m + \sum_{r=1}^{n} \frac{B_{2r} m^{-2r+1}}{2r(2r-1)} + R_n$$

leads to a divergent series if R_n is omitted and $n \to \infty$. But the series is asymptotic in the sense that R_n is always numerically less than the last term in the sum which involves B_{2n}. Thus for large m, a few terms of the sum give a good approximation. This leads to Stirling's approximation to $m!$, $m! \sim \sqrt{2\pi m}\, m^m e^{-m}$. For $m \to \infty$, the absolute error becomes infinite, but the percentage error is of the order of $e^{1/12m} \sim {}^1\!/_{12}\, m$ which is small even for moderate m.

Applications of series. Series sometimes appear in disguised form in arithmetic. Thus the approximation of a rational number by an infinite repeating decimal is really a geometric series, ${}^1\!/_3 = 0.3$ being a series with $u_n = {}^1\!/_3 > (10^{n+1})$ with sum ${}^1\!/_3$.

It is possible to use $1 = 0.9$ to find the fraction represented. For example, $0.285714 = 285714/999999 = {}^2\!/_7$.

Roots are often conveniently found by the binomial series. For instance, for small x, $(1 + x)^{1/n} \sim 1 + (x/n)$, $\log_e(1 + x) \sim x$.

Solution of equations. Sometimes algebraic or transcendental equations are best solved by reverting series. By translations of x and y, $y = y_0 + Y$, $x = x_0 + X$, it is possible to make $Y = 0$ correspond to $X = 0$ near the points of interest. Also, a change of scale, $X = kX'$, makes the first coefficient one. Then with new notation, $y = x + bx^2 + cx^3 + dx^4 + ex^5 + \cdots$, with x and y small. The reverted series is $x = y - by^2 + (2b^2 - c)y^3 - (5b^3 - 5bc + d)y^4 + (14b^4 - 21b^2c + 6bd + 3c^2 - e)y^5 + \cdots$. For example, with $d = 0$, $e = 0$, $y = -q$, this gives a series solution of the cubic equation $cx^3 + bx^2 + x + q = 0$ for small values of q.

Tables. The values of logarithms for tables may be computed by judicious use of the series

$$\log_e \frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right)$$

The expansions

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

$$\sin^{-1} x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{x^5}{5} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\frac{x^7}{7} + \cdots$$

or

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$$

are useful in computing tables of trigonometric functions. The last series makes it possible to compute π easily by using relations like

$$\frac{\pi}{4} = 2\tan^{-1} {}^1\!/_3 + \tan^{-1}\ {}^1\!/_7$$

or

$$\frac{\pi}{4} = 4\tan^{-1} {}^1\!/_5 - \tan^{-1}\ {}^1\!/_{239}$$

Integrals may often be found from series. For example,

$$\int_0^x e^{-x^2}\, dx = x - \frac{x^3}{3} + \frac{x^5}{5 \cdot 2!} - \frac{x^7}{7 \cdot 3!} + \cdots$$

makes it possible to evaluate the probability integral. For large values of x it is easier to use the divergent but asymptotic expression

$$\int_0^x e^{-x^2}\, dx = \frac{\sqrt{\pi}}{2} - e^{-x^2}\left(\frac{1}{2x} - \frac{1}{2^2x^3} + \frac{3}{2^3x^5} - \frac{3 \cdot 5}{2^4x^7} + \cdots\right)$$

Ordinary differential equations. An ordinary point of a linear differential equation is a value of x at which all the coefficients are analytic, with the first coefficient not zero. For x near such a value, the complete solution may be expressed in terms of Taylor series. For example, $x = 0$ is an ordinary point of

$$(1 - x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + n(n+1)y = 0$$

which is Legendre's equation. For n zero or an integer, one solution is

$$P_n(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(2n - 2k)!}{2^n k!(n-k)!(n-2k)!}$$

the Legendre polynomial of degree n.

Fourier-Legendre series. If, for $-1 < x < 1$, the function $f(x)$ satisfies the conditions for expansion in a Fourier series, there is an expansion

$$f(x) = \sum_{n=0}^{\infty} a_n P_n(x)$$

where

$$a_n = \frac{2n+1}{2} \int_{-1}^{1} f(x) P_n(x)\, dx$$

The sum to $(n + 1)$ terms gives the polynomial of the nth degree best approximating $f(x)$ in the sense of least square error.

Regular singular point. Let a differential equation have the form

$$A_2 \frac{d^2y}{dx^2} + A_1 \frac{dy}{dx} + A_0 y = 0$$

where $A_2(x)$ is analytic and not zero at x_0, $(x - x_0)$ $A_1(x)$ and $(x - x_0)^2 A_0(x)$ are each analytic at x_0. Then if x_0 is not an ordinary point, it is a regular singular point. Near such a point, at least one solution may be found in the form $(x - x_0)^s$ times a Taylor series, where s is some real or complex value. For example, for n zero or any positive integer, the Bessel function of order n

$$J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{n+2k}}{2^{n+2k} k!(n+k)!}$$

is a solution of Bessel's differential equation

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right) y = 0$$

Partial differential equations. Certain boundary-value problems in partial differential equations may be solved by the use of series. Thus, in two dimensions, Laplace's equation in polar coordinates is

$$\frac{\partial^2 U}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2} + \frac{1}{r}\frac{\partial U}{\partial r} = 0$$

Let the values on the boundary of a circular region $r = a$ be given in the form $f(\theta)$, and let the Fourier expansion of this function of period 2π be

$$f(\theta) = A + \sum_{n=1}^{\infty} (A_n \cos n\theta + B_n \sin n\theta)$$

Then the solution of Laplace's equation with $U(a,\theta) = f(\theta)$ is

$$U(r, \theta) = A + \sum_{n=1}^{\infty} \left(\frac{r}{a}\right)^n (A_n \cos n\theta + B_n \sin n\theta)$$

Again, the solution of Laplace's equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0$$

with $U(0,y) = 0$, $U(L,y) = 0$, $U(x, +\infty) = 0$, $U(x,0) = f(x)$, with Fourier sine expansion

$$f(x) = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L}$$

is given by

$$U(x, y) = \sum_{n=1}^{\infty} B_n e^{-n\pi y/L} \sin \frac{n\pi x}{L}$$

See FOURIER SERIES AND TRANSFORMS.

Philip Franklin; Salomon Bochner

Bibliography. W. Kaplan, *Advanced Calculus*, 4th ed., 1991; I. S. Sokolnikoff and R. M. Redheffer, *Mathematics of Physics and Modern Engineering*, 2d ed., 1966; S. K. Stein, *Calculus and Analytical Geometry*, 5th ed., 1992.

Series circuit

An electric circuit in which the principal circuit elements have their terminals joined in sequence so that a common current flows through all the elements.

The circuit may consist of any number of passive and active elements, such as resistors, inductors, capacitors, electron tubes, and transistors.

The algebraic sum of the voltage drops across each of the circuit elements of the series circuit must equal the algebraic sum of the applied voltages. This rule is known as Kirchhoff's second law and is of fundamental importance in electric circuit theory. *See* KIRCHHOFF'S LAWS OF ELECTRIC CIRCUITS.

When time-varying voltages and currents are involved, it is necessary to employ differential or integral equations to express the summation of voltages about a series circuit. If the voltages and currents vary sinusoidally with time, functions of a complex variable are used in place of the calculus. *See* ALTERNATING-CURRENT CIRCUIT THEORY; CIRCUIT (ELECTRICITY); DIRECT-CURRENT CIRCUIT THEORY.

Robert Lee Ramey

Serology

The study of antibodies in the serum or other body fluids as well as the study of the antigens present on cells of the blood. It is the oldest branch of immunology. *See* ANTIBODY; ANTIGEN; ANTIGEN-ANTIBODY REACTION; IMMUNOLOGY.

Blood consists of red and white cells, platelets, and plasma (the fluid medium). After blood coagulates, serum (consisting of plasma minus the clotting proteins) separates from the clot. It is generally easier to prepare and more convenient to study in a laboratory than plasma. Even before antibodies were identified, it was noted that the serum of a person who had recovered from a bacterial infection could agglutinate, or clump, those bacteria, so a simple test became available to help identify the organisms with which a person was, or had been, infected. It is easier and more reliable to measure antibodies in serum than the microorganisms themselves, which often are present at low or undetectable numbers. *See* AGGLUTINATION REACTION; SERUM.

Serologic tests. In modern laboratories there are many techniques for identifying antibodies against

microorganisms. Agglutination is still used, as well as complement fixation, enzyme-linked immunosorbent assays (ELISA), virus neutralization assays, and immunofluorescence. Most serologic tests can be set up quickly in the face of an emerging infectious disease epidemic. *See* COMPLEMENT-FIXATION TEST; IMMUNOASSAY; IMMUNOFLUORESCENCE.

Medicine. Large reference laboratories may offer serologic tests for the detection of over 100 pathogens. Some of the familiar tests include the oldest, such as the serologic test for syphilis, which is required to obtain a marriage license in some jurisdictions. Similarly, many governments require serologic proof of immunity to rubella (German measles) before marriage, because rubella infection early in pregnancy can result in serious damage to the fetus. Serology is also of great importance in veterinary medicine to monitor the health of animal herds as well as individual companion animals. *See* RUBELLA; SYPHILIS; VETERINARY MEDICINE.

Severe acute respiratory syndrome. Within weeks of the first cases of severe acute respiratory syndrome (SARS), an ELISA was available. In this test, a synthetic peptide derived from the sequenced SARS-CoV virus genome is attached chemically to the bottom of a well in a plastic dish. The patient's serum is then added, and the well washed so that only antibody, if present, remains bound to the peptide. Next a second antibody (often from an animal) that is able to generically recognize human antibodies is added; it binds only if SARS antibody has previously bound. This second antibody comes tagged with an enzyme. A colorless substrate for this enzyme is then added, which becomes a colored product whose intensity depends on how much antibody is present in the patient's serum. This test is very sensitive, and it is safe for laboratory workers because they do not have to be exposed to a potentially dangerous pathogen. *See* ANIMAL VIRUS; VIRUS.

Human immunodeficiency virus. Human immunodeficiency virus (HIV) infection is confirmed by serology, using an ELISA test. Since the ELISA test for HIV can sometimes yield false-positive results, a more precise test called a Western blot is used to confirm positive results. HIV highlights the limitations of serology, because infected persons usually do not produce detectable antibodies for about 6–9 weeks, during which time they contain the virus and may be contagious. Although expensive, there are tests that can detect the virus itself and are applied when necessary. *See* ACQUIRED IMMUNE DEFICIENCY SYNDROME (AIDS).

Seroconversion. Seroconversion is the process whereby antibodies to a pathogen become detectable over time due to infection or immunization; thus, initial serologic test results are negative, but subsequent tests yield positive results. In most diseases, concentrations of antibody in the serum (called titers) can remain elevated for months or years after recovery from the illness. This means that a single positive serologic result indicates that the patient has encountered the pathogen but does not mean the person is currently ill from it; thus, a second test is ordered, usually about 2 weeks later, to confirm the results. If the infection is quite recent, the second test may show a rise in the titer; if less recent, the titer may be lower; and if the infection was contracted a long time ago, there is likely to be no change in the titer level.

Forensic science. Serology plays a major role in forensic science. Police serologists can type the cells in various body fluids using similar methods to those used in blood banks. Using enough markers, with high probability they can either rule out or implicate suspects. Forensic serologic techniques can even distinguish between the blood of identical twins: Although the cell antigens and markers of the twins are the same, the pathogens they have been exposed to are likely to differ over a lifetime. By comparing antibody titers in the crime scene samples with those in blood taken from the twins, the stains can be definitively ascribed to one or the other. *See* FORENSIC BIOLOGY.

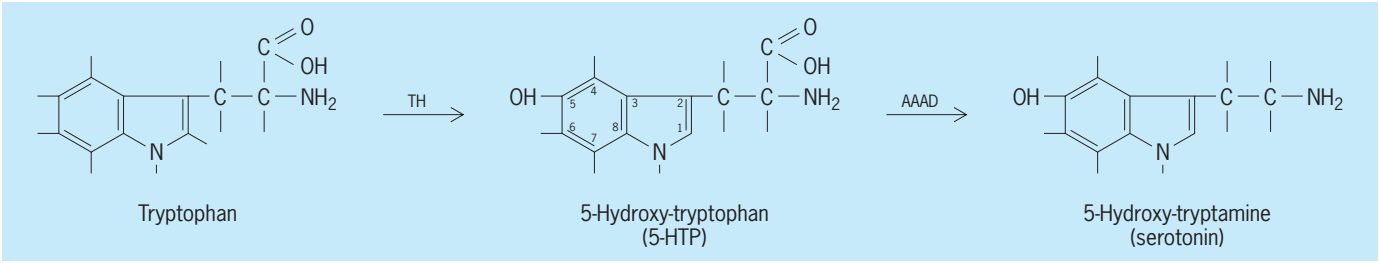
J. John Cohen

Bibliography. R. Burns (ed.), *Immunochemical Protocols* (*Methods in Molecular Biology*), 3d ed., Humana Press, 2005; A. P. Johnstone and M. W. Turner (eds.), *Immunochemistry: A Practical Approach.*, Oxford University Press, 1997; K. Landsteiner, *The Specificity of Serological Reactions*, Dover Publications, 1990 (reprint of original 1945 publication).

Serotonin

A neurotransmitter, also known as 5-hydroxytryptamine (5-HT), that is released throughout most of the brain and spinal cord by the terminals of neurons originating in the raphe nuclei of the brain stem. (These neurons also release serotonin from their dendrites, where it can act on nearby autoreceptors to control the frequency with which the neurons fire.) Serotonin is also released from platelets to promote blood clotting and from enterochromaffin cells in the intestines as a hormone; moreover, in the pineal gland serotonin is an intermediate in the synthesis of the hormone melatonin (5-methoxy *N*-acetyltryptamine).

Serotonin is made from tryptophan, an essential amino acid, and the concentration of tryptophan in the central nervous system (CNS) normally controls the rate at which serotonin is synthesized. Thus, eating carbohydrate foods that elevate brain tryptophan levels (by increasing the amino acid's transport across the blood-brain barrier) can increase serotonin's synthesis and its release into synapses. This unusual ability of serotonin-releasing neurons to couple the release of their neurotransmitter to the composition of each meal allows these neurons to participate in maintaining nutritional homeostasis. However, this ability also underlies the "carbohydrate craving" and weight gain associated with a number of mood disturbances, in which patients learn to use dietary carbohydrates as though they were antidepressant drugs. A large number of protein molecules—including enzymes, receptors,



The synthesis of serotonin from tryptophan. AAAD, aromatic L-amino acid decarboxylase; TH, tryptophan hydroxylase.

and binding proteins—are involved in serotonin's actions, allowing an impressive array of relatively specific drugs to be developed for treating serotonin-related disorders.

Synthesis, and transport across blood-brain barrier. The initial and rate-limiting step in serotonin's synthesis is tryptophan's hydroxylation to the amino acid 5-hydroxytryptophan (5-HTP), a process catalyzed by the enzyme tryptophan hydroxylase. 5-HTP is then decarboxylated to serotonin by a ubiquitous and nonspecific enzyme, aromatic-L-amino acid decarboxylase (AAAD) [see **illustration**]. Tryptophan hydroxylase is a low-affinity enzyme, and the tryptophan levels found in the brain are normally too low to saturate it with substrate. Hence, any process that increases or decreases brain tryptophan levels can rapidly affect the synthesis of serotonin, as well as the amounts released into synapses each time the neuron fires. The main such process is eating: a meal that is rich in carbohydrates but contains little or no protein will cause brain tryptophan levels to rise; however, a high-protein meal or snack will fail to elevate brain tryptophan. These relationships are paradoxical, since carbohydrates contain no tryptophan and do not elevate blood tryptophan levels, while all proteins do contain tryptophan and thus do elevate blood tryptophan levels. These food-induced effects occur due to a characteristic of the transport protein, within the endothelial cells lining brain capillaries, which mediates the transport of tryptophan from the blood into the brain.

The transport protein carries numerous additional large neutral amino acids (LNAAs) besides tryptophan—particularly leucine, isoleucine, valine, tyrosine, and phenylalanine—across the blood-brain barrier, and these compounds all compete with tryptophan for access to the transport sites. Eating most simple and complex carbohydrates (but not fructose) causes insulin to be secreted, which then lowers blood levels of the competing LNAAs by allowing them to enter muscle cells. However, insulin has very little effect on blood tryptophan levels, so tryptophan's ability to compete for transport sites is enhanced, causing brain tryptophan levels to rise. In contrast, most proteins decrease tryptophan's ability to compete for transport sites because they contain 20 times (or more) of the other LNAAs than of tryptophan, so brain tryptophan levels fail to rise and may actually fall. (Eating the Thanksgiving turkey does not raise brain tryptophan and serotonin; however, eating the dessert probably does.) Brain tryptophan can also be raised by administering tryptophan itself; however, this product was withdrawn from the market in 1989, when an impure batch caused a new disease, eosinophilia-myalgia syndrome [characterized by high levels of eosinophils (circulating white blood cells) and debilitating muscle pain], that killed many people.

Metabolism. Serotonin is metabolized by the enzyme monoamine oxidase (MAO), which is present in virtually all cells, including the neurons that produce serotonin. Its action generates an intermediate aldehyde which is then either oxidized to an acid (5-hydroxyindole acetic acid; 5HIAA) or reduced to an alcohol. Although none of these metabolites retains serotonin's ability to transmit signals, the process by which "active" serotonin—that is, serotonin present within a synapse—is inactivated involves not its metabolism but its reuptake into the neuron that released it. Inhibition of MAO (for example, by the original antidepressant drugs), like consumption of a carbohydrate-rich meal or snack, elevates serotonin levels within the terminals of raphe neurons and, consequently, the amounts released into synapses when the neurons fire. *See* MONOAMINE OXIDASE; NEUROSECRETION.

Role in the CNS. Once released into synaptic clefts, serotonin (1) can interact with any one of at least 14 postsynaptic receptors (5-HT receptors) to affect second-messenger synthesis or, via the 5-HT_3 receptor, ion channels, (2) can be inactivated by reuptake, a process inhibited by selective serotonin reuptake inhibitors (SSRIs), such as the antidepressants fluoxetine (Prozac®), sertraline (Zoloft®), and Paroxetine (Paxil®), or (3) can act on autoreceptors on the terminals or dendrites of serotonin-releasing neurons to slow its firing frequency or diminish the amount of serotonin subsequently released per firing. (The time needed to desensitize these autoreceptors may explain why antidepressant drugs can require weeks to become clinically effective, even though they block serotonin's reuptake or inhibit MAO activity soon after they are administered.)

Serotonin in the CNS has numerous known physiological and behavioral roles. Drugs that increase its levels within brain synapses, or interact with its receptors, are used to relieve depression, anxiety, excessively aggressive behavior, and bulimia; to enhance satiety and suppress carbohydrate craving; to treat premenstrual syndrome; to suppress

vomiting (for example, in patients receiving cancer chemotherapy); and to treat migraines. Serotonin release within the brain also promotes sleep, and within the spinal cord, it lowers blood pressure and acts as a "gate," limiting pain-generated signals ascending to the brain. *See* BRAIN; NEUROBIOLOGY; PSYCHOPHARMACOLOGY; PSYCHOTOMIMETIC DRUG; SYNAPTIC TRANSMISSION.

Blood serotonin. Serotonin is present within the blood plasma and platelets. Plasma serotonin can be thought of as a hormone, synthesized by specialized cells, the enterochromaffin cells of the intestine, and secreted into the bloodstream. Plasma serotonin is the source of platelet serotonin, since platelets have a serotonin-uptake protein, such as the one in terminals of raphe neurons, but lack the enzyme (tryptophan hydroxylase) required to synthesize serotonin. When platelets coagulate, for example at the site of an injury, they release serotonin, causing local capillaries to constrict and decrease bleeding. *See* BLOOD.

Serotonin levels in plasma are normally quite low. If they are made to rise substantially—for example, in patients suffering from carcinoid tumors, which release very large quantities of serotonin into the bloodstream—this can cause flushing of the skin, diarrhea, wheezing, and sometimes severe damage to the heart valves. Two biochemical pathways exist for removing serotonin from the plasma: (1) uptake into platelets and (2) destruction by MAO in the liver, kidneys, and elsewhere. If both pathways are blocked by the concurrent administration of an SSRI and a MAO inhibitor, plasma serotonin levels will also rise. This mechanism may explain the heart valve lesions and pulmonary hypertension observed in some patients treated for obesity with fenfluramine plus phentermine ("fen-phen"), since fenfluramine suppressed serotonin uptake while phentermine inhibited MAO. Administration of either drug alone was very rarely associated with such side effects, and their incidence may have been no greater than that found in untreated, control populations. It is illegal to give patients both an SSRI and an MAO inhibitor, but unfortunately not all compounds with some MAO-inhibiting activity are labeled as such.

Melatonin. The mammalian pineal gland can also produce serotonin from circulating tryptophan; however, it further transforms the serotonin to the hormone melatonin. Most of the melatonin made during each 24-h period is synthesized during the hours of darkness. Because of its great lipid solubility, it rapidly diffuses across the cellular membrane to enter the cerebrospinal fluid and the bloodstream without being actively secreted. Hence, there are marked circadian rhythms in human plasma melatonin levels, rising from about 8–10 pg/ml during the day to 100–200 pg/ml at night. The mammalian pineal gland is informed about what time it is, and whether or not light is available, by a special neuronal network that runs from the retinas, through the brain and upper spinal cord, to sympathetic nerves, which stimulate melatonin synthesis by releasing norepinephrine onto the pineal cells. The nocturnal secretion of melatonin is an important factor in initiating and sustaining sleep: Many older people whose largely calcified pineal glands release subnormal quantities of melatonin at night exhibit a characteristic type of insomnia, causing them to awaken frequently during the latter half of the night and then experience difficulty falling back asleep. This type of insomnia can usually be treated by administration of oral melatonin, in the tiny doses needed to restore nighttime blood melatonin levels to what they probably were earlier in life. Melatonin is known to act on several types of receptors within the brain; however, the particular receptor which underlies its sleep-promoting effects (as opposed to its effects on other circadian rhythms) awaits discovery. *See* BIOLOGICAL CLOCKS; HORMONE; PINEAL GLAND.

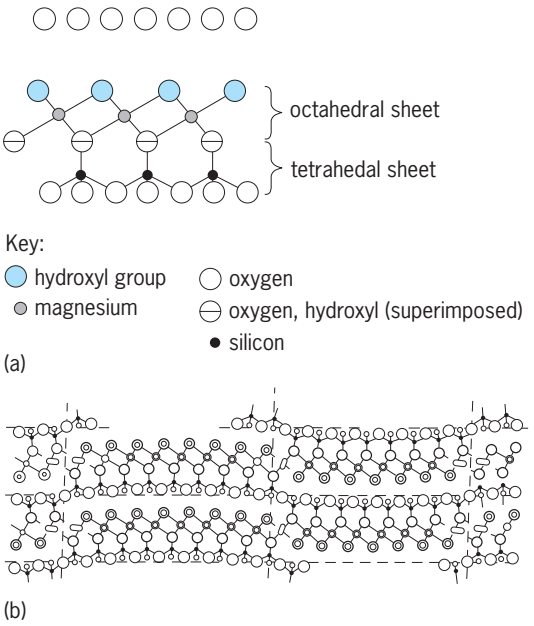
Richard J. Wurtman

Bibliography. J. Cooper, F. Bloom, and R. Roth, *The Biochemical Basis of Neuropharmacology*, 8th ed., 2003; R. J. Wurtman et al., Effects of normal meals rich in carbohydrates or proteins on the plasma tryptophan ratio, *Am. J. Clin. Nutrition*, 77:127–132, 2003; R. J. Wurtman and J. J. Wurtman, Carbohydrates and depression, *Sci. Am.*, 260(1):68–75, 1989; A. F. Schatzberg and C. B. Nemeroff, *Textbook of Psychopharmacology*, 3d ed., American Psychiatric Publishing, Washington, D.C., 2004; I. V. Zhdanova et al., Melatonin treatment for age-related insomnia, *J. Clin. Endocrin. Metab.*, 86(10):4727–4730, 2001.

Serpentine

The name traditionally applied to three hydrated magnesium silicate minerals, antigorite, chrysotile, and lizardite. All have similar chemical compositions [ideally $Mg_3Si_2O_5(OH)_4$] but with three different but closely related layered crystal structures. Serpentine also has been used as a group name for minerals with the same layered structures but with a variety of compositions. The general formula is $M_3T_2O_5(OH)_4$, where M may be magnesium (Mg), ferrous iron (Fe^{2+}), ferric iron (Fe^{3+}), aluminum (Al), nickel (Ni), manganese (Mn), cobalt (Co), chromium (Cr), zinc (Zn), or lithium (Li); and T may be silicon (Si), Al, Fe^{3+}, or boron (B).

Basic crystal structure. The basic layered structure is made up of an octahedral sheet, ideally composed of divalent cations of the M group bonded in octahedral coordination to oxygens (O) and hydroxyls (OH), linked through shared oxygens to a tetrahedral sheet composed of cations of the T group, dominantly silicon atoms, bonded in tetrahedral coordination to oxygens (see **illus.**). Hydrogen bonds between the hydroxyls in the outer surface of the octahedral sheet and the basal oxygens of the overlying tetrahedral sheet bond successive structural layers. All the octahedral sites are occupied in serpentine, unlike structurally related kaolinite, in which only two-thirds of the sites are occupied. In the magnesium-serpentine minerals, this produces an octahedral sheet with slightly larger lateral dimensions than the tetrahedral sheet, and this



Serpentine structures. (*a*) Crystal structure drawing of lizardite showing the octahedral and tetrahedral sheets and the planar nature of the layered structure. (*b*) Crystal structure drawing of antigorite showing the alternating wave structure (*after G. Kunze, Die gewellte struktur des antigorites, I. Zeit., Krystallographica, 108:82–107, 1956*).

misfit produces the different serpentine structures. *See* HYDROGEN BOND; HYDROXYL.

Magnesium-serpentine minerals. Lizardite has a planar structure (illus. *a*), with the misfit accommodated by slight adjustments of the atomic positions within the layers. This structure is similar to that of other layer silicates and clay minerals, and it has a well-developed cleavage parallel to the layers. Curvature of the sheets also relieves the misfit along the direction of curvature, and there are two curved serpentine structures. Chrysotile has a cylindrical structure in which the layers are either concentrically or spirally rolled to produce fiber commonly ranging from 15 to 30 nanometers in diameter, and micrometers to centimeters in length. These fibers have great strength and flexibility and are the most abundant and commonly used form of asbestos. Successive layers in both lizardite and chrysotile can be offset by series of systematic shifts, so that each mineral has a series of slightly different structures, or polytypes. Antigorite has a modulated wave structure, with wavelengths generally varying between 3 and 5 nm (illus. *b*). At each point of inversion, magnesium and hydroxyl groups are omitted relative to the silicate sheet, so that the composition of antigorite is slightly and systematically different from lizardite and chrysotile. *See* ASBESTOS.

Serpentine minerals often form in an environment that does not reach equilibrium, and transmission electron microscope studies often show one structure type passing into another laterally along the layers. Thus, the fine details of structural relationships can be varied and complex. Polygonal serpentine, or polygonal fibers, is a common and interesting form of lizardite. It is composed of either 15 or 30 polygonal sectors of lizardite, often around a chrysotile fiber core. A fourth mechanism for overcoming the misfit has been recognized in the structure of the mineral carlosturanite. Although this is not a layer silicate in the strict sense, it has been identified in several serpentinites, and it may be a commonly occurring mineral. The structure is believed to be composed of a octahedral sheet linked to a tetrahedral sheet modified into strips of tetrahedra, with the breaks between strips occupied by hydroxyls and water (H_2O). This produces a composition with higher magnesium and H_2O and lower silicon contents than the ideal composition.

The magnesium-serpentine minerals usually contain some iron and aluminum substituting for magnesium, and aluminum or ferric iron for silicon, but they may also contain lesser amounts of nickel, manganese, chromium, and cobalt, reflecting the composition of the parent rock. Chrysotile has the smallest range of substitution and antigorite is intermediate, both limited by the curvature of the structure. Lizardite has the greatest range, and it is possible that a solid solution exists between lizardite and amesite $[Mg_2Al[SiAl]O_5(OH)_4]$. The coupled substitution of aluminum for silicon and magnesium reduces the misfit between the octahedral and tetrahedral sheets, making the former smaller than the latter and producing a fifth variation of the layered structure. *See* SERPENTINITE.

Other serpentine minerals. The serpentine group of minerals include, in addition to the magnesium-serpentines, nepouite and pecoraite $[Ni_3Si_2O_5(OH)_4]$, the nickel analogs of lizardite and chrysotile respectively; and brindleyite $[Ni_2Al[SiAl]O_5(OH)_4]$, berthierine $[Fe_2Al[SiAl]O_5(OH)_4]$, kellyite $[Mn_2Al[SiAl]O_5(OH)_4]$, fraipontite $[Zn_2Al[SiAl]O_5(OH)_4]$, and cronstedtite $[Fe^{2+}_2Fe^{3+}[SiFe^{3+}]O_5(OH)_4]$, which can be considered analogs of amesite. In greenalite $[Fe_3Si_2O_5(OH)_4]$ and caryopilite $[Mn_3Si_2O_5(OH)_4]$ the substitutions of iron and manganese produce misfits that are too large to be accommodated in the regular lizardite-type structure, and they have modulated structures. Manandonite $[Li_2Al_4[Si_2AlB]O_{10}(OH)_8]$ is a rare mineral that has come to be recognized as a serpentine mineral. *See* CHEMICAL BONDING; CLAY MINERALS; CRYSTAL STRUCTURE; KAOLINITE; SILICATE MINERALS.

Frederick J. Wicks

Bibliography. S. W. Bailey (ed.), *Hydrous Phyllosilicates (Exclusive of Micas)*, 1988; G. D. Guthrie and B. T. Mossman (eds.), *Health Effects of Mineral Dusts*, 1993.

Serpentinite

A common rock composed of serpentine minerals, serpentinite is usually formed through the hydration of ultramafic rocks, dunites, and peridotites in a process known as serpentinization. The result is the formation of hydrated magnesium-rich minerals, such as antigorite, chrysotile, or lizardite, commonly with magnetite or, less frequently, brucite. *See* DUNITE; PERIDOTITE.

Origin. Serpentinites can be distinguished by, and are named for, the dominant serpentine mineral in

the rock, that is, antigorite-serpentinite, chrysotile-serpentinite, and lizardite-serpentinite. Lizardite-serpentinites are the most abundant. They have been formed in retrograde terrains and are characterized by the pseudomorphic replacement of the original olivine, pyroxenes, amphiboles, and talc by lizardite with or without magnetite or brucite. Each host mineral produces an easily recognized characteristic pseudomorph in which relicts of the original minerals may remain if serpentinization is not complete. Thus, the textures of the preserpentinization rock are preserved by the pseudomorphs; so that a great deal of information on the original grain shapes, fabrics, deformation features, and distribution of the original minerals is retained.

Antigorite-serpentinites can form directly from minerals such as olivine, pyroxene, and so forth in retrograde terrains similar to lizardite, but at a high temperature. However, antigorite-serpentinites most commonly have been formed in prograde terrains, often by the recrystallization of lizardite-serpentinites. The recrystallization of the rock obliterates the previous minerals and textures, replacing them with easily recognized interpenetrating blades of antigorite.

Chrysotile-serpentinites usually occur only in chrysotile asbestos deposits. The complex sequence of serpentinization, recrystallization, folding, and faulting that produce chrysotile asbestos deposits produce serpentinites that are mineralogically and texturally complex. The typical sequence begins with serpentinization of an anhydrous peridotite or dunite to produce a lizardite-serpentinite. Subsequent recrystallization produces a second generation of lizardite, chrysotile, or antigorite, with all three often occurring in a single rock. Obviously, such rocks did not reach equilibrium conditions. *See* ASBESTOS.

Accessory minerals. Magnetite is the main accessory mineral, a by-product of the serpentinization of olivine, and it may outline the original grains, making them easier to recognize. However, it may not be formed or may be resorbed by a later recrystallization event. The serpentinization of dunite often produces brucite rather than magnetite. In lizardite-serpentinites the brucite is intimately associated with the lizardite on a submicrometer scale, but in antigorite and chrysotile-serpentinites it occurs as discrete easily recognized grains. In the reducing conditions that produced some serpentinites, nickel/iron alloys are a minor accessory mineral. *See* BRUCITE; MAGNETITE; OLIVINE.

Calcium in pyroxenes is released during serpentinization, but because it cannot be accommodated in the serpentine crystal structures, calcium-rich fluids are produced. Dikes and blocks of other rocks such as granite, diabase, and gabbro, or rocks at the contacts with serpentinites may be altered by these fluids to form calcium-rich rocks called rodingites. The strong chemical gradient between the magnesium-rich serpentinites and the silica-rich country rock can produce other reactions that result in nephrite jade, jadeite, chlorite/talc, or, if carbon dioxide is present, talc/carbonate reaction zones along the contacts. *See* METAMORPHISM.

The occurrence of serpentinites is widespread, particularly in greenstone belts, mountain chains, and mid-ocean ridges, where they have formed through the serpentinization of ultramafic rocks. Serpentinites also are formed infrequently during the intrusion of diabase dikes into dolomitic marbles. This type of serpentinite is unusual, because it contains no magnetite; and the serpentine minerals contain little or no iron, a reflection of the very low iron content of the original marble. *See* AMPHIBOLE; MID-OCEANIC RIDGE; PYROXENE; SERPENTINE.

Frederick J. Wicks

Bibliography. S. W. Bailey (ed.), *Hydrous Phyllosilicates (Exclusive of Micas)*, 1988; G. D. Guthrie and B. T. Mossman (eds.), *Health Effects of Mineral Dusts*, 1993; D. S. O'Hanley, *Serpentinites: Records of Petrologic and Tectonic History*, 1966.

Serum

The liquid portion that remains when blood is allowed to clot spontaneously and is then centrifuged to remove the blood cells and clotting elements. It is approximately of the same volume (55%) as plasma and differs from it only by the absence of fibrinogen which, after conversion into fibrin, is incorporated in the separated clot. *See* FIBRINOGEN.

Blood serum contains 6–8% solids, including macromolecules such as albumin, antibodies and other globulins, and enzymes; peptide and lipid-based hormones; and cytokines; as well as certain nutritive organic materials in small amounts, such as amino acids, glucose, and fats. Somewhat less than 1% of the serum consists of inorganic substances such as sodium, chloride, sulfates, phosphates, calcium, potassium, bicarbonate, magnesium, iodine, zinc, and iron. Small amounts of respiratory gases such as oxygen and carbon dioxide are dissolved in the serum, as is the gas nitric oxide, which serves as a chemical messenger and vasodilator. Small amounts of waste material such as urea, uric acid, xanthine, creatinine, creatine, bile pigments, and ammonia are also present in the serum. These substances, along with other small molecules which are not bound to blood proteins, are filtered out as blood flows through the kidney. *See* BLOOD; CLINICAL PATHOLOGY; KIDNEY.

Certain types of sera, both human and animals, are used in clinical medicine. Immune serum and hyperimmune serum either are developed by naturally occurring disease or are deliberately prepared by repeated injection of antigens to increase antibody titer for either diagnostic tests or the treatment of active disease. These sera are referred to as antisera, since they have a specific antagonistic action against specific antigens. This ability of antibodies to detect and/or immobilize specific antigenic molecules provides a very important tool for biomedical research. There are also homologous and heterologous sera, the former pertaining to serum prepared and used in animals of the same species, and the latter referring to sera prepared and used on animals of different

species. *See* ANTIBODY; ANTIGEN; BIOLOGICALS; IMMUNITY.

By custom, the clear portion of any liquid material of animal origin separated from its solid or cellular elements is also referred to as sera, for example, the fluid in a blister and the fluids present in certain natural cavities of the body, such as the peritoneal, pleural, or pericardial sacs. These fluids are more properly referred to, however, as effusions. *See* SEROLOGY.

Rueben Straus; Bruce A. Stanley

Servomechanism

A system for the automatic control of motion by means of feedback. The term servomechanism, or servo for short, is sometimes used interchangeably with feedback control system (servosystem). In a narrower sense, servomechanism refers to the feedback control of a single variable (feedback loop or servo loop). In the strictest sense, followed in this article, the term servomechanism is restricted to a feedback loop in which the controlled quantity or output is mechanical position or one of its derivatives (velocity and acceleration). *See* CONTROL SYSTEMS.

The purpose of a servomechanism is to provide one or more of the following objectives: (1) accurate control of motion without the need for human attendants (automatic control); (2) maintenance of accuracy with mechanical load variations, changes in the environment, power supply fluctuations, and aging and deterioration of components (regulation and self-calibration); (3) control of a high-power load from a low-power command signal (power amplification); (4) control of an output from a remotely located input, without the use of mechanical linkages (remote control, shaft repeater).

Many of these objectives could be attained by a nonfeedback system if accurately calibrated components were available. The salient advantage of the closed-loop feedback method of control is that it permits economical design and production of accurate systems using primarily inaccurate components. A typical servomechanism requires only a few of its components to have high accuracy.

Servomechanisms were first used in speed governing of engines, automatic steering of ships, automatic control of guns, and electromechanical analog computers. Today, servomechanisms are employed in almost every industrial field. Among the applications are cutting tools for discrete parts manufacturing, rollers in sheet and web processes (such as steel rolling, paper calendering, and high-speed printing), elevators, automobile and aircraft engines, robots, remote manipulators and teleoperators, telescopes, antennas, space vehicles, mechanical knee and arm prostheses, and tape, disk, and film drives. *See* COMPUTER STORAGE TECHNOLOGY; FLIGHT CONTROLS; GOVERNOR; MAGNETIC RECORDING; MEDICAL CONTROL SYSTEMS; REMOTE MANIPULATORS; ROBOTICS; VIDEO DISK.

Servo Loops

Figure 1 shows the basic elements of a servomechanism and their interconnections; in this type of block diagram the connection between elements is such that only a unidirectional cause-and-effect action takes place in the direction shown by the arrows. The arrows form a closed path or loop; hence this is a single-loop servomechanism or, simply, a servo loop. More complex servomechanisms may have two or more loops (multiloop servo), and a complete control system may contain many servomechanisms. *See* BLOCK DIAGRAM.

In Fig. 1 the objective of controlling the output variable c is accomplished if the feedback signal b can be made identically equal to the reference input signal r. If b and r are not equal, a deviation or actuating signal e is produced. This signal e, suitably amplified, is impressed on the servomotor which corrects the output variable c so as to bring b into coincidence with r. If the amplifier gain is large, the deviation e required for a given corrective action is quite small. Adequate control action can be achieved provided the following conditions are met: (1) The servomotor is large enough to drive the load and perform the required corrections. (2) The measuring devices and their connection to the shafts have adequate accuracy. (3) The summing device is accurate. (4) The amplifier gain is sufficiently large. *See* GAIN.

The accuracy of the forward controlling elements and of the load does not enter into these considerations. In other words, with a sufficiently

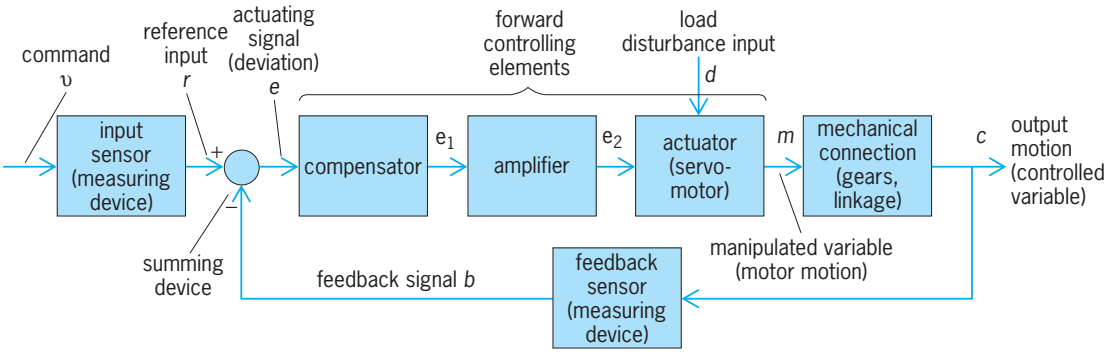


Fig. 1. Servo loop elements and their interconnections. Cause-and-effect action takes place in the directions of arrows. (*After American National Standards Institute, Terminology for Automatic Control, ANSI C85.1*)

large motor and sufficiently high amplifier gain, the accuracy of the system is dependent primarily on the accuracy of the measuring and summing devices and is independent of the characteristics of the high-power portion of the servo, amplifier, motor, and load. Variations in amplifier, motor, and load characteristics, either from unit to unit in mass production or in one unit over a period of time, have little effect on the servo accuracy. The devices which determine the overall accuracy, the measuring and summing elements, are low-power components, and they can be fabricated to a high order of accuracy. A high-grade control requires accurate means of measurement and instrumentation. The quality of the servomechanism can never exceed that of the instruments which measure input and output motion.

The use of high amplifier gain in feedback systems tends to cause overcorrection, and the resulting system may be poorly damped or unstable. The purpose of the compensator in Fig. 1 is to minimize or eliminate these instabilities and to provide accurate, rapid, and smooth response. The functions of the summing device and compensator are usually provided by a single component called the controller. *See* CONTROL SYSTEM STABILITY; DAMPING.

Servomechanisms may be classified in several ways. The following classifications show some of the many available variations:

1. Positional or velocity servo, depending on whether the variable to be controlled is the output position or the output velocity.
2. Electric, hydraulic, pneumatic, or mixed, depending on the form of the energy used in actuation, measurement, and control.
3. Analog, pulse, digital, or mixed, depending on the way the measurement, actuation, signal conditioning, and control functions are implemented.
4. Proportional or on-off (relay) servo, depending on the number of levels in the motor drive signal.
5. Translational or rotational, depending on the type of actuator and on the output motion.
6. Direct-drive or geared, depending on the manner of connection between actuator and load.
7. Direct-current (dc), alternating-current (ac), pulse-modulated, or mixed, depending on the electrical signal format.
8. Absolute value or incremental motion, depending on whether the input command and feedback sensor possess a datum or zero level.
9. Power or instrument servo, the latter referring to a small servo whose load consists primarily of displays and sensors.

A few of these combinations are described more fully below.

Positional servomechanism. The feedback sensor generates a signal representing a measured position of the output shaft. **Figure 2** is an example of a positional servo employing a dc motor as the actuator and a resistance potentiometer as a feedback sensor. The input is a shaft position which is measured by an identical potentiometer. A change in the input command causes a change in the reference voltage r, which produces a deviation e. This in turn causes a voltage to be applied to the motor armature, which then repositions the load so as to reduce the deviation to zero. This servo is called a shaft position repeater, because the output and input shaft positions are equal at equilibrium. This is an example of an analog dc servo with proportional actuation.

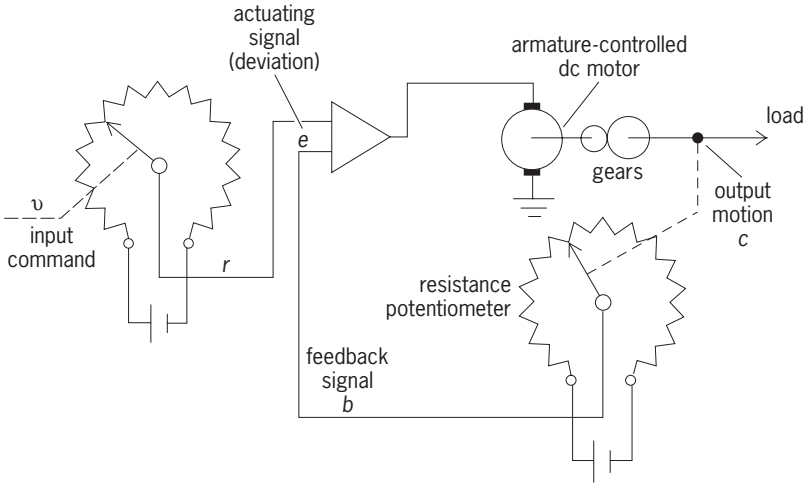


Fig. 2. Positional servomechanism, employing a dc motor as the actuator and a resistance potentiometer as a feedback sensor.

Velocity servomechanism. This is also called a rate servo. The feedback-measuring device generates a signal representing a measured value of the velocity of the output shaft. **Figure 3** is an example of a rate servo employing a dc motor as actuator and a dc generator as a feedback sensor. The speed is controlled by the electric reference voltage r. A transient disturbance such as a change in load causes a transient change in speed, after which the speed recovers to its original value. In the example shown, the servo does not have any positional reference; that is, the shaft speed returns to its original value after a transient disturbance, but the shaft may not be in the proper position. More sophisticated digital speed control systems are able to provide both velocity and position control (phase lock).

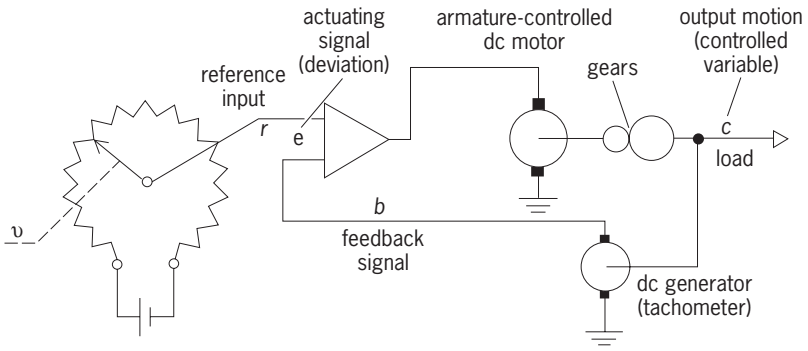


Fig. 3. Velocity or rate servomechanism, employing a dc motor as actuator and a dc generator as a feedback sensor.

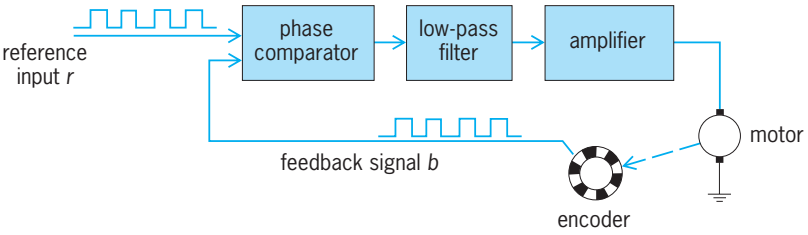


Fig. 4. Phase-locked servo loop.

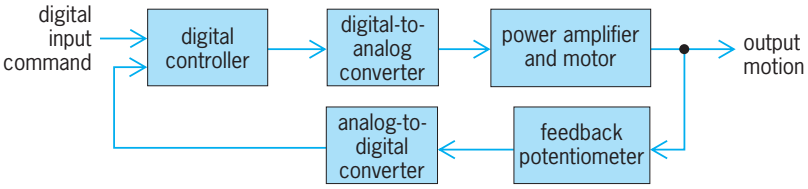


Fig. 5. Digitally controlled positional servomechanism.

The phase-locked servo of **Fig. 4** employs an encoder or pulse tachometer as the feedback sensor. The command signal and the feedback signal are square waves or sine waves of variable frequency. The servo is phase-locked when the frequencies and phases of the two signals become identical. As long as the servo is phase-locked, the motor velocity follows the frequency of the command independently of component parameter variations. The phase-locked servo is really a positional control system used for velocity control: in the event of a disturbance, such as a change in load torque, the motor will quickly recover to its original velocity and position.

Digital servomechanism. Digital control has many advantages over the analog control schemes of Figs. 2 and 3. A digitally controlled positional servo (**Fig. 5**) has digital-to-analog and analog-to-digital converters that interface the digital microcontroller with the motor and the feedback potentiometer, both of which are analog components. *See* ANALOG-TO-DIGITAL CONVERTER; DIGITAL CONTROL; DIGITAL-TO-ANALOG CONVERTER.

Stepping servomechanism. The stepping servo employs a special kind of motor, a stepping motor (also called a step motor or stepper). Any electric motor will move in discrete steps when it is energized from a sequence of discrete voltage pulses, but the size of these steps cannot be exactly controlled. The stepping motor turns by a precisely determined angle with each pulse. It is basically an ac synchronous motor with special construction features for stepping service. *See* STEPPING MOTOR.

Because of its precise response, positional control with a stepping motor is often done open-loop, except for an index mark that determines a starting point or zero positional reference. Starting at the index, the stepper turns by a predetermined angle for each pulse. An error due to a lost or spurious pulse corrects itself as soon as the stepper returns to the index position.

A closed-loop stepping servo is used when errors due to lost or spurious pulses are unacceptable, such as in speed control. Feedback makes it possible to drive the motor at a higher stepping rate, and permits rapid resumption of positional control after a power interruption. The feedback transducer can be an analog component, such as a resistance potentiometer or a dc tachometer generator, or it can be an optical encoder. The stepping servo with encoder feedback is sometimes called an all-digital servo, but the stepper and the incremental encoder are really pulse analog rather than digital devices. Pulse analog signals are much easier to interface with digital controllers than the customary voltage-amplitude analog signals.

Alternating-current servomechanism. It is sometimes desirable to use inductive and capacitive transducers, such as synchros and differential transformers, which require ac excitation. The mechanical input to such a transducer results in a suppressed-carrier modulated output voltage (double-sideband amplitude modulation without carrier). Currently a mixture of ac-dc analog signals or of ac analog with digital signals is used. A typical application is in 360° position control, for example, an antenna azimuth drive, which employs a synchro as a position sensor and a dc motor as an actuator. Synchronous detectors (phase-sensitive rectifier circuits) convert the synchro output ac voltage to dc. Synchro-to-digital and digital-to-synchro converters interface with digital controllers and computers. *See* MODULATION; SINGLE SIDEBAND.

Hydraulic and pneumatic servomechanisms. Hydraulic and pneumatic drives have a number of advantages over electric drives. The principal one is their ability to develop a force at zero speed without power dissipation; on the other hand, obtaining a standstill or holding force from an electric motor or solenoid entails power dissipation and temperature rise which severely limit the available force. This

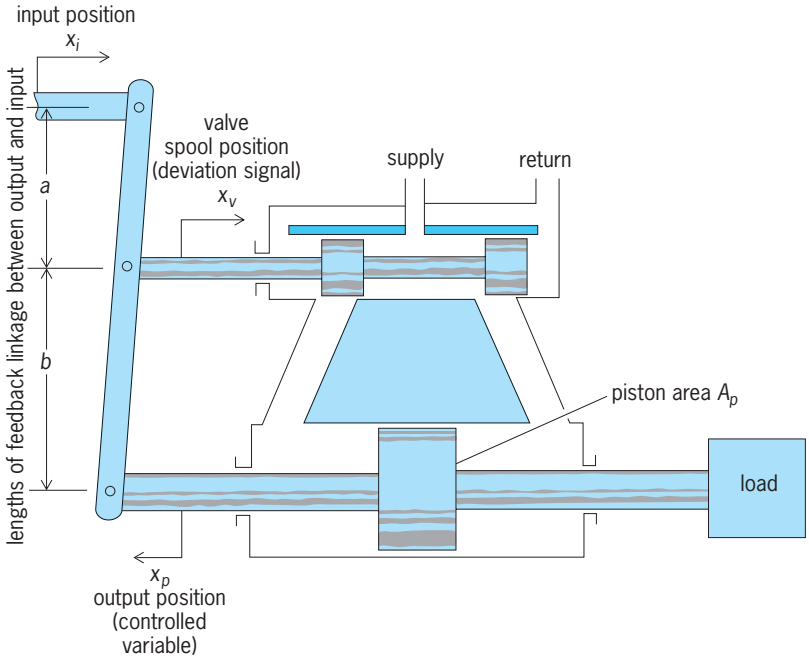


Fig. 6. Hydromechanical servomechanism, with two-stage servovalve. (*After H. E. Merritt, Hydraulic Control System, John Wiley and Sons, 1967*)

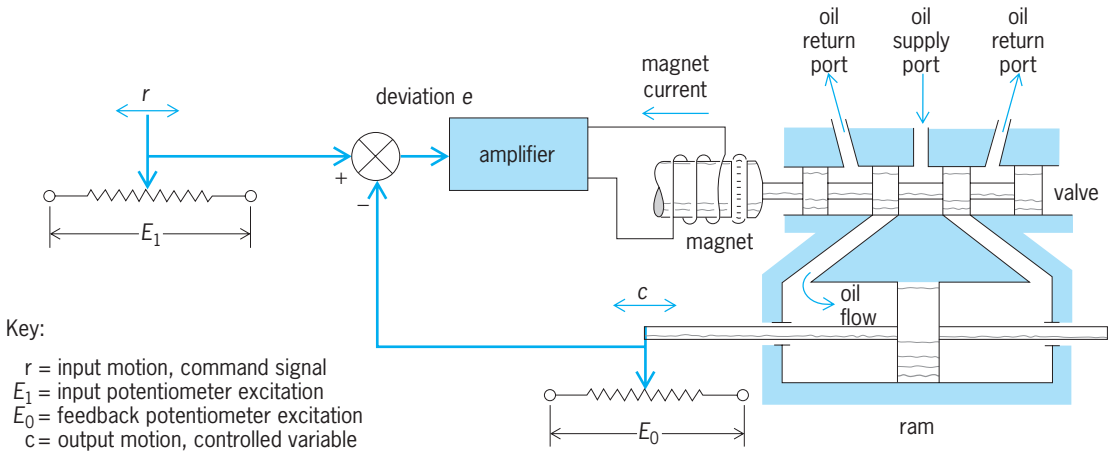


Fig. 7. Electrohydraulic servomechanism. (*After R. N. Clark, Introduction to Automatic Control Systems, John Wiley and Sons, 1962*)

advantage is a major consideration in positional servomechanisms where operation at standstill is often the rule. For a hydraulic drive, other advantages are smaller weight per output power and faster response time. *See* HYDRAULIC ACTUATOR.

Hydraulic power systems are either valve-controlled or pump-controlled. A valve-controlled system consists of a servovalve controlling the flow from a constant-delivery pump to the motor. A pump-controlled system has a variable-delivery pump supply fluid to the motor. Valve control is faster, but less efficient because servovalves are dissipative. Pump control is slow, bulky, but power-efficient, and may be preferred at very high power levels.

Figure 6 shows a valve-controlled hydraulic or hydromechanical servo. The input motion x_i opens and closes the ports of the servovalve, the upper part of the mechanism, which in turn controls the pressure applied to the sides of the piston, the lower portion. The piston motion is fed back and subtracted from the input motion so as to null the servovalve. A hydraulic pump (not shown) supplies the fluid pressure.

The electrohydraulic servo of **Fig. 7** uses resistance potentiometers for input and feedback motion sensing. The electric error signal is amplified and applied to the winding of an electromagnet, generating a force which positions the hydraulic valve spool.

An interesting variation of the above two schemes is an electrohydraulic servo with two-stage servovalve. This two-stage valve is precisely the hydromechanical servo of Fig. 6.

Figure 8 shows a pneumatic servo, called a valve positioner, which is used as a component in process control systems. The input signal, a variable air pressure from the process controller, is converted to motion of a linkage by the air bellows. This adjusts the opening of the pilot valve, which in turn controls the pressure applied to the spring-loaded diaphragm of the process valve or valve motor. The diaphragm converts this pressure change into a motion of the valve stem and thereby controls the flow of the process stream through the main valve body. The main valve stem motion is fed back through the lever and subtracted from the input motion so as to null the opening of the pilot valve.

The purpose of such a servo is to improve the operating characteristics of valves used in industrial process control. Such valves must have good seals between stem and body to isolate the process stream from the instrument air, and these seals exhibit high Coulomb friction or "stiction." Sticky valves are responsible for a phenomenon called mechanical hysteresis, which makes it difficult to control the performance of the system. The servomechanism, by enclosing the sticky valve in a feedback loop, reduces the hysteresis and improves smoothness of operation. *See* HYSTERESIS; PROCESS CONTROL.

The dynamic characteristics of hydraulic and pneumatic servos depend on the interaction of motor and load and on their interconnection. The effect of fluid compressibility cannot be ignored in hydraulic systems, particularly in the case of large masses and long fluid columns. Hydraulic liquids are

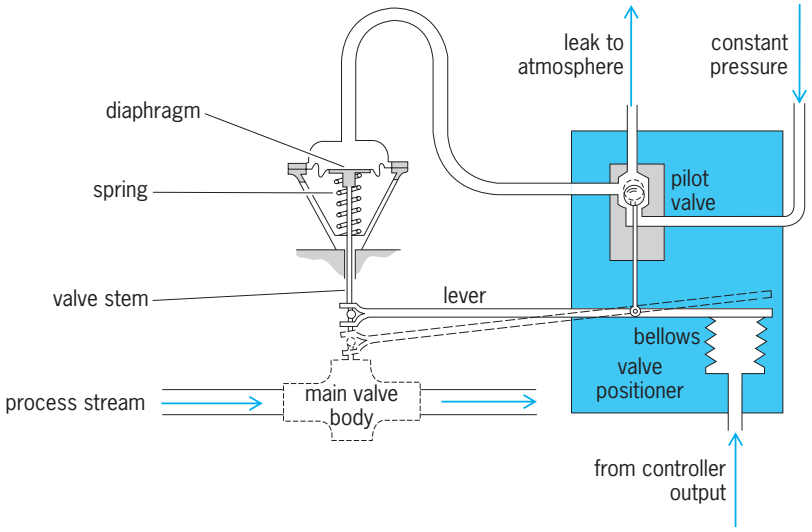


Fig. 8. Pneumatic servomechanism, or valve positioner. (*After W. A. Hadley and G. Longobardo, Automatic Process Control, Addison-Wesley, 1963*)

selected for low compressibility and must be kept free from entrained air. In pneumatic systems the compressibility is always one of the major factors in the system dynamics.

All-hydraulic and all-pneumatic servos such as the ones shown in Figs. 6 and 8 have rather limited applications. Mixed electric systems are the rule in fast-acting, high-accuracy servomechanisms. Electric motors (torquers) or electric servos are used to move the spool of the servovalve or the lever of the delivery pump. The power package (drive) is hydraulic or pneumatic, but all low-level components are electric.

Servomotors

The servomotor, actuator, or final control element is the "muscle" of the servomechanism. It must be large enough to move the load over the required range of velocities and accelerations. Load may mean inertia, friction, or external torque, or any combination thereof.

The motor is the middle part of a three-part power package consisting of (1) the power amplifier which raises the low-level controller output signal to the required power level, (2) the motor itself, and (3) the connection between motor and load, such as gears, cables, or, in some cases, merely a direct solid shaft connection. The combination of these elements is called a power drive. The drive itself often accounts for most of the weight and volume of the complete servomechanism. Servo dynamic performance, such as stability and response time, is determined primarily by the combined parameters of the power drive and load. All these elements interact, and they cannot be studied in isolation. In some servos the load is negligible and the servo dynamics depends primarily on the drive alone.

Electric motors employed in servomechanisms are usually standard motors developed for general (open-loop) use. This is particularly so in the case of large motors, 1 hp (0.75 kW) and more, where special modifications would be very costly. For small sizes, manufacturers have developed special motors for servomechanisms, originally as a matter of necessity, and subsequently to optimize designs. Three areas critical to servomechanisms are stickiness, response time, and shaft compliance.

Stickiness in electric servomotors has the same deleterious effect on system performance as described above for the pneumatic valve positioner. In the electric motor the stickiness is due not so much to mechanical friction but to an effect called magnetic cogging. The nonuniformity of the motor air gap due to the winding slots causes torque fluctuations (torque ripple) as the motor rotates; the motor has a tendency to lock itself into a position of minimum energy.

The response time of a motor is limited by the moment of inertia of its rotor, by the inductance of its winding, and by the maximum torque which can be generated without overheating. Inertia and inductance must be minimized, and torque per watt maximized.

Shaft compliance is reduced and backlash is minimized if as many moving components as possible can be assembled and mounted on a common shaft. One step in this direction is the integrally mounted motor-tachometer or the integrally mounted motor-encoder. Another approach is the gearless or direct-drive connection from motor to load.

Direct current servomotor. This type is favored in servo applications because its torque and speed can be readily controlled by varying the applied voltage. Small- and medium-size dc servomotors have permanent-magnet fields and are controlled by a voltage applied to the armature circuit (armature control). Large dc motors require field windings, and can be controlled by varying the applied field voltage (field control), by armature control or by a combination. A principal drawback of the dc motor is the complexity, limited life, unreliability, and electromagnetic interference associated with its commutator and brushes. *See* COMMUTATOR.

The amplifier that controls the dc servomotor from the available ac power source can have a number of forms. Linear amplifiers are employed at power levels up to about 100 W. Large dc motors are driven from switching amplifiers, thyristors (controlled rectifiers), or magnetic amplifiers. For very large drives, a scheme called the Ward-Leonard system uses a separately excited dc generator to supply the armature voltage of the dc servomotor. A small change in generator field voltage produces a large change in armature voltage and power; hence the term rotating amplifier. The shaft of the dc generator is mechanically driven by an ac motor, or perhaps by a diesel engine, which acts as a power source and runs at near-constant speed. *See* AMPLIFIER; DIRECT-CURRENT MOTOR; MOTOR; SEMICONDUCTOR RECTIFIER.

Besides the conventional dc servomotor, a number of specialized dc motors tailored to servomechanism applications are in use, such as the moving-coil motor, the disk-armature motor (printed motor); and the brushless dc motor.

Moving-coil motor. This type of motor superficially resembles a standard dc motor, with a permanent-magnet-field structure on the stator and an armature on the rotor, but the armature structure (**Fig. 9***a*) is hollow (cup or shell armature). Inside this rotating shell is a stationary iron core, which serves as a magnetic return path for the field flux. Thus, the magnetic path is stationary in its entirety, and torque ripple is virtually eliminated. Integrally mounting the armatures of a moving-coil motor and a dc tachometer-generator on a common shaft (Fig. 9*b*) eliminates the backlash between motor and tachometer and minimizes the compliance.

Brushless dc motor. In this type of motor, the mechanical commutator and brushes of the conventional dc motor are replaced by position sensors and electronic switches. A practical design (**Fig. 10**) has its armature winding carried on the stator and its permanent-magnet field on the rotor. This reversal of the usual dc motor arrangement eliminates all sliding contacts and improves heat transfer out of the motor.

Fig. 9. Cup armatures of a moving coil servomotor. (*a*) Open basket-type construction designed for high power dispersion when air-cooled. (*b*) Resin-molded moving coil winding and integrally coupled tachometer-generator armature. (*Robbins Myers/Electro-Craft Corp.*)

Alternating current servomotor. This type does not require a commutator, and is thus generally more reliable and economical than the dc motor. It is inherently fixed-speed, and its use as a control motor requires a variable-frequency power supply. The frequency synthesizer, a microprocessor-controlled electronic switching amplifier, converts the available fixed-frequency power into a controllable variable-frequency source. The two-phase servo induction motor is a special design with variable speed characteristic, but its inefficiency limits its size to a few watts. Even in these low-power applications, it has been largely superseded by the dc servomotor or by the stepping motor. As stated above, the stepping motor is essentially an ac synchronous motor, and its speed is determined by the applied pulse rate. *See* ALTERNATING-CURRENT MOTOR; INDUCTION MOTOR; SYNCHRONOUS MOTOR.

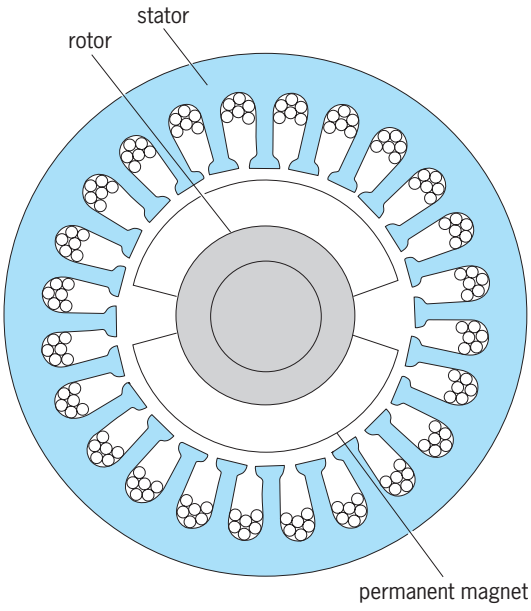


Fig. 10. Cutaway view of brushless dc motor assembly. (*After Electro-Craft Corp. Staff, Handbook of DC Motors, Speed Controls, and Servo Systems, 5th ed., Robbins & Myers/ Electro-Craft Corp., 1980*)

Unconventional motors. When the motion to be controlled is rotary but limited in angle, or when it is translational (rectilinear), an unconventional electric motor may be superior, particularly in small instruments such as cameras, magnetic-disk players, video-disk players, and gyroscopes.

A motor for limited rotational control is required to supply a specified torque, but it never really picks up full speed. Such a motor is called a torquer (or torque motor or torque generator). In direct-drive (gearless) designs, the torquer is often supplied in the form of separate rotor and stator pieces without bearings or housing, for direct installation in the user equipment.

Translational motors can take a variety of forms. The voice-coil motor resembles a conventional electrodynamic loudspeaker movement. A piezoelectric crystal is often employed as the actuator (bender, pusher, puller) for tiny motions, such as the positioning of lenses and prisms in optical instruments. Magnetostriction is also used in motors for fine motion control. The linear induction motor is a stretched-out, translational form of the ac induction motor that has been used in shuttle mechanisms and for proposed high-speed electric railroads. *See* LOUDSPEAKER; MAGNETOSTRICTION; PIEZOELECTRICITY.

Measuring Devices (Sensors)

The quantities measured in motion control are displacement, velocity, acceleration, and force, both in translation and in rotation. Many physical phenomena have been exploited in instrument design. Most sensors and transducers used in laboratory and industrial practice have been or can be adapted to servos. Electrical and optic devices predominate. A few of the most common transducers will be described. *See* PHYSICAL MEASUREMENT; TRANSDUCER.

Analog transducers. The resistance potentiometer is used primarily to measure limited-angle rotation, but can also be designed for translation. The servo potentiometer is a precision component with high-quality bearings. Resistance linearity may be as good as 0.05% of full scale. *See* POTENTIOMETER.

The differential transformer is frequently employed in the control of small translational motion. It is an inductive transducer and requires ac excitation. The generation of the ac input with the conversion of the output to dc is available packaged inside the transformer. *See* DIFFERENTIAL TRANSFORMER.

Synchros are angular position sensors that can handle the measurement of limitless 360° motion. They are inductive devices with ac output, and come in two kinds. Torque synchros, always operated in pairs, are induction motors with interconnected rotors which make them self-aligning or self-synchronous. They require no further electronic

amplifiers, but their accuracy is poor. They have been largely superseded by the family of control synchros: transmitters, differentials, control transformers, and resolvers. The control synchro is an outstandingly accurate component (typically 3 arc-minutes, equivalent to 0.014% of full travel).

The dc tachometer is the most common analog transducer for velocity (rate generator). Physically it is a small permanent-magnet dc generator, designed to obtain high linearity, high stability with temperature, and low output ripple. Linearity of 0.2% of full scale is achievable. Servo tachometers are supplied as individual bearing-mounted components, or as part of an integral motor-tachometer unit with a common shaft and, possibly, a common field magnet. *See* TACHOMETER.

A pulse tachometer can be constructed by depositing regularly spaced tick marks along the edge of the body whose speed is to be measured, be it in rotation or translation. The tick marks are detected optically or magnetically and used to generate pulses. The observed pulse rate or pulse repetition frequency is proportional to speed. When the pulse tachometer is supplied as a separate bearing-mounted unit, it is effectively an encoder, described below.

Gyroscopic and seismic instruments sense and measure absolute motion in space, and are employed in the control of systems mounted on moving vehicles and stabilized platforms, such as cameras, antennas, and guns. *See* ACCELERATION MEASUREMENT; ACCELEROMETER; GYROSCOPE.

Optical encoders. Encoders come in two kinds, absolute encoders and incremental encoders. The absolute encoder is a position transducer with output in the form of parallel binary digits; it is in effect an electromechanical analog-to-digital converter. The incremental encoder produces a single output pulse for a given displacement. It is in effect a pulse tachometer and can be used to measure either position or velocity. The encoder can be constructed to measure rotation (rotary encoder; formerly known as shaft encoder) or translation (linear encoder).

Absolute encoder. The heart of the absolute encoder is a patterned code disk with a number of tracks, one track per binary digit (**Fig. 11**) constructed of alternating transparent and opaque segments and read by means of a light source, focusing system, and photosensor array. Typical resolution is from 6 to 22 binary digits (bits) in 1- to 10-in. (2.5- to 25-cm) disk sizes. (Twenty-two bits represent a 0.3-arc-second angular resolution.) Because of the difficulty of fabricating small disks with many tracks, higher-resolution encoders are sometimes multiturn units, with two disks geared together.

It is not possible to construct an encoder with such mechanical perfection that all sensors will consistently operate in perfect synchronism. As a result, the natural binary code illustrated in Fig. 11 results in errors when there is a transition which requires a change in more than one digit, for example, in going from binary 011 to binary 100. All digital codes in which adjacent numbers can differ in more than one digit (polystrophic codes) suffer from the same

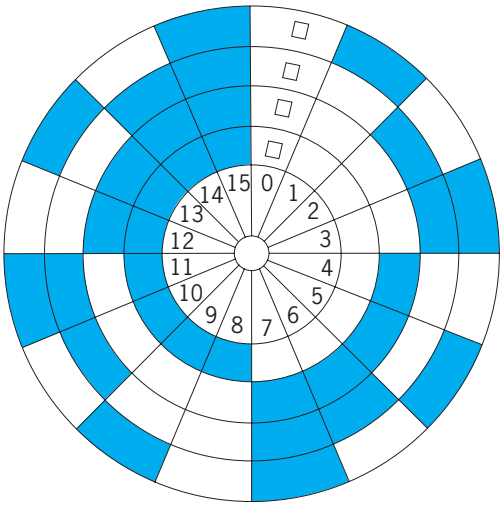


Fig. 11. Four-digit encoder disk. (*After E. O. Doebelin, Measurement Systems, 3d ed., McGraw-Hill, 1983*)

switching ambiguity. There are two solutions to this problem, the addition of extra sensors, or the employment of a special binary code which changes only one digit at a time (monostrophic code), such as a reflected or cyclic code, of which the Gray code is the most common. The **table** shows the Gray code for a three-bit disk; all transitions between adjacent numbers require only one bit change, including the transition from the last to the first number. The output of an encoder may actually be in any of the codes employed in computers, such as binary-coded-decimal (BCD) or excess-3. Logic circuits to convert from the disk code to the output code are then packaged inside the encoder housing. *See* INFORMATION THEORY; LOGIC CIRCUITS.

Older encoders were read by electric contacts, (brushes), magnetically, or capacitatively. A brush encoder disk has conductive and nonconductive segments read by a set of brushes riding on each track. A magnetic code disk has magnetized segments sensed by tiny toroidal coils serving as reading heads. To avert switching ambiguities, brush encoders have two brushes per track, arranged in various configurations known as U-scan and V-scan. Diode logic circuits select the proper sequence of brushes to provide ambiguity-free output.

Incremental encoder. In its simplest form, the incremental encoder has only one track with on-off segments. The number of pulses produced by the sensor is a measure of the displacement, while the pulse

Natural binary and Gray codes

Decimal	Natural binary	Gray
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

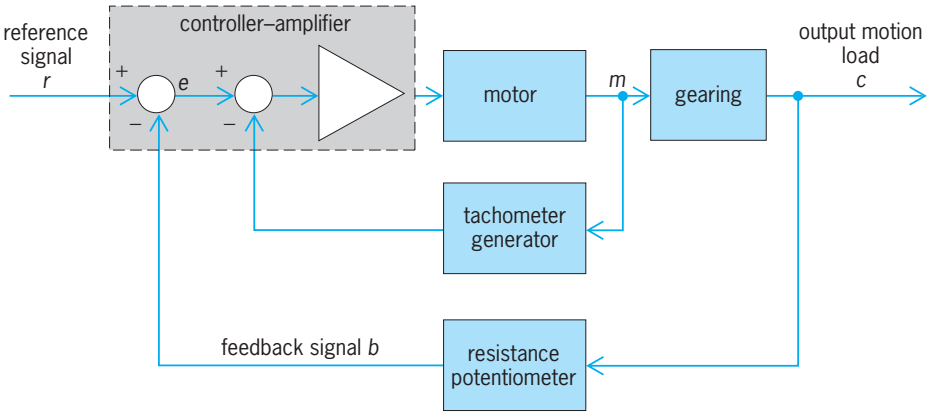


Fig. 12. Tachometer-damped positional servomechanism.

rate indicates the velocity. For absolute motion measurement, the incremental encoder is used with an electronic counter which totalizes the output pulses. Bidirectional sensing is accomplished by employing two identical tracks offset by one-half of a segment length. A third track which generates one pulse per revolution is sometimes provided as a zero marker or index. In the event of a signal interruption or power failure, the error due to lost pulses can be corrected when the index is sensed. The incremental encoder is also used as a unidirectional (one track) or bidirectional (two tracks) pulse tachometer.

Compensation (Control Algorithm)

A servomechanism must be stable and exhibit specified static and dynamic accuracy. These requirements are usually incompatible; the trade-off is that an increase in loop gain improves static accuracy but makes the servo less stable. The process of overcoming this conflict and achieving simultaneously the desired degree of accuracy and stability is called compensation (or stabilization) in analog technology; the term control algorithm has a similar meaning for digital systems.

Static accuracy depends on the zero-frequency (static) loop gain, which becomes infinite if the feedback loop contains an integration. This is indeed the case in the positional servo which uses a conventional motor with output speed proportional to input voltage. The desired output variable, motor displacement, is the integral of motor speed. Hence, the positional servo contains an integration and can have zero static error for a constant input. In Fig. 1, the actuating signal e that represents this error will be zero when the command v is constant and the disturbance d is zero. The velocity servo, on the other hand, does not include an inherent integration, and the loop error as exhibited by the signal e will not be zero. Feedback loops are classified as types 0, 1, 2, and so forth, according to the number of integrations enclosed in the loop. The higher the type number the better the static accuracy, but the poorer the stability.

Standard analog or digital controllers in the compensator block of Fig. 1 implement the proportional-integral-derivative (PID) function and can produce the control modes proportional control (P), proportional-plus-integral (PI), proportional-plus-derivative (PD), or combined PID control. Compensators can also be characterized in terms of their frequency response, the simplest functions being lag, lead, and lag-lead. To some extent, this is only a change in terminology: phase lead compensation is equivalent to PD control, phase lag is an approximation to PI control, and lag-lead approximates PID control. However, integral control adds an integration to the loop, while phase lag does not. This small difference may be significant in the positional servo because PI control converts it into a type-2 loop, which could be unstable. For large loads with structural resonances, more elaborate control algorithms such as notch filters (band-rejection filters) may be required. *See* ELECTRIC FILTER.

Some other methods of compensation do not fit into the simple framework of Fig. 1. For positional servos, rate feedback (**Fig. 12**) is often very effective. Here, a second, inner, loop is added that feeds back a signal proportional to output velocity; this scheme is a form of linear-state feedback. The tachometer for the auxiliary feedback need not be accurate but must be tightly coupled to the motor. In positional or velocity servos with large torque load, it is possible to improve the performance by feeding a signal proportional to torque into the controller. This is, in fact, a feed-forward control scheme. *See* LINEAR SYSTEM ANALYSIS.

Gerald Weiss

Bibliography. B. W. Andersen, *The Analysis and Design of Pneumatic Systems*, 1967, reprint 2001; G. Biernson, *Principles of Feedback Control*, 2 vols., 1988; E. B. Canfield, *Electromechanical Control Systems and Devices*, 1965, reprint 1977; Y. Dote and S. Kinoshita (eds.), *Brushless Servomotors: Fundamentals and Applications*, 1990; Electro-Craft Corp. (Eden Prairie, Minnesota) Staff, *Handbook of DC Motors, Speed Controls, and Servo Systems*, 5th ed., 1980; H. E. Merritt, *Hydraulic Control Systems*, 1967, reprint 1991; R. T. Pickett, *Feedback Control Systems for Technicians*, 1988; J. J. Pippenger, *Hydraulic Valves and Controls*, 1984; D. P. Sante, *Automatic Control System Technology*,

1980; R. B. Walters (ed.), *Hydraulic and Electro-Hydraulic Control Systems*, 1992.

Set theory

A mathematical term referring to the study of collections or sets. Consider a collection of objects (such as points, dishes, equations, chemicals, numbers, or curves). This set may be denoted by some symbol, such as X. It is useful to know properties that the set X has, irrespective of what the elements of X are. The cardinality of X is such a property.

Cardinality of sets. Two sets A and B are said to have the same cardinal, written $C(A) = C(B)$, provided there is a one-to-one correspondence between the elements of A and the elements of B. For finite sets this notion coincides with the phrase "A has the same number of elements as B." However, for infinite sets the above definition yields some interesting consequences. For example, let A denote the set of integers and B the set of odd integers. The function $f(n) = 2n - 1$ shows that $C(A) = C(B)$. Hence, an infinite set may have the same cardinal as a part or subset of itself.

Subset. A is called a subset of B if each element of A is an element of B, and it is expressed as $A \subset B$. The collection of odd integers is a subset of the reals. Also, each set is a subset of itself.

Continuum hypothesis. An infinite set is called uncountable if it cannot be put in a one-to-one correspondence with the positive integers. Here is one of the unsolved problems of set theory. It is of particular interest since so many mathematicians have tried unsuccessfully to solve it. If X is an uncountable subset of the reals R, is $C(X)$ equal to $C(R)$? The conjecture that the answer is in the affirmative is called the continuum hypothesis. It has been shown that the answer cannot be decided by using the ordinary axioms of set theory.

Comparing cardinality of sets. One says that $C(A) \leqq C(B)$ if there is a one-to-one correspondence between the elements of A and a subset of the elements of B. One useful theorem that can be proved states that any two sets A, B are comparable, that is, either $C(A) \leqq C(B)$ or $C(B) \leqq C(A)$ (possibly both). Another theorem states that if $C(A) \leqq C(B)$ and $C(B) \leqq C(A)$, then $C(A) = C(B)$. Each of these results may be proved by using well orderings of A and B.

Ordering. An ordering is one way of setting up a one-to-one correspondence between two sets of the same cardinality. A relation $<$ is an order relation for a set X if it satisfies the following conditions:

1. If x_1, x_2 are two elements of X, either $x_1 < x_2$ or $x_2 < x_1$ (any two elements are related).
2. $x_1 \not< x_1$ (no element is less than itself).
3. If $x_1 < x_2$ and $x_2 < x_3$, then $x_1 < x_3$ (the order relation is transitive).

The following are examples of ordered sets: a horizontal line where $<$ means "to the left of"; the reals where $<$ means "is less than"; and the collection of words in the dictionary where ordering is alphabetical.

Well ordering. An ordering of a set is called a well ordering if it satisfies the additional condition:

4. Each non-null subset Y of X has a first element; that is, there is an element y_o of Y such that if y' is another element of Y, $y_o < y'$.

The natural ordering of the positive integers is a well ordering, but the natural ordering of neither the integers nor the reals is a well ordering. A well ordering for the integers is 0, 1, 2, . . . , −1, −2, ... Since a well ordering of the reals cannot be written down, one might guess that there is none. This guess is shown to be false by the theorem that states that any set X has a well ordering. In proving this, one considers the collection Z of all non-null subsets of X, selects a point $x_a = f(z_a)$ from each element z_a of Z, and well orders X so that if S_a is the set of all elements that precedes x_a, then $x_a = f(Z - S_a)$.

Some of the theorems proved by well ordering are so strange that their truths do not seem intuitively obvious. Well ordering is also used to construct pathological examples which serve as counterexamples to various conjectures. These counterexamples are useful since they show that the conjectures are false and it is useless to try to prove them.

Formation of sets. One approved method of forming a set is to consider a property P possessed by certain elements of a given set X. The set of elements of X having property P may be considered as a set Y. The expression $p \in X$ is used to denote the fact that p is an element of X. Then $Y = \{p | p \in X$ and p has property $P\}$. Another approved method is to consider the set Z of all subsets of a given set X. It may be shown in this case that $C(X) < C(Z)$.

Paradoxically, it is not permissible to regard the collection of all sets as a set. If such a collection X were called a set, and Z were used to denote the set of all subsets of X, one would arrive at the absurdity that $C(X) < C(Z)$.

Operations with sets. In set theory, one is interested not only in the properties of sets but also in operations involving sets: addition, subtraction, multiplication, and mapping.

Sum or union. The sum of A and B ($A + B$ or $A \cup B$) is the set of all elements in either A or B; that is, $A + B = \{p | p \in A$ or $p \in B\}$.

Intersection, product, or common part. The intersection of A and B ($A \cdot B$, $A \cap B$, or AB) is the set of all elements in both A and B; that is, $A \cdot B = \{p | p \in A$ and $p \in B\}$. If there is no element which is both A and B, one says that A does not intersect B and writes $A \cdot B = 0$.

Difference. The expression $A - B$ is used to denote the collection of elements of A that do not belong to B; that is, $A - B = \{p | p \in A$ and $p \notin B\}$. If $A \subset B$, it is expressed as $A - B = 0$.

An example. If a person were to squirt some black ink on a plane, the set A of points in the dark spot would be an example of a point set. Suppose a set B is determined by squirting some red ink on the plane. Then $A + B$ designates the set of points covered by ink, $A \cdot B$ designates the set covered by both kinds of ink, and $A - B$ designates those covered by black but not red ink (**Fig. 1**).

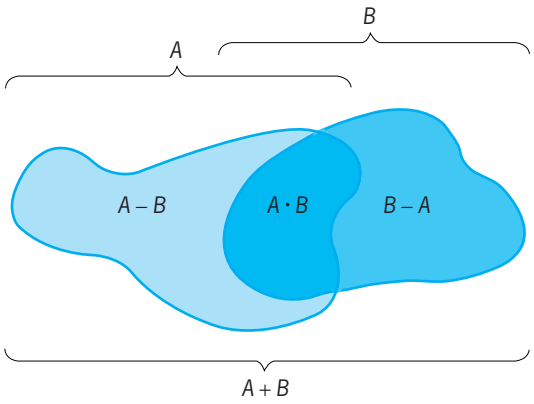


Fig. 1. Sets of points on a plane.

Boolean algebra. By using the previous notation, it follows that the sets of Fig. 1 satisfy some of the familiar laws of algebra as

$$A + B = B + A$$
$$A \cdot (B + C) = A \cdot B + A \cdot C$$

However, other identities are not so familiar:

$$X - (A + B) = (X - A) \cdot (X - B)$$
$$X - A \cdot B = (X - A) + (X - B)$$

See BOOLEAN ALGEBRA.

Transformations. A transformation of a set X into a set Y is a function that assigns a point of Y to each point of X. The transformation shown in **Fig. 2** is the vertical projection of a set X onto a segment Y. The point assigned to X under a transformation f is called the image of x and denoted by $f(x)$. Also the set of all points x sent into a particular point y of Y is called the inverse of y and denoted by $f^{-1}(y)$. The inverse of y shown in Fig. 2 is the sum of three segments. The equation $f(x) = x^2$ represents a transformation that takes each real number into its square, for example, 2 to 4, -5 to 25, and so on. Rotations, congruences, and similarities are examples of transformations from geometry. However, in general, a transformation may change both the size and shape of an object.

Topology. Topology is one of the branches of mathematics that makes extensive use of set theory. Not only does one have sets of points for consideration but also collections of interiors of spheres or neighborhoods. These neighborhoods enable one to study limit points and the continuity of transformations. *See* CONFORMAL MAPPING; RING THEORY; TOPOLOGY.

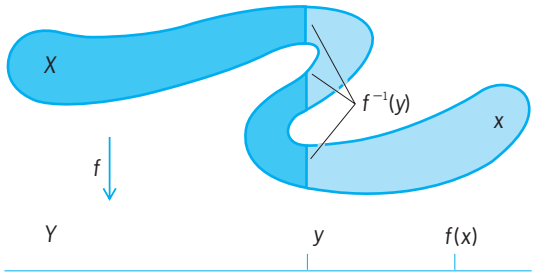


Fig. 2. Vertical projection of set *X* onto segment *Y*.

R. H. Bing

Bibliography. K. Devlin, *The Joy of Sets: Fundamentals of Contemporary Set Theory*, 2d ed., 1994; M. D. Potter, *Sets: An Introduction*, 1991; J. Roitman, *Introduction to Modern Set Theory*, 1990; R. L. Vaught, *Set Theory: An Introduction*, 2d ed., 1994.

Sewage

Water-carried wastes, in either solution or suspension, that flow away from a community. Also known as wastewater flows, sewage is the used water supply of the community. It is more than 99.9% pure water and is characterized by its volume or rate of flow, its physical condition, its chemical constituents, and the bacteriological organisms that it contains. Depending on their origin, wastewaters can be classed as sanitary, commercial, industrial, or surface runoff.

The spent water from residences and institutions, carrying body wastes, ablution water, food preparation wastes, laundry wastes, and other waste products of normal living, are classed as domestic or sanitary sewage. Liquid-carried wastes from stores and service establishments serving the immediate community, termed commercial wastes, are included in the sanitary or domestic sewage category if their characteristics are similar to household flows. Wastes that result from an industrial process or the production or manufacture of goods are classed as industrial wastes. Their flows and strengths are usually more varied, intense, and concentrated than those of sanitary sewage. Surface runoff, also known as storm flow or overland flow, is that portion of precipitation that runs rapidly over the ground surface to a defined channel. Precipitation absorbs gases and particulates from the atmosphere, dissolves and leaches materials from vegetation and soil, suspends matter from the land, washes spills and debris from urban streets and highways, and carries all these pollutants as wastes in its flow to a collection point. Discharges are classified as point-source when they emanate from a pipe outfall, or non-point-source when they are diffused and come from agriculture or unchanneled urban land drainage runoff. *See* HYDROLOGY; PRECIPITATION (METEOROLOGY).

Wastewaters from all of these sources may carry pathogenic organisms that can transmit disease to humans and other animals; contain organic matter that can cause odor and nuisance problems; hold nutrients that may cause eutrophication of receiving water bodies; and may contain hazardous or toxic materials. Proper collection and safe, nuisance-free disposal of the liquid wastes of a community are legally recognized as a necessity in an urbanized, industrialized society.

In urban areas, wastewaters are generally conveyed away through a series of buried pipes. The inlets, manholes, pipes, and other appurtenances constitute the sewerage system. A system carrying only domestic wastes is known as a sanitary sewer system; if only storm runoff is collected in the drains, it is a

storm drainage system; but if both storm water and wastewater are carried in the same sewerage system, the system is known as a combined system. In rural or suburban areas, sanitary sewage from individual buildings may be collected and disposed of on-site through the use of septic tanks and disposal fields. *See* SEPTIC TANK.

Sewage flows. The used water supply leaving a community is directly related to the quantity and timed use of water entering that community. To establish values for the engineered design of collection systems, pump stations, and treatment plants, the water consumed within a community must be measured; or it must be evaluated, taking into account climate, extent of industrialization, types of residential districts, socioeconomic status, water cost, water quality, and delivery pressure. Not all water delivered to a community ends up as flow into the sewerage system. Water is lost through public use, fire fighting, street washing, and park and home irrigation, and is consumed in commercial and industrial production and for steam generation and cooling. Per-capita water use tends to rise with increased community size (**Table 1**). *See* WATER SUPPLY ENGINEERING.

Some of the lost flow is replaced by additions from ground-water infiltration, from inflows and, at times, from private wells and industrial activities. Infiltration is ground-water leakage through poor joints, fittings, and connections into the sewerage system lying below the ground-water level. Inflow is surface water, or directed water entering the sewerage system through manhole covers, illegal roof leaders, foundation and area drains, cooling water discharges, and street washing. *See* GROUND-WATER HYDROLOGY.

A community's wastewater flow may be only 60–75% of the average water supply when new sewerage systems with tight joints and no illegal connections are installed, and infiltration/inflow is kept to a minimum. However, in time, infiltration/inflow as well as unauthorized connections increase; so engineers frequently assume the community's wastewater flow is equal to community water supply.

Human activities follow a diurnal pattern; thus fluctuations in water supply and consequently in sewage flows occur according to time of day, day of the week, weather conditions, holidays, and seasons (**Fig. 1**). Size of community, as well as the presence and type of industry, and institutions such as colleges, prisons, or hospitals all affect the quantity and fluctuation of flows. Within the same community, wastewater peaks are smaller and lag behind water supply peaks. Water supply is a pressure system, with demand anywhere in the service area being met instantaneously. Flow into the distribution network is measured at the point of supply to the total pressurized supply system. The sewerage collection system is essentially a gravity operating system whose flow is usually measured at a final collection point, the treatment plant. The wastewater, collected from all parts of the community, requires time (lag time) to flow from the user through the pipes to the treatment plant. The flow is further detained by storage within access holes and within the pipes themselves. On any given day the flow is lowest during the night, increasing to peaks at breakfast time and in the evening. In general, the smaller the community, the greater the daily variation from the average. Water use, and hence sewage flow, not only varies with time of day but also varies with time of year. During hot, dry months the flows are greater than during cold, wet periods. Infiltration/inflow varies with rainfall frequency and intensity.

For proper engineering design to safely control and contain all anticipated variations in flow, adequate capacity must be provided within the sewerage system, pumping stations, diversions, treatment plants, and outfalls. To develop practical design flow values, not only the diurnal flow variations but also peak flows must be considered. Peak flows are the maximum flows that may occur over a relatively short period of time, usually time of concentration for storm flows and up to 1 h for other wastewater flows. The total annual flow for a community, divided by 365 days, is termed the average daily flow, and this value is taken as unity. Ranges of ratios of anticipated flow fluctutations are frequently used in wastewater system design (**Table 2**).

Peaking factors, combined with infiltration/inflow values, must be analyzed for the total community, as well as for each major district within the community. The smaller and more homogeneous the study area, the more intense the peaking factor. Peaking factors are vital for proper design of pumping stations, treatment plants, and sewerage systems. Many regulatory agencies have set 100 gpd (378 lpcd) as the average daily flow for the design of sewers and plants, with peak flows into sewers at 400 gpd (1500 lpcd) for small populations, decreasing to 200 gpd (750 lpcd) for populations over 100,000.

Storm wastewater. This is that portion of precipitation that, in modern urban environments, must be safely collected and economically removed through a controlled system. Precipitation follows three paths at the Earth's surface. In any rainfall occurrence a portion of that storm's water is removed as it initially wets and cools the ground, is retained on foliage, is temporarily stored in ground depressions, is evaporated from the surface, or is transpired by vegetation. As more rain falls and overcomes this initial abstraction, a portion of the precipitation infiltrates the soil mantle and percolates downward to the ground-water table. If the rainfall exceeds the rate that can percolate into the ground, this excess water flows overland and is conveyed away as storm water runoff. *See* SURFACE WATER.

For the proper engineering design of storm water drainage systems, the volumes of storm wastewater to be handled must be determined. In the rational method, runoff is related to rainfall through the formula $Q = cIA$, where Q is the peak runoff rate from a drainage area in ft^3/s, c is a dimensionless ratio incorporating all abstractions and losses and is an indication of the portion of the rainfall that can be anticipated to actually run off from the subject drainage area, I is rainfall intensity in in./h based on the time of

TABLE 1. Average wastewater flow from various sources*

Source	Unit	Flow, gallons per unit per day† Range	Typical
Residential sources			
Apartment			
High-rise	Person	35–75	50
Low-rise	Person	50–80	65
Hotel	Guest	30–55	45
Individual residence			
Typical home	Person	45–90	70
Better home	Person	60–100	80
Luxury home	Person	75–150	95
Older home	Person	30–60	45
Summer cottage	Person	25–50	40
Motel			
With kitchen	Unit	90–180	100
Without kitchen	Unit	75–150	95
Trailer park	Person	30–50	40
Commercial sources			
Airport	Passenger	2–4	3
Automobile service station	Vehicle served	7–13	10
	Employee	9–15	12
Bar	Customer	1–5	3
	Employee	10–16	13
Departmental store	Toilet room	400–600	500
	Employee	8–12	10
Hotel	Guest	40–56	48
	Employee	7–13	10
Industrial building (sanitary waste only)	Employee	7–16	13
Laundry (self-service)	Machine	450–650	550
	Wash	45–55	50
Office	Employee	7–16	13
Restaurant	Meal	2–4	3
Shopping center	Employee	7–13	10
	Parking space	1–2	2
Institutional sources			
Hospital medical	Bed	125–240	165
	Employee	5–15	10
Hospital, mental	Bed	75–140	100
	Employee	5–15	10
Prison	Inmate	75–150	115
	Employee	5–15	10
Rest home	Resident	50–120	85
School, day			
With cafeteria, gym, and showers	Student	15–30	25
With cafeteria only	Student	10–20	15
Without cafeteria and gym	Student	5–17	11
School, boarding	Student	50–100	75
Recreational sources			
Apartment, resort	Person	50–70	60
Cabin, resort	Person	8–50	40
Cafeteria	Customer	1–3	2
	Employee	8–12	10
Campground (developed)	Person	20–40	30
Cocktail lounge	Seat	12–25	20
Coffee shop	Customer	4–8	6
	Employee	8–12	10
Country club	Member present	60–130	100
	Employee	10–15	13
Day camp (no meals)	Person	10–15	13
Dinning hall	Meal served	4–10	7
Dormitory, bunkhouse	Person	20–50	40
Hotel, resort	Person	40–60	50
Store, resort	Customer	1–4	3
	Employee	8–12	10
Swimming pool	Customer	5–12	10
	Employee	8–12	10
Theater	Seal	2–4	3
Visitor center	Visitor	4–8	5

*After Metcalf & Eddy, Inc., G. Tchobanoglous, and F. L. Burton, *Wastewater Engineering: Treatment, Disposal, and Reuse,* 3d ed., 1991.
†1 gal = 3.7854 liters.

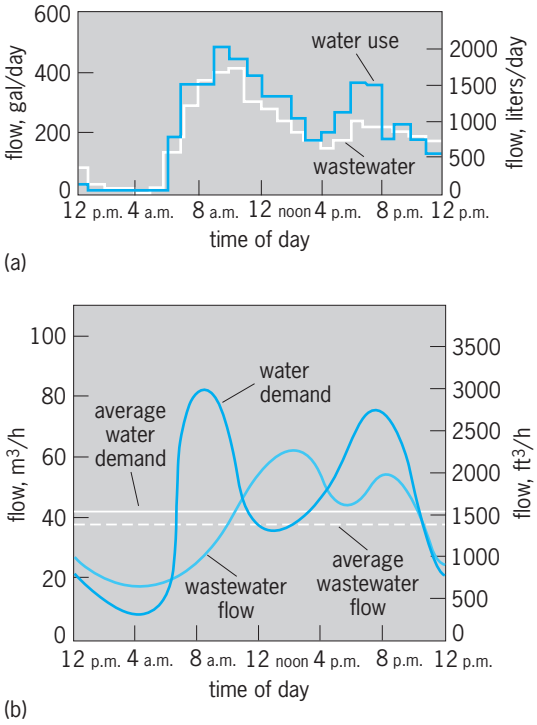


Fig. 1. Comparison of water use and wastewater flows (*a*) for a residential community (*after W. Viessman, Jr., and M. J. Hammer, Water Supply and Pollution Control, 4th ed., Harper & Row, 1985*) and (*b*) for a municipal system (*after S. R. Qasim, Wastewater Treatment Plants: Planning, Design and Operation, Holt, Rinehart and Winston, 1985*).

concentration t_c, and A is the drainage area in acres. If a rainfall with an intensity of 1 in./h falls on a drainage area of 1 acre that is totally impervious ($c = 1.0$), the peak runoff is 1 ft^3/s (28.317 liters/s). Since all the units in the U.S. Customary System work out without any conversions, this formula has been called the rational formula. The rational formula is used for urban areas of under 1 mi^2. The rational formula can be used in the International System (SI) with $Q = 2.78 \times cIA$, where Q is in liters/s, I is in mm/h, and A is in hectares. The runoff coefficient c can never be unity, because no surface is totally impervious and there are always abstractive losses. However, many regulatory agencies use a value of 1.0 for building roof areas and for concrete surfaces to simplify calculations and to incorporate a factor of safety. The wide range of runoff coefficients (**Table 3**) demands realistic evaluation of local conditions and use of engineering judgment.

Time of concentration t_c is the total time required for a drop of water to travel from the hydraulically farthest point of the tributary area to the point of collection. The term t_c is composed of two components, inlet flow time t_0 and pipe flow time t_p. Inlet (overland) flow time is the maximum time required for surface water to reach the inlet, and that time is dependent on ground slope, surface texture, and physical distance. Pipe flow time is established by dividing the length of sewer run, from inlet to point of concentration, by the velocity of the wastewater flow within the pipe. The time of concentration is equal to the sum of the two times ($t_0 + t_p$).

Rain gages at various locations in an area measure and record the quantity of precipitation collected per unit time. The rate and quantity of rainfall varies during any one storm and among the gages. Intensity–duration curves are developed by statistical analysis of rainfall data (**Fig. 2**). From accumulated gaged values, over extended time periods, a probability of frequency of occurrence can be developed. The return period is the anticipated average time interval in years between occurrences of a specified magnitude. Rainfall intensity values, correlated with the time of concentration t_c, are obtained from these intensity–duration–frequency curves. *See* PRECIPITATION MEASUREMENT.

Urban storm water volumes to be collected and carried away are evaluated on a cost-benefit basis. The greater the return period, the greater the probability of a longer time interval between flood

TABLE 2. Ranges of the ratios of flow fluctuation to average annual value

Design factor	Ratio
Maximum day in a year	(1.7–2.0) : 1
Maximum hour any day	(1.4–1.8) : 1
Peak hour in year (maximum)	(2.85–3.25) : 1
Minimum day in a year	(.50–.67) : 1
Minimum hour any day	(.35–.50) : 1
Minimum hour in year	(.20–.30) : 1
Daily average during maximum month (late summer)	(1.25–1.40) : 1
Daily average during minimum month (late winter)	(.65–.80) : 1
Daily average during maximum week (dry summer)	(1.30–1.65) : 1

TABLE 3. Recommended runoff coefficients for a specified type of area or surface*

Area or surface	Runoff coefficients†
Business	
Downtown	0.70–0.95
Neighborhood	0.50–0.70
Residential	
Single-family	0.30–0.50
Multiunits, detached	0.40–0.60
Multiunits, attached	0.60–0.75
Residential (suburban)	0.25–0.40
Apartment	0.50–0.70
Industrial	
Light	0.50–0.80
Heavy	0.60–0.90
Park, cemetery	0.10–0.25
Playground	0.20–0.35
Railroad yard	0.20–0.35
Unimproved	0.10–0.30
Pavement	
Asphalt and concrete	0.70–0.95
Brick	0.70–0.85
Roof	0.75–0.95
Lawn, sandy soil	
Flat, 2%	0.13–0.17
Average, 2–7%	0.18–0.22
Steep, 7%	0.25–0.35

* After American Iron and Steel Institute, *Modern Sewer Design*, 1980.
† The coefficients are applicable for storms of 5–10-year frequencies. Less-frequent, higher-intensity storms require the use of higher coefficients because infiltration and other losses have a proportionally smaller effect on runoff. The coefficients are based on the assumption that the design storm does not occur when the ground surface is frozen.

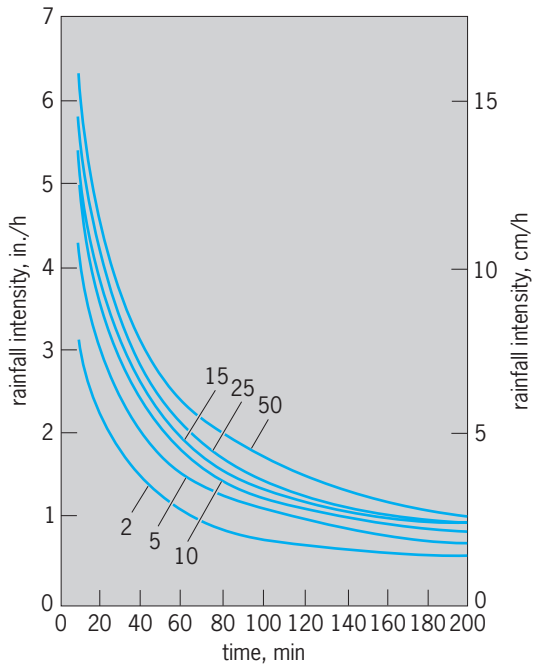


Fig. 2. Rainfall intensity-duration-frequency curves. Frequency (in years) is indicated by numbers on curves. (*After Design of Wastewater and Stormwater Pumping Stations, Manual of Practice no. FD-4, Water Pollution Control Federation, 1981*)

damage or hazardous conditions. A larger return period also means larger design flow volumes, greater required capacity of the installed system, and consequently greater construction costs. Engineering judgment must be used to balance construction cost against reduction in future inconvenience, traffic delays, and potential flooding damages. The more serious or costly the anticipated flooding, the greater the justification for the initial higher installed costs. In areas with small structural development, a 5-year return period might be acceptable, while in a highly developed area with an extensive infrastructure or a major transportation hub, a more conservative design with a 25-year, or even a 50-year, return might be reasonable.

Design period. This is the time span for which a proposed system or installed infrastructure should be able to function in a safe manner, that is, be capable of providing the desired service without structural failure. The design period is also known as project life expectancy. The U.S. Internal Revenue Service provides estimated life expectancies for many industrial components, with hydraulic items including piping among them. Many nontechnical issues such as legal or regulatory constraints, bonding limitations, politics, or other issues may influence the design periods. Sewerage systems are designed to handle the anticipated peak load that will develop under anticipated population growth within the subject area for a period of 40–50 years. Appurtenances and mechanical equipment are not expected to last that long, and 15–20-year design periods are usual.

Pumping sewage. Pumping facilities become necessary when it is uneconomical or physically impossible to convey sewage or storm water by gravity to the desired point of discharge. Pumps can be installed in deep basements of buildings to raise the wastes to the level of street sewers, or they can be used to overcome cost and construction problems of deep excavation or topographic impediments like crossing a ridge or bypassing a watercourse. Pumping facilities are sited after evaluating topography, subsurface conditions, flooding hazards, and economic feasibility.

Pumping facilities can be factory-assembled or built in place. They can be either wet-pit (**Fig. 3**) or wet-well/dry-well (**Fig. 4**) types. Wet-pit stations are generally employed for small flows up to 700 gal/min (0.044 m^3/s). They are usually factory-assembled. The wet-pit stations are automated for routine unattended operation, using pneumatic ejectors, suction lift pumps, vertical wet-pit pumps, or submersible pumps that are nonclogging or self-cleaning. Built-in-place stations, using a separate wet well to receive and store the incoming wastes, have centrifugal pumps installed in an adjacent dry well, and typically handle flows in excess of 1000 gal/min (63 m^3/s). Since a pump station is usually installed at a topographic low point, the design must incorporate flood protection and access under all weather conditions. Many states require flood protection for recurrence intervals (return periods) of 25–100 years. *See* PUMP; PUMPING MACHINERY.

Sewage pumps, even though they are fitted with open impellers designed to pass the large solids anticipated in the flow, must still be protected from clogging. Pumps are additionally protected by grit chambers, comminutors, or bar racks (screens), which must be accessible for service and maintenance. Because organic matter is present in wastewater and decomposes with time, malodorous, corrosive, toxic, and hazardous gases frequently develop in pump station wet wells. Proper ventilation, explosion-proof construction, and personnel safety procedures must be adhered to stringently. Pumping facilities are provided with two or more pumps of the size required to handle the peak anticipated inflow when one pumping unit is out of service. Pumps, with electric drive units set above flood level or otherwise protected, should have dual power sources, either from two separate and distinct electric supplies or through the use of a fuel-fed auxiliary system. Pump stations, particularly unattended ones, must be protected against freezing and should be provided with an alarm system to notify appropriate personnel in the event of any type of power interruption or high-water conditions at the facility. *See* SEWAGE TREATMENT.

Examination. Tests and analysis are employed to establish the characteristics of wastewaters such as composition; condition; and physical, chemical, and biological constituents. Water, as used within a community, undergoes only a slight delay in its passage along the hydrologic cycle. In the modern world, water separated from sewage may eventually be directed into a body of water that may be used as a source for a potable water supply. As such, knowledge about what is being returned to

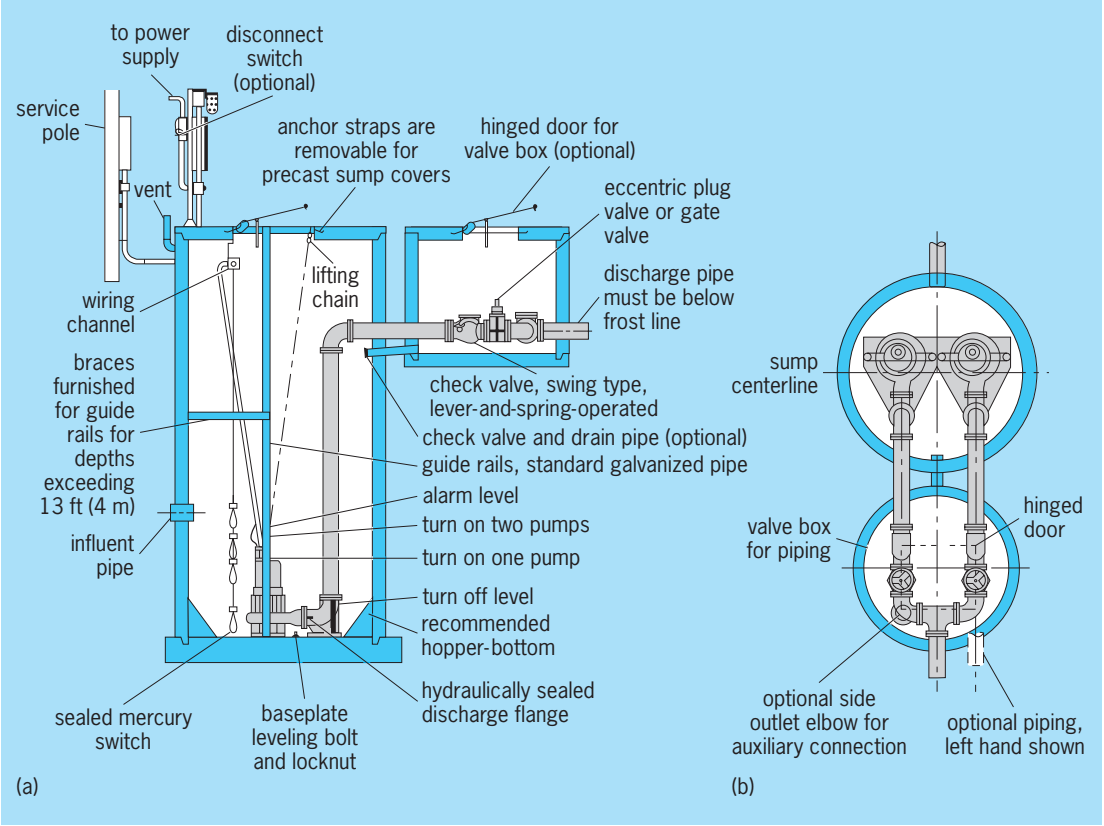


Fig. 3. Wet-pit pumping station with submersible pumps shown (*a*) in section and (*b*) in plan view. (*After Metcalf & Eddy, Inc., and G. Tchobanoglous, Wastewater Engineering: Collection and Pumping of Wastewater, McGraw-Hill, 1981*)

the watercourse and into the immediate environment is essential. Wastewaters contain a myriad of dissolved and suspended materials, including minerals, salts, organic compounds, heavy metals, bacteria, pathogenic organisms, sediments, and toxic substances. It is society's responsibility to return the spent waters to the water-use cycle with safe and acceptable constituents. Tests are used to determine treatability of the waste, amounts of solids and methods to handle these residues, economical amounts of necessary or required additives, and presence of pathogenic organisms, and to establish safe limits for constituents found in effluents that are to be returned to the environment. Technologies and mensuration methods now employed for testing wastes can establish quantities as low as 1 part per trillion (ppt). *See* ANALYTICAL CHEMISTRY; PUBLIC HEALTH.

Results of chemical laboratory tests are usually expressed in mg/liter or g/m^3. For much smaller values, micrograms/liter are used. Since wastewater is more than 99.9% water, mg/liter can be interchanged with parts per million (ppm) and μg/liter with parts per billion (ppb). For solids and sludges, where concentrations are greater than 1 kg/liter, percentages are commonly used. Flows are reported in liters/s, m^3/s, or 10^6 gal/day (mgd). In U.S. Customary units, lb/day can be evaluated from the flow rate in mgd $\times$ dose rate in ppm, or mg/liter $\times$ 8.33 lb/gal.

Since any examination is only as useful and as valid as the sample, planning and care must be exercised in selecting the sampling point, the type and frequency of collection, and the actual manner in which the sample is obtained. Sampling may be a single, random sample, known as a grab sample, or it can be a more representative sample that is made up of a repeated number of collections that are composited by either flow or time. All samples collected must be preserved so that there is no significant change in quality between time of sampling and actual testing.

Tests are characterized as physical, chemical, for solids, bioassay, or bacteriological. The physical tests made for turbidity, color, oils, suspended solids, dissolved oxygen, and temperature relate to the esthetic senses. Sewage may be categorized by its dissolved-oxygen content as fresh, stale, or septic. Fresh sewage has a dissolved oxygen content well in excess of 2 mg/liter, is grayish in color, usually is turbid containing visible solids, but has little or no odor. As organic matter undergoes bacterial decomposition, the dissolved oxygen content drops, color darkens, and a musty odor is noted. When all the dissolved oxygen is used up, sewage becomes dark, odors develop, and the waste is septic. Sewage in this last phase is esthetically unacceptable and also more difficult to treat. *See* STREAM POLLUTION; WATER POLLUTION.

Various tests are employed to determine the type and quantity of organics present. Aerobic organisms present in wastewater require dissolved-oxygen and organic matter for growth. As they multiply, they

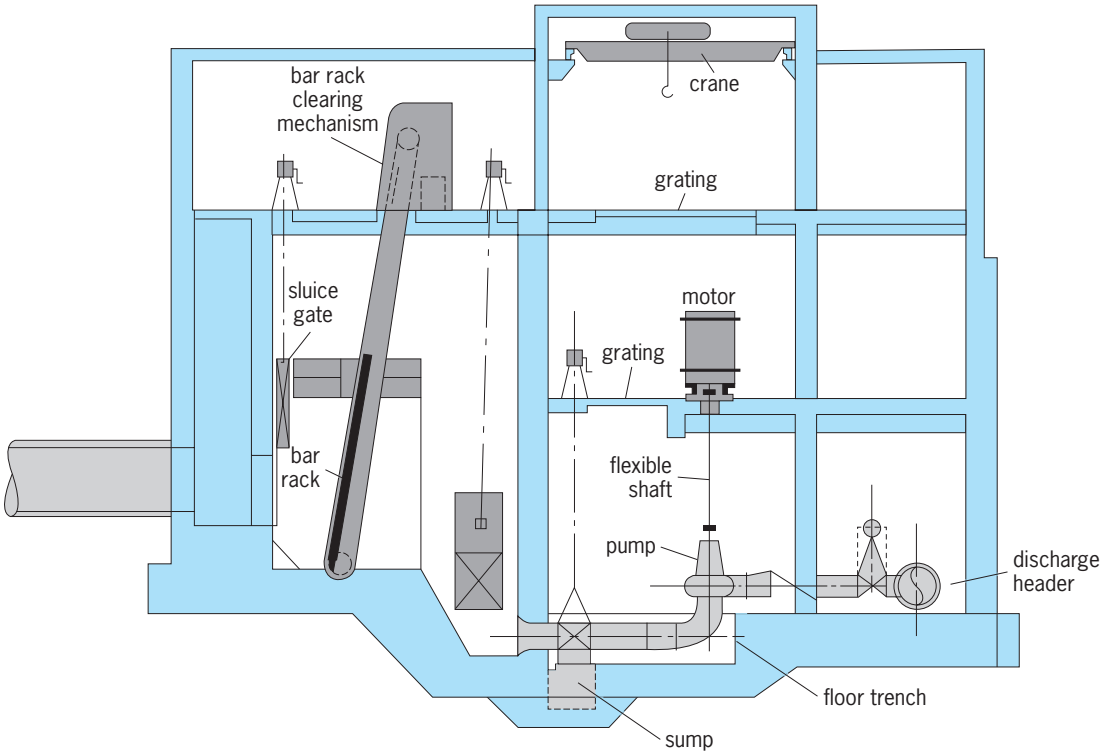


Fig. 4. Section view of conventional wet/dry-well pumping station with protective devices. (*After Metcalf & Eddy, Inc., and G. Tchobanoglous, Wastewater Engineering: Collection and Pumping of Wastewater, McGraw-Hill, 1981*)

use up the dissolved oxygen but stabilize the organic matter and reduce its putrescibility. By measuring the rate at which these bacteria deplete the dissolved-oxygen during 5 days at 68°F (20°C), the organic strength, reported as the biochemical oxygen demand, is determined. The chemical oxygen demand test uses a strong oxidizing agent instead of bacteria. This quicker method indicates that portion of the organic matter that is susceptible to chemical oxidation. Though both tests provide a measure of oxygen required by organics in wastes, they are different processes producing results that may differ widely and not provide good correlation.

In a small waste sample or where there is a low concentration of organic matter, the organics can be determined by the total organic carbon test. In this test, a known quantity of the sample is injected into a specially designed high-temperature retort. The resulting carbon dioxide produced is automatically read by an infrared analyzer. Still another test for organic content is the total oxygen demand test.

Wastewater solids provide important information about the strength, character, and the quantity of residue that will ultimately have to be disposed of. A measured quantity of a waste that is evaporated to dryness and weighed establishes the total residue. Passing another measured quantity of the same waste through a filter and weighing the trapped residue indicates the suspended solids. Subtracting the weight of the suspended solids from the total residue yields the weight of dissolved solids. Settleable solids are determined by placing a liter of the same waste in an Imhoff cone, letting it stay quiescent for an hour, and measuring what has settled out. Once dry weights of the tested waste have been determined, each sample is burned in a muffle furnace. Organic matter is volatilized, and inert matter or ash remains. The weight loss represents the organic component in each constituent portion of the solids found in that waste (**Table 4**). *See* SEWAGE SOLIDS.

Many tests are used to evaluate plants involved in pollution control operations and to define the residuals discharged in the effluent. Tests are conducted for pH, alkalinity, residual chlorine, nitrogen in its various forms, and phosphorus. Bioassay tests permit rapid, realistic interactions between minute quantities of difficult-to-detect toxic components on test organisms in their local environment. Bioassays, using fish as test organisms, establish maximum concentrations of toxics, and also provide the means of establishing required dilutions of effluents, so that they will not have significant impact on the local environment. *See* BIOASSAY; TOXICOLOGY.

The sources of most waterborne pathogenic diseases are human and animal fecal discharges. Sewage contains these discharges, but the pathogenic organisms are both few in number and difficult to test for. A human voids $2–3 \times 10^{11}$ coliform organisms daily. Since these harmless organisms are hardy, plentiful, and relatively easy to test for, and emanate from the body together with any pathogens, they are used as the index organisms for fecal pollution. Consequently when coliforms are found, it is presumed pathogenic organisms are also present. There are two different bacteriological procedures to determine the number of coliform organisms in a sample. The first method, known as the most probable number, employs the fact that coliform organisms can

TABLE 4. Typical composition of untreated domestic wastewater*

Constituent	Concentration†		
	Strong	Medium	Weak
Solids, total	1200	720	350
Dissolved, total	850	500	250
Fixed	525	300	145
Volatile	325	200	105
Suspended, total	350	220	100
Fixed	75	55	20
Volatile	275	165	80
Settleable solids, ml/liter‡	20	10	5
Biochemical oxygen demand, 5-day, 68°F (20°C)	400	220	110
Total organic carbon (TOC)	290	160	80
Chemical oxygen demand (COD)	1000	500	250
Nitrogen (total as N)	85	40	20
Organic	35	15	8
Free ammonia	50	25	12
Nitrites	0	0	0
Nitrates	0	0	0
Phosphorus (total as P)	15	8	4
Organic	5	3	1
Inorganic	10	5	3
Chlorides§	100	50	30
Alkalinity (as $CaCO_3$)§	200	100	50
Grease	150	100	50

* After Metcalf & Eddy, Inc., and G. Tchobanoglous, *Wastewater Engineering: Treatment, Disposal, Reuse*, 2d ed., McGraw-Hill, 1979.
† All values except settleable solids are expressed in mg/liter; mg/liter = 0.001 oz/ft^3.
‡ ml/liter = 11.5 in.^{3}/ft.^{3}
§ Values should be increased by amount in domestic water supply.

ferment lactose broth and produce gas. The presence or absence of gas, following 24–48 h of incubation at 95°F (35°C) in multiple fermentation tubes of broth that contain inoculations of serial dilutions of a sample, is noted. A statistical estimate of the probable density of coliforms can be established from the data. Results are recorded as the most probable number of coliforms per 100 ml of sample. *See* WATER-BORNE DISEASE.

The second procedure, the membrane filter method, is a direct enumeration method; a known volume of sample is passed through a fine-pore membrane that traps the bacteria present on its surface. The filter is then placed in a special petri dish containing requisite nutrients and incubated at 95°F (35°C) for 24 h. Each trapped coliform organism develops into a colony large enough to be identified and counted directly.

Microscopic examinations, frequently performed as a part of evaluation of plant operations, seek particular organisms within the activated sludge floc or in the zoogloea (a gelatinous mass characteristic of bacteria in fixed films). Microscopic and biologic examinations of receiving waters are made for evaluation of water quality. *See* SEWAGE COLLECTION SYSTEMS.

Gerald Palevsky

Bibliography. American Public Health Association, American Water Works Association, Water Pollution Control Association Joint Committee, *Standard Methods for the Examination of Water and Wastewater*, 20th ed., 1999; American Water Works Association, *Water Quality and Treatment*: *A Handbook of Public Water Supplies*, 5th ed., 1999; R. A. Corbitt, *Standard Handbook of Environmental Engineering*, 2d ed., 1998; M. J. Hammer, *Water and Wastewater Technology*, 5th ed., 2003; S. R. Qasim, *Wastewater Treatment Plants: Planning, Design, and Operation*, 2d ed., 1998; G. Tchobanoglous, F. L. Burton, and H. D. Stensel (eds.), *Wastewater Engineering: Treatment and Reuse*, 4th ed., 2002; W. Viessman, Jr., and M. J. Hammer, *Water Supply and Pollution Control*, 7th ed., 2004.

Sewage collection systems

Configurations of inlets, catch basins, manholes, pipes, drains, mains, holding basins, pump stations, outfalls, controls, and special devices to move wastewaters from points of collection to discharge. Sewage collection systems must be designed and constructed to function without nuisance to the public, with a minimum of maintenance, with maximum safety for workers, without damage due to corrosion or erosion, with all components sized large enough to be capable of adequately serving the community through a justifiable design period with assured durability, and with an economically acceptable cost. The system of pipes and appurtenances is also known as the sewerage system. Wastewaters may be sanitary sewage, industrial wastes, storm runoff, or combined flows.

A sewer is a constructed ditch or channel designed to carry away liquid-conveyed wastes discharged by houses and towns. Modern sewer systems typically are gravity-flow pipelines installed below the ground surface in streets and following the ground slope. The depth of cover over pipelines is controlled by

factors such as the location of rock and ground water, the ability to receive flows from all buildings by gravity, depth to frost line, economics of maintaining gravity flow as compared with pumping, and location and elevation of other existing utilities and infrastructures.

Sewerage systems are designed to carry the liquid wastes smoothly, without deposition, with a minimum of wasted hydraulic energy, and at minimum costs for excavation and construction; they should provide maximum capacity for future populations and flows. Engineered construction, controlled by availability of time, material, personnel, and finances, affects the choice and use of individual components within sewerage systems.

Components. Sewage collection systems include pipes, joints, manholes, house sewers, street inlets, catch basins, inverted siphons, backwater gates, diversions and regulators, outlets, and pumps.

Pipes. Pipes, house sewers, laterals, trunks, mains, and culverts are constructed of various materials to meet diverse design conditions. Sewerage systems are composed of commercially manufactured pipe sections joined together; in special circumstances they are constructed in place by using concrete or masonry, or the pipe sections are assembled in place from fabricated component pieces. Pipelines are usually circular, but they can also be elliptical, horseshoe, or arched shapes, boxes, or specially fabricated shapes (**Fig. 1**).

Criteria for choice of system piping include friction factors, size and availability of materials, types of joints, crushing strength, ease of installation, resistance to degradation from corrosion and erosion, and total costs of construction. Vitrified clay, concrete, cast iron, fabricated steel, and plastic are the materials most commonly used for pipes. Standards for uniformity, material quality control, size tolerances, guaranteed strength, and jointing conditions are established by standards organizations and the federal government, as well as industrial, trade, and professional organizations.

Different materials are manufactured in specified diameters, in various standard lengths, and with defined types of ends or joints. The shorter the length of pipe, the more joints are necessary. This can lead to higher labor costs, increased possibility of misalignment, leakage, chance of failure, and lack of durability. Increasing pipe length increases section weight, making pieces heavy, unwieldy, and difficult to place and join in a trench, thus requiring special equipment and, often, more personnel. The choice of pipe frequently depends upon economics, and it may be controlled by construction conditions.

Vitrified clay, generally 4–36 in. (10–91 cm) in diameter, is nonabsorptive and resistant to corrosion, erosion, and scour. It is available in lengths of 1–8 ft (0.3–2.4 m), and it varies in strength with diameter and category (**Table 1**). Clay pipes with bell-and-spigot joints formerly were used in traditional house drains and street sewer lines. *See* CLAY, COMMERCIAL.

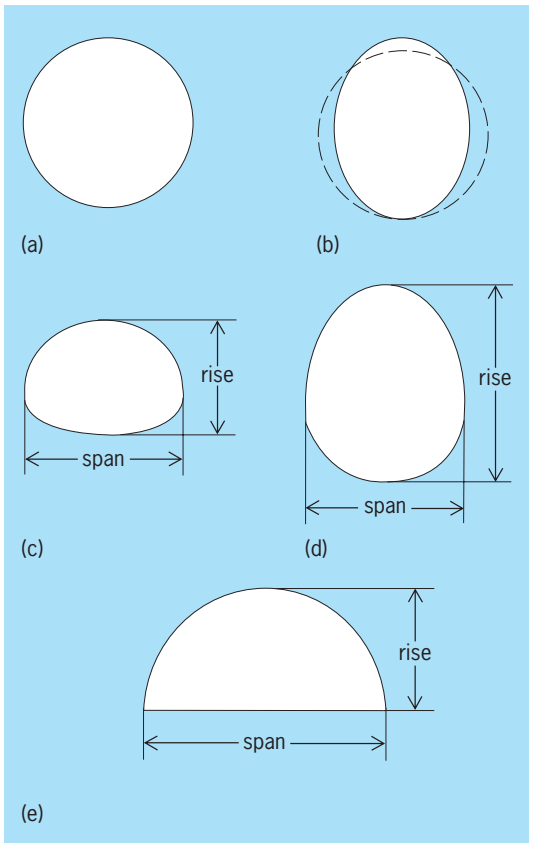


Fig. 1. Various pipe shapes. (*a*) Round. (*b*) Vertically elongated (ellipse). (c) Pipe-arch. (*d*) Underpass. (e) Arch. (*After American Iron and Steel Institute, Handbook of Steel Drainage and Highway Construction Products, 2d ed., 1971*)

Cast-iron soil pipe is a bell-and-spigot pipe of limited strength. It is used in plumbing work to carry sanitary wastes from water closets and other fixtures to the street sewer. Soil pipe is often used for house sewers. *See* CAST IRON.

Concrete pipe is manufactured in both circular and elliptical shapes that are 4 in. to 12 ft (10 cm to 3.6 m) in diameter, either plain or reinforced. Various methods, including spinning, casting, and packing, are used in its manufacture. The laying lengths and strength of concrete pipe can be varied and specifically controlled by placing different styles, configurations, and thicknesses of metal reinforcement, as well as by adjusting wall thickness (Table 1). Concrete is subject to corrosion from acids and hydrogen sulfide, and also possibly to scour or erosion where there is excessive velocity and heavy sediment load. *See* CONCRETE; REINFORCED CONCRETE.

Cast-iron pipe and ductile iron pipe 2–48 in. (5–122 cm) in diameter and with lengths to 20 ft (6 m) and various types of joints provide very tight connections, flexibility, and greater strength than concrete. These pipes, as well as those made of welded steel, are most often used for pipe crossings, in areas of high water, for special load conditions, and for pressure lines. Linings and wrappings are applied inside and out to protect these metal pipes from corrosion. *See* CORROSION.

TABLE 1. Crushing strength of clay and concrete pipe by the three-edge bearing method, lb/linear ft*

Internal diameter, in.†	Clay		Nonreinforced concrete			Reinforced concrete, ultimate strength				
	Standard	Extra strength	Class I	Class II	Class III	Class I	Class II	Class III	Class IV	Class V
3	—	2000								
4	1200	2000	1500	2000	2400					
6	1200	2000	1500	2000	2400					
8	1400	2200	1500	2000	2400					
10	1600	2400	1600	2000	2400					
12	1800	2600	1800	2250	2600	—	1500	2000	3000	3750
15	2000	2900	2000	2600	2900	—	1875	2500	3750	4690
18	2200	3300	2200	3000	3300	—	2250	3000	4500	5620
21	2400	3850	2400	3300	3850	—	2625	3500	5250	6560
24	2600	4400	2600	3600	4400	—	3000	4000	6000	7500
27	2800	4700	2800	3950	4600	—	3375	4500	6750	8440
30	3300	5000	3000	4300	4750	—	3750	5000	7500	9380
33	3600	5500	3150	4400	4875	—	4125	5500	8250	10,220
36	4000	6000	3300	4500	5000	—	4500	6000	9000	11,250
39	—	6600	—	—	—	—	4825	6500	9750	—
42	—	7000	—	—	—	—	5250	7000	10,500	13,120
48	—	—	—	—	—	—	6000	8000	12,000	15,000
54	—	—	—	—	—	—	6750	9000	13,500	16,880
60	—	—	—	—	—	6000	7500	10,000	15,000	18,750
66	—	—	—	—	—	6600	8250	11,000	16,500	20,620
72	—	—	—	—	—	7200	9000	12,000	18,000	22,500
—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—
108	—	—	—	—	—	10,800	13,500	18,000	27,000	33,750

*1lb/linear ft = 0.0146 kN/m.
†1 in. = 2.54 cm.
SOURCE: *Annual Book of ASTM Standards*, American Society for Testing and Materials, 1977.

Fabricated steel pipe is constructed of lightweight corrugated flat sheets that provide strength and flexibility; they can be fabricated into a variety of cross-sectional shapes. Circular, elliptical, partial sections, arches, and pipe arches can be either fabricated in the shop or erected in the field in sizes up to 25 ft (7.5 m). As the corrugations increase in depth for increased strength and size opening, hydraulic coefficients become lower and flow characteristics change. Corrugated pipes are frequently coated and paved to fill the corrugations in order to increase flow efficiency and reduce erosion and corrosion problems.

Plastic pipe of poly(vinyl chloride) (PVC), polypropylene, and acrylonitrile-butadiene-styrene resin (ABS), plastic truss pipe, and other novel materials are being used in sewerage systems. These pipes are very light, are resistant to corrosion, and are available in extended lengths. *See* POLYMER.

Joints. The joints or junctions between two adjacent pipes or between a pipe and an appurtenance must be watertight, possess a degree of flexibility, and have a long life expectancy. Most joints utilize elastomeric O rings that compress between two adjacent pipes to form a tight, yet somewhat flexible, seal. Mortar or grout is used to fill openings or spaces in order to prevent roots from growing into the joints or to prevent soil constituents from interacting with the sealing rings. Corrugated pipe typically uses a band coupling that compresses the O rings set in the corrugated grooves as bolts draw the band tight to the outside of the pipe (**Fig. 2**).

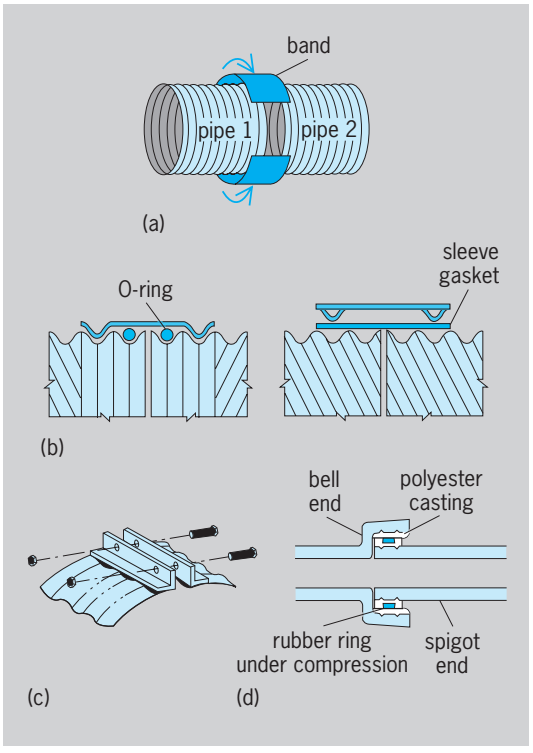


Fig. 2. Band-coupled pipe joints. (*a*) Typical band coupling; (*b*) standard corrugated steel pipe band connectors; (*c*) band angle connector (*after American Iron and Steel Institute, Modern Sewer Design, 1980*). (*d*) Details of an O-ring compression joint for clay, cast iron, and concrete pipes.

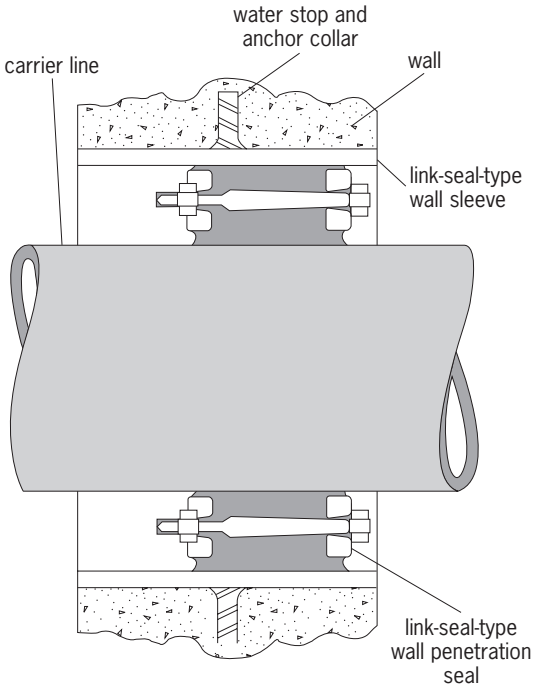


Fig. 3. Diagram of a mechanical compression water stop and seal. (*Link Seal, Thunderline Corp., Wayne, Michigan*)

Appurtenances and structures that permit connection to, entry into, or maintenance of sewerage system piping require seals at openings where pipes penetrate these structures. The openings can be grouted closed, or a mechanical compression type of closure may be used (**Fig. 3**).

Manholes. Also known as access holes, manholes are openings from the ground surface that are fitted with removable covers and permit personnel to enter, inspect, maintain, and repair the gravity flow system (**Fig. 4**). Manholes are placed at ends of lines and wherever a pipe must be changed in size, elevation, direction, or slope. On straight runs, for up to 24-in. (600-mm) diameter, they are placed no more than 350 ft (100 m) apart to permit mechanical cleaning of the intervening pipe. Spacing increases with diameter. For sewer lines 8–24 in. (20–61 cm) in diameter, manholes 4 ft (1.2 m) in diameter sit directly over the pipe, with the steps of the manhole set into the vertical barrel side. For pipes larger than 24 in. (61 cm), manholes may be offset to the side of the pipe; or larger structures known as junction chambers may be used.

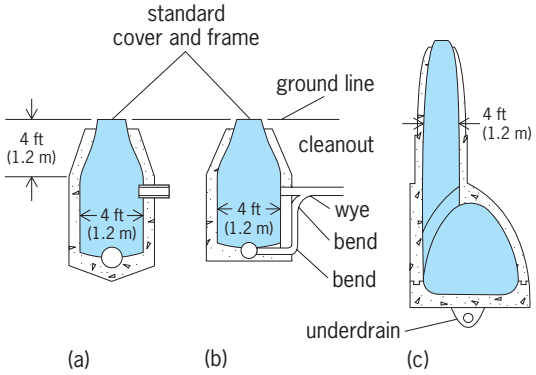


Fig. 4. Some typical sewer manholes with standard 24-in. (61-cm) cover and frame. (*a*) Inside drop. (*b*) Outside drop. (*c*) Unit for a large sewer. (*After R. L. Linsley and J. B. Franzini, Water Resources Engineering, 3d ed., McGraw-Hill, 1979*)

The channel within the bottom of the manhole should provide a smooth passage for the flow entering and leaving the structure. The paved invert in the manhole permits transitions from a smaller pipe entry to a larger pipe leaving, and it offers the means to change direction by developing a paved curve. Where the entering line is considerably above the existing pipe, a drop pipe is constructed to permit a smooth transition and yet maintain a clear straight line for ease of cleaning (**Fig. 5**).

Manhole covers, usually in roadway traffic lanes, are generally made of cast iron, set in fabricated frames, and circular in shape so that they cannot fall through the opening when they are lifted for service entry. Covers must lie flat with the pavement, seat well to prevent rattle or movement, be capable of supporting typical traffic loads without damage, and prevent storm water inflow. Covers are frequently provided with special locking devices to prevent removal, vandalism, unauthorized entry, or illegal dumping of wastes into the system.

House sewers. House or building sewers, connections, or service laterals are the small-diameter sewer lines that connect the individual residence or property with the street sewer or lateral (**Fig. 6**). House connections must be made with tight joints, and care must be taken to support the pipe adequately so that subsequent construction does not break the pipe, open joints, or develop pockets in the line. Unused stubs, the openings that have been provided in street sewers for future house sewer connections, must be tightly plugged to control infiltration. Special appurtenances such as flushing manholes, terminal cleanouts, relief overflows, lamp holes, and bypasses are installed in sewerage systems to overcome special topographic or design conditions.

Street inlets. Street inlets are structures that intercept, collect, and transfer storm runoff from street surfaces to the underground storm drain system. Street inlets are located along paved streets and are used to keep pedestrian walkways and crossways relatively free of flowing storm runoff. There are three types: gutter, curb or combined. A gutter inlet has a flat grate at the paved surface that allows flowing storm water to fall through to the inlet structure below. A curb inlet has a vertical opening in the curb and is depressed below the gutter line, allowing water to flow and drop laterally into the inlet structure. A combined inlet has both a gutter and curb opening with the two aligned over the structure. Special designs for areas of very heavy gutter flows, or where gutter flow velocities are very high, use multiple openings, deflector inlets with notches, grooves, or ridges. The outlet line from the inlet to the storm drainline manhole drains the entire structure (**Fig. 7**).

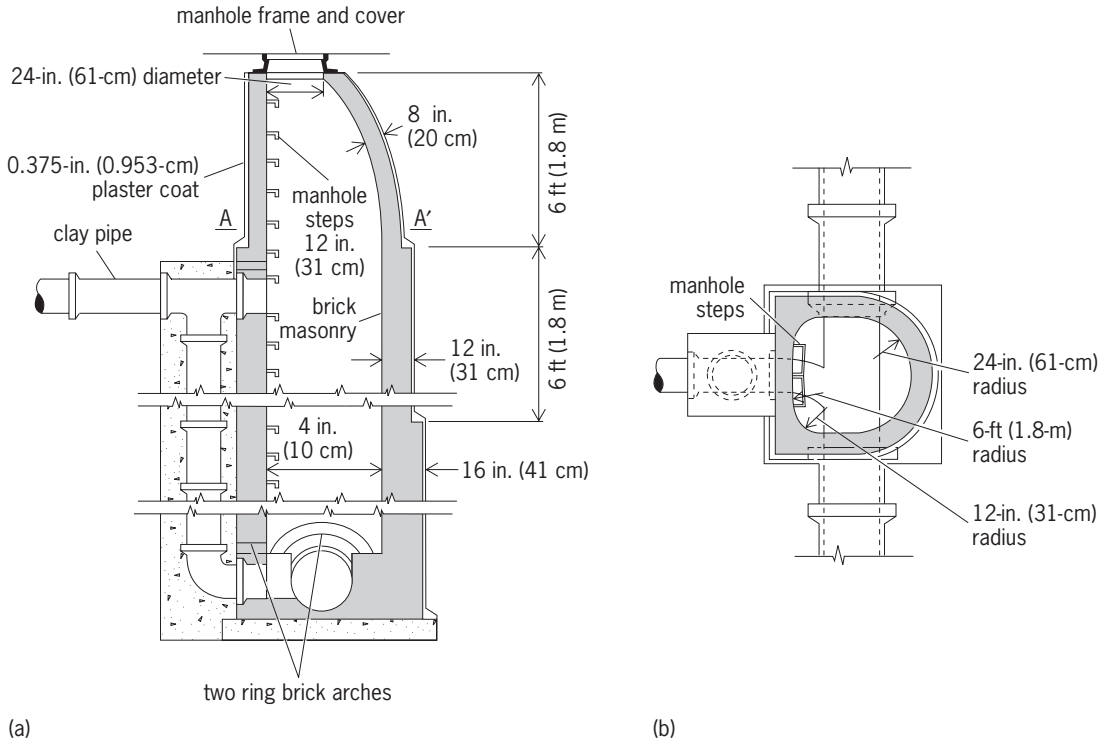


Fig. 5. Drop manhole detail. (*a*) Vertical section. (*b*) Horizontal section (A-A in part a). (*After National Clay Pipe Institute, Clay Pipe Engineering Manual, 1962*)

Catch basins. At one time, catch basins were an integral part of combined systems. They existed for storm drainage systems when streets were poorly paved or not maintained and drain lines were not constructed to maintain self-cleaning velocities. Catch basins are constructed with a sump below the outfall line to trap all grit, sand, street sweepings, and other debris within the structure. Catch basins have hatches, covers, or easily removable gutter grates to permit access for periodic maintenance and cleaning of accumulated solids. Special hoods or overflow designs are required on combined systems to prevent sewer gases from venting into the street (**Fig. 8**).

Inverted siphons. An inverted siphon is a depressed or suppressed sewer. The pipe enters the upper end chamber and usually leaves in a number of smaller lines. The multiple lines pass under the obstruction, structure, or stream with the outlet chamber being at a lower invert elevation so that inflow enters through the inlet and flows out the outlet structure. The section of pipe that is below the hydraulic grade line is always full of liquid, even when there is no flow into

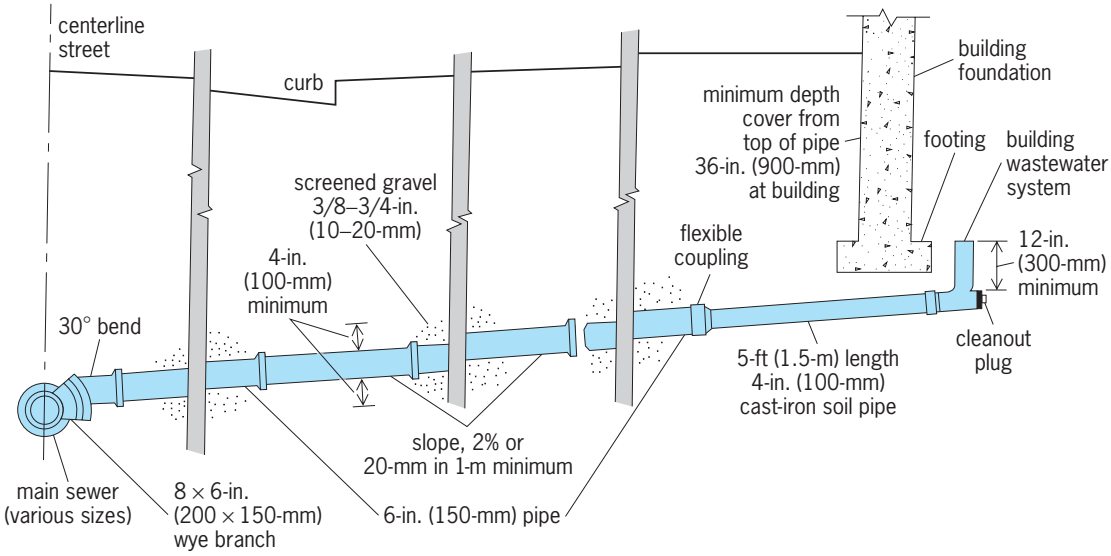


Fig. 6. Typical house sewer connection. (*After Metcalf & Eddy, Inc., and G. Tchobanoglous, Wastewater Engineering: Collection and Pumping of Wastewater, McGraw-Hill, 1981*)

the line. The multiple pipes of the inverted siphon are arranged so that as increased flow enters the inlet structure, successive lines are pressed into use, thus maintaining self-cleaning velocities above 3.5 ft/s (1.0 m/s) to prevent clogging.

Backwater gates. Backwater gates are devices that function like check valves to provide unidirectional flow and prevent the reversal of water flow in a line when the water elevation in the receiving water body rises above the pipeline's hydraulic gradient. They are also known as backflow valves, tide gates, or flap valves.

Diversions and regulators. These structures bypass, divert, or dump excess combined flow through use of weirs, of relief siphons, or by float, electrical, or pressure-operated valves and gates. Treatment plants are designed for flows up to three times the volumes of sanitary flows that occur in dry weather but they cannot cope with the huge volumes of water delivered by combined sewers during storms. Excess storm flow, above the design flow that is to continue to the treatment plant, is diverted by these structures to relief sewers, to overflows to nearby watercourses, or to holding facilities for later treatment.

Outlets. At the end of any sewerage system—be it from storm water drains, diversion or regulator outlet line, an overflow, or the end of the treatment facility—the treated or untreated flows must be discharged to the environment without causing nuisance, offense, or damage. Outlet structures vary from a single pipe retained by a straight headwall, to flared pipes, structures with special wingwalls, riprapped aprons, or specially constructed devices to dissipate hydraulic energy and prevent erosional damage as water discharges from an outlet and spills over the surface of the land. When the outlet flows are large and velocities high, special devices are used to dissipate energy. They include hydraulic jumps, hydraulic buckets, side channel discharges, and other structures designed specifically for energy dissipation.

Treatment plant outfalls usually discharge below the surface of the receiving water body. Discharges from outfall pipes are generally diffused through a series of manifolded lines fitted with multiple ports. Lines are placed deep enough to allow for good mixing and dilution, prevent so-called water boils, and maintain sufficiently low exit velocities to preclude erosion. They are anchored to prevent undermining or movement by currents or tides, and they are placed in locations where they will not interfere with normal use of the receiving body of water. *See* EROSION; HYDRAULIC JUMP; HYDRAULICS.

Pumps. Pumping facilities must be included in sewage collection systems when topography or soil conditions preclude typical piping layouts or construction, and where normal gravity flow cannot be employed.

Sewer system design. Sewer design is composed of several phases: initial, preliminary, and final. The initial conceptual phase defines the specific problem with generalized solutions, based on population, area, and financing. Conceptual design addresses broad issues of politics, public policy, probable costs, design period, and phasing of construction. An environmental assessment is necessary. In addition, evaluations should be made of bonded indebtedness, how bonds will be issued, availability of state and

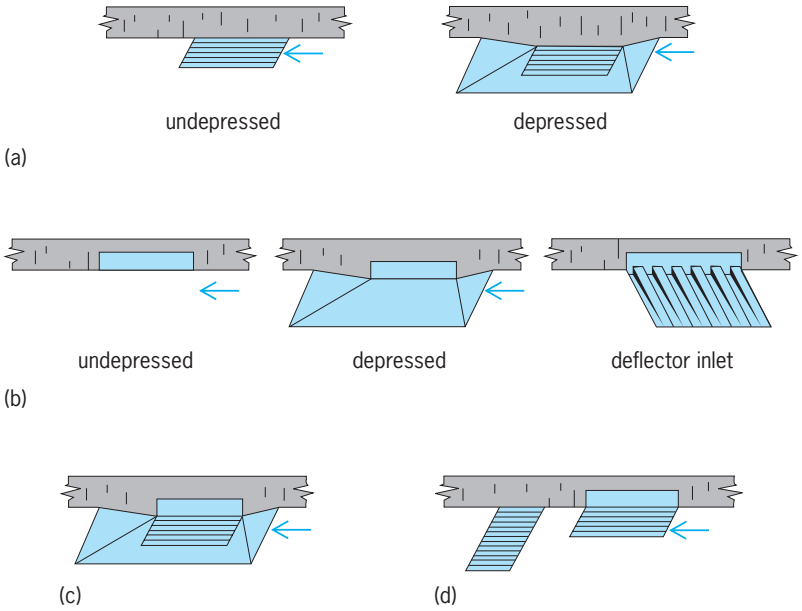


Fig. 7. Typical stormwater street inlets. (*a*) Gutter. (*b*) Curb. (*c*) Combination; grate placed directly in front of curb opening and depressed. (*d*) Multiple; undepressed. (*After Joint Committee of American Society of Civil Engineers and Water Pollution Control Federation, Design and Construction of Sanitary and Storm Sewers, ASCE MOP 37, WPCF MOP 9, 1969*)

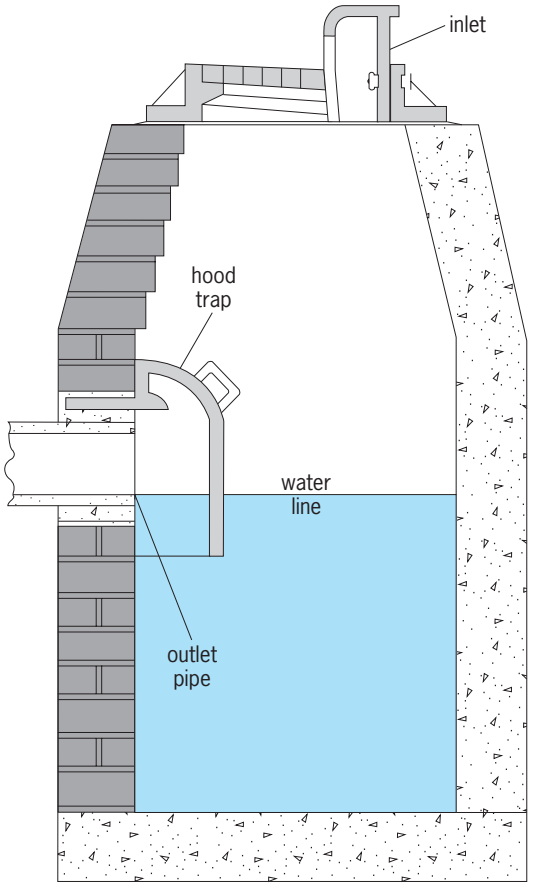


Fig. 8. Typical catch basin with removable hood trap for cleanout. (*Neenah Foundry Co., Neenah, Wisconsin*)

federal assistance programs, and preliminary determination regarding tax structures and rates which may be set by the public service commission.

Conceptual design is comprehensive and is based on physical, demographic, economic, and legal information. Projected population distribution and density, dependent upon zoning regulations and economic development within a political subdivision, must be investigated. Topography, rainfall, flood information, soil conditions, ground-water elevations, extent of existing road systems and paving, extent of existing and future infrastructure, and locations for collection, treatment, and outfall must be established. Environmental regulations, regional or multimunicipal systems, intra- and interstate compacts and legal requirements of flood control, flood plain insurance, and final disposal of liquid and solid effluents must all be resolved. *See* ENVIRONMENTAL ENGINEERING; LAND-USE PLANNING.

The preliminary design phase uses reliable data to permit development of realistic evaluations of the proposed project. The area to be served and the initial and future quantities of flow to be handled are established, and treatment and discharge points are determined. Tentative pipeline runs are prepared by using topographic maps and aerial photographs to locate major concentrations of buildings, roads, possible easements, streams, and special features. Preliminary sewer lines and grades are approximated, and appurtenances, diversions, pump stations, overflows, treatment plants, and outfalls spotted. Preliminary costs based on projected lengths of pipe, number of manholes, special structures, appurtenances, and approximate excavation are estimated. *See* AERIAL PHOTOGRAPH; TOPOGRAPHIC SURVEYING AND MAPPING.

Final design is based on refined information of the topography and soil that has been obtained by field surveys and borings. Design flows are established by detailed evaluation of areas and population projections based on demographics. Pipelines and appurtenances are located, sized, and drawn on contract plans. Specifications for construction are prepared, and cost estimates for the designed system are made. *See* SURVEYING.

Final construction plans for a sewerage system contain both plan and profile sheets (**Fig. 9**). Plan sheets provide graphical information about the location, dimensions, and clearances of all proposed pipes and appurtenances with respect to topography, legal grades, existing and proposed infrastructure, water mains, overhead power lines,

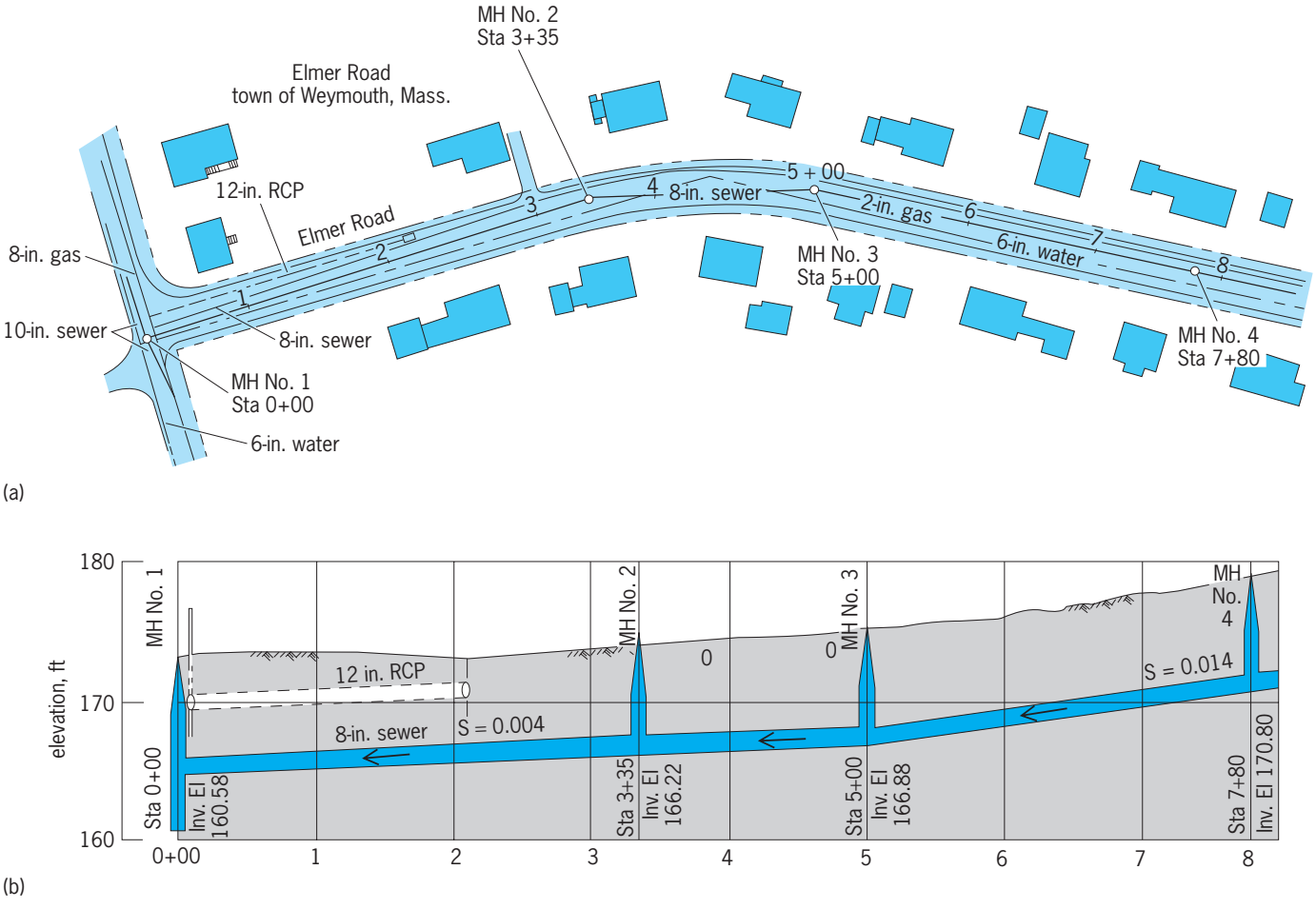


Fig. 9. Typical contract plan and profile sheet for sewer lateral. (*a*) Plan. (*b*) Profile; numbers on horizontal axis refer to sites with same numbers in part *a*. MH = manhole. Sta = Station. Inv. El. = Invert elevation. RCP = reinforced concrete pipe. S = slope in ft/ft or m/m. 1 ft = 0. 3 m. 1 in. = 2.54 cm. (*After Metcalf & Eddy, Inc., and G. Tchobanoglous, Wastewater Engineering: Collection and Pumping of Wastewater, McGraw-Hill, 1981*)

trees, and special situations. The profile sheet, a developmental plan along the centerline of the proposed sewerage system, with stationing and references to the plan sheet, clearly shows initial and final ground elevations; diameter, material, length, and slope of pipes; location of and connections to appurtenances; all control invert elevations; all known belowground infrastructure crossings; clearances and cover over proposed pipe; and special bedding conditions under the pipe.

Additional sheets showing typical details, standard castings, construction limits, trench conditions, and pipe bedding, together with designs of special structures and appurtenances for erosion and sediment control during construction, complete the contract drawings. Specifications defining the quality of material, workmanship of construction, esthetics of finished work, testing of system, and guarantees are an integral portion of the public works contract documents.

The flows established in the design of a sanitary sewer system are dependent upon population. Storm drainage systems are based on rainfall-runoff values coupled with risk evaluation of flood damage. Combined systems are more complex, since they are designed on values derived from both of the above. Combined systems must carry dry weather flows without deposition, must carry limited storm water flow without surcharging treatment facilities, and must provide for diversion and overflows when heavy storm flows develop.

Pipe flow. Liquid flows from a point of higher energy to a point of lower energy. Energy at a point is made up of position, pressure, and kinetics. Position is elevation above a measured datum; pressure is due to applied external force; and kinetic energy is developed from the velocity of the moving liquid. The difference in energy levels between an upstream and a downstream point is known as the available head.

Within the normal range of sizes used, a pipe section between manholes is a gravity sewer that is an open-channel system. The pipe section between manholes is of a constant size, placed on a continuous slope, with the liquid quantity between manholes assumed constant. Under these conditions, since there is no external pressure, uniform flow exists, and velocity is constant.

The difference in physical elevation between two points along the bottom of the pipe (difference in invert elevation), divided by the horizontal distance between the two points (reach), is the invert slope. Since quantity of flow and pipe size are constant, the depth of flow in the pipe (normal depth) is constant, and the water surface slope, or hydraulic gradient, is parallel to the invert slope. Because the liquid is moving, it develops kinetic energy that is known as the velocity head and is given by Eq. (1), where h_v is

$$h_v = \frac{V^2}{2g} \tag{1}$$

the velocity head in ft or m, V is velocity in ft/s or m/s, and g is the gravitational acceleration constant. The energy gradient is the theoretical line depicting the total energy in the fluid at every point along the pipe. Under uniform flow, the energy gradient, the hydraulic gradient, and the bottom of the pipeline are all parallel, and their slopes are identical. **Figure 10**

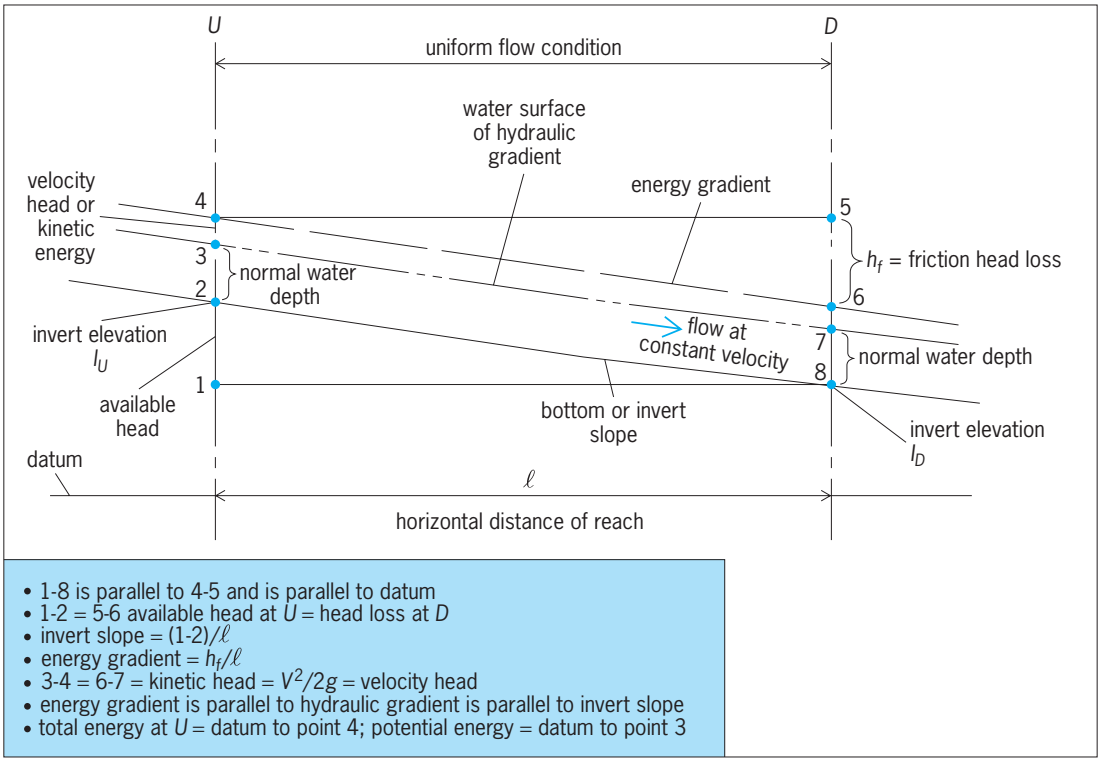


Fig. 10. Hydraulic profile for uniform flow condition, showing values at points *U* and *D*.

TABLE 2. Coefficients of roughness values of *n* for use in the Manning equation

Surface	Best	Good	Fair	Poor
Uncoated cast-iron pipe	0.012	0.013	0.014	0.015
Coated cast-iron pipe	0.011	0.012*	0.013*	
Commercial wrought-iron pipe, black	0.012	0.013	0.014	0.015
Commercial wrought-iron pipe, galvanized	0.013	0.014	0.015	0.017
Smooth brass and glass pipe	0.009	0.010	0.011	0.013
Smooth lockbar and welded O.D. pipe	0.010	0.011*	0.013*	
Riveted and spiral steel pipe	0.013	0.015*	0.017*	
Vitrified sewer pipe	{0.010 / 0.011}	0.013*	0.015	0.017
Common clay drainage tile	0.011	0.012*	0.014*	0.017
Glazed brickwork	0.011	0.012	0.013*	0.015
Brick in cement mortar; brick sewers	0.012	0.013	0.015*	0.017
Neat cement surfaces	0.010	0.011	0.012	0.013
Cement mortar surfaces	0.011	0.012	0.013*	0.015
Concrete pipe	0.012	0.013	0.015*	0.016
Wood stave pipe	0.010	0.011	0.012	0.013
Plank flumes				
Planed	0.010	0.012*	0.013	0.014
Unplaned	0.011	0.013*	0.014	0.015
With battens	0.012	0.015*	0.016	
Concrete-lined channels	0.012	0.014*	0.016*	0.018
Cement-rubble surface	0.017	0.020	0.025	0.030
Dry-rubble surface	0.025	0.030	0.033	0.035
Dressed-ashlar surface	0.013	0.014	0.015	0.017
Semicircular metal flumes, smooth	0.011	0.012	0.013	0.015
Semicircular metal flumes, corrugated	0.0225	0.025	0.0275	0.030
Canals and ditches				
Earth, straight and uniform	0.017	0.020	0.0225*	0.025
Rock cuts, smooth and uniform	0.025	0.030	0.033*	0.035
Rocks cuts, jagged and irregular	0.035	0.040	0.045	
Winding sluggish canals	0.0225	0.025*	0.0275	0.030
Dredged-earth channels	0.025	0.0275*	0.030	0.033
Canals with rough stony beds, weeds on earth banks	0.025	0.030	0.035*	0.040
Earth bottom, rubble sides	0.028	0.030*	0.033*	0.035
Natural-stream channels				
1. Clean, straight bank, full stage, no rifts or deep pools	0.025	0.0275	0.030	0.033
2. Same as 1, but some weeds and stones	0.030	0.033	0.035	0.040
3. Winding, some pools and shoals, clean	0.033	0.035	0.040	0.045
4. Same as 3, lower stages, more ineffective slope and sections	0.040	0.045	0.050	0.055
5. Same as 3, some weeds and stones	0.035	0.040	0.045	0.050
6. Same as 4, stony sections	0.045	0.050	0.055	0.060
7. Sluggish river reaches, rather weedy or with very deep pools	0.050	0.060	0.070	0.080
8. Very weedy reaches	0.075	0.100	0.125	0.150

*Values commonly used in designing.
SOURCE: Metcalf & Eddy, Inc., and G. Tchobanoglous, *Wastewater Engineering: Collection and Pumping of Wastewater,* McGraw-Hill, 1981.

shows a typical hydraulic profile. *See* FLUID-FLOW PRINCIPLES; FLUID MECHANICS; OPEN CHANNEL.

Energy is required to move a fluid from one point to another. This energy is used to overcome the resistance or friction of the liquid along the surface of the conduit carrying the flow. The rougher the surface texture, and the more joints or imperfections in alignment, the greater will be the resistance to flow. Energy for flow requires available head, which is the change in elevation, or slope of the energy gradient from manhole to manhole.

An equation used to compute the velocity of uniform flow in an open channel is known as the Manning equation. It provides the formula that is used the most for sewerage system design. The Manning equation (2) relates velocity within a pipe or open channel to slope, to a roughness coefficient that is a measure of the surface friction, and to the hydraulic radius. Here V is velocity in ft/s [Eq. (2*a*)] or m/s [Eq. (2*b*)]; n is the coefficient of roughness and is dependent upon the material and finish of the open channel (**Table 2**); R is the hydraulic radius, obtained by dividing the cross-sectional area of water by the wetted perimeter in the open-channel flow zone (R in pipes flowing full is pipe diameter ÷ 4); and S is the slope in ft/ft [Eq. (2*a*)] or m/m [Eq. (2*b*)].

$$V = \frac{1.486}{n} \times R^{2/3} \times S^{1/2} \qquad (2a)$$

$$V = \frac{1}{n} \times R^{2/3} \times S^{1/2} \qquad (2b)$$

Sewerage systems should be designed with a velocity high enough to prevent deposition of solids, but not so high that the flow causes erosion or excess thrust or movement of pipe or appurtenances (**Table 3**). The basic Manning equation is developed for pipes flowing full; design charts and nomographs are used for sizing pipes. Many regulatory agencies limit maximum flow within pipes to some proportion of full pipe capacity as a safety factor. In addition, inflow to sewerage systems is not a constant; it varies with time and season for volume, depth, and velocity. Since the hydraulic radius R is dependent on cross-sectional area and wetted perimeter, and these do not change in a constant or uniform manner, actual

TABLE 3. Minimum slope for gravity flow in sanitary sewers

Size		Slope, m/m*	
mm	in.	$n = 0.013$	$n = 0.015$
200	8	0.0033	0.0044
250	10	0.0025	0.0033
300	12	0.0019	0.0026
375	15	0.0014	0.0019
450	18	0.0011	0.0015
525	21	0.0009	0.0012
600	24	0.0008	0.0010
675	27	0.0007†	0.0009
750	30	0.0006†	0.0008†
900	36	0.0004†	0.0006†

*Based on Manning's equation with a minimum velocity of 0.6 m/s. Where practicable, steeper slopes should be used. 1 m = 3.2808 ft.
†The minimum practicable slope for construction is about 0.0008 m/m.
SOURCE: Metcalf & Eddy, Inc., and G. Tchobanoglous, *Wastewater Engineering: Collection and Pumping of Wastewater,* McGraw-Hill, 1981.

flows within the pipes must be determined for these different conditions. Various graphs for determining hydraulic elements for partial flow are available (**Fig. 11**); they are used to determine depths of flow and velocities in sections having different shapes.

Good design of a gravity sewer line attempts to keep the slope of the energy gradient as flat as possible in order to minimize excavation and prevent loss of available energy, which cannot be recovered. Proper calculations for manhole losses must be considered, since wherever pipes change slope, direction, or size within a manhole energy losses occur. A general rule of sewerage system construction has been that the soffit (inner top) of the outflow line from a manhole should be at least 0.1 ft (3 cm) lower than the soffits of all lines entering the manhole. This is done to prevent the energy gradient from sloping upward or against the direction of flow, a condition that produces deposition of solids.

When pipe slope is significantly flatter than the overlying ground, insufficient cover over pipes may develop. This is overcome by the use of drop manholes (Figs. 4 and 5). A localized energy loss is sustained as the energy gradient is made rapidly discontinuous. Special designs for inside and outside drop manholes may become necessary to absorb excess turbulence and energy, direct the flows, and prevent scour or erosion from the falling water. Where the pipe is installed on steep grades, with steep slopes and high velocities, special collars, anchorages, thrust blocks, and joint restraints are designed and provided to prevent movement or undermining of the pipe. *See* PIPE FLOW; PIPELINE.

Pipe installation. Considerations in installing pipe include trenches, bedding, maintenance, and safety.

Trenches. Sewerage system pipelines must be installed accurately, to proper line and grade. Construction methods for normal cut and cover, or open-trench excavation, are similar to other types of excavation for general construction. Since sewerage lines are most frequently installed under pavements within the rights of way for roadways, adjacent to existing buildings, with overhead and underground utilities, special care must be exercised to protect work and workers from traffic, cave-ins, flooding, and electrical shock. Professional construction practices and regulations of the Occupational Health and Safety Administration (OSHA) must be followed for shoring, bracing, and dewatering excavations. Spoil, the excavated material from the trench, should be placed on the uphill side to prevent surface water from flowing into the trench; yet it should be placed far enough away from the trench so that it does not add weight and cause a cave-in. If the spoil is of satisfactory material, it may be used for backfilling the trench after the pipe is bedded in place. Trench widths are kept to a minimum to reduce transmitted load to the pipe. If the trench is too narrow, workers cannot work within it, make proper joints, and backfill around the pipe safely. Sheeting, shoring, and bracing, used to maintain stability of the trench excavation and prevent cave-ins, are removed after the pipe is in place and backfilling is under way. Since piping is installed on dry, stable trench bases, dewatering is necessary where ground-water levels are high. Removed water must be disposed of in a safe manner. Pipe must be properly joined with watertight joints, so that a smooth, continuous invert, on

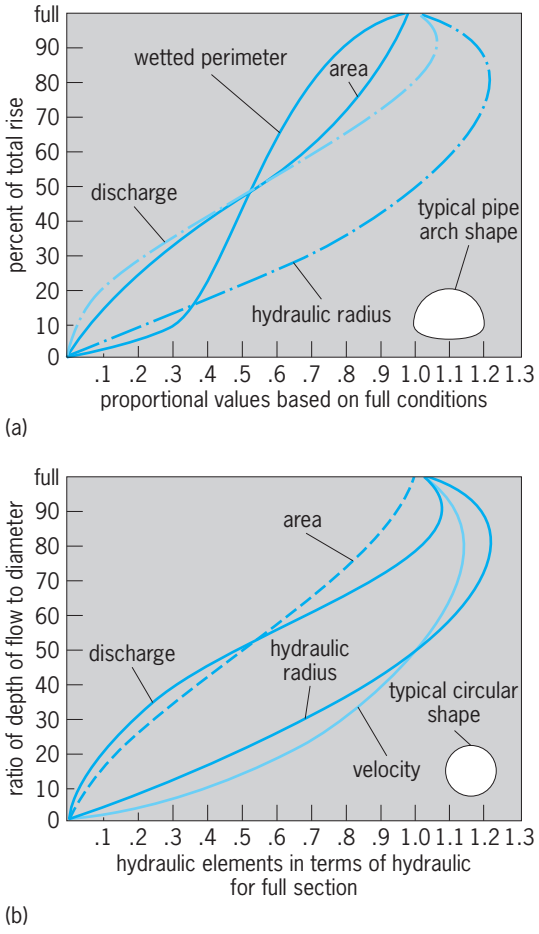


Fig. 11. Hydraulic element graphs for conditions of partial flow in terms of full section. (*a*) Pipe-arch section. (*b*) Circular section. (*After American Iron and Steel Institute, Modern Sewer Design, 1980*)

TABLE 4. Values of load coefficients C_d for trench condition

Fill material:	Sand and gravel	Saturated topsoil	Clay	Saturated clay
Specific weight, lb/ft^3 (kN/m^3):	100 (15.7)	100 (15.7)	120 (18. 9)	130 (20.4)
Ratio of cover depth to trench width				
1.0	0.84	0.86	0.88	0.90
2.0	1.45	1.50	1.55	1.62
3.0	1.90	2.00	2.10	2.20
4.0	2.22	2.33	2.49	2.65
5.0	2.45	2.60	2.80	3.03
6.0	2.60	2.78	3.04	3.33
7.0	2.75	2.95	3.23	3.57
8.0	2.80	3.03	3.37	3.76
9.0	2.88	3.11	3.48	3.92
10.0	2.92	3.17	3.56	4.04
12.0	2.97	3.24	3.68	4.22
14.0	3.00	3.28	3.75	4.34

SOURCE: R. K. Linsley and J. B. Franzini, *Water Resources Engineering,* 2d ed., McGraw-Hill, 1979.

the proper slope, with no pockets, depressions, or offsets is developed. Stubs for house sewer connections and manhole openings must be made at proper locations and elevations. Stubs or openings must be sealed to prevent infiltration. Pipe must be properly backfilled to finished grade; the excavation must be paved to provide a safe surface for carrying traffic.

Pipe tunneling and jacking are other construction techniques used for pipe installations. These are generally used to pass under highways, existing embankments, railroads, and streams, or for other special situations. New sewer lines are tested for watertightness, generally with low air pressure, to establish conformance with infiltration and exfiltration limits.

Bedding. Installed sewer lines must be able to withstand the external loads superimposed on them due to dead load from soil backfill and overburden, and from live loads of traffic and impact. The pipe's ability to withstand these loads is based on the crushing strength of the pipe material used, the method of installation, and the type and style of bedding provided. The standard for the crushing strength to which the pipe is manufactured is defined by the laboratory three-edge bearing test (Table 1).

The method of installation is defined as either trench or embankment condition. In the trench condition, pipe is installed in a relatively narrow trench, excavated into firm soil, with trench walls extending above the top of the pipe. Backfill is then placed around and over the pipe and carried up to the final ground surface. An embankment condition occurs when pipe is essentially placed on the existing ground surface or in a very wide trench, with backfill placed over the pipe and carried up to the final ground condition. Bedding, used to reduce stress concentrations and increase pipe support, is the prepared foundation between the bottom of the pipe and the firm trench or support ground surface. The safe support strength or load carrying ability of the pipe is increased with better bedding conditions.

The Marston formula [Eq. (3) for trench condition and Eq. (4) for embankment condition] provides a

$$W_d = C_d w B_d^2 \tag{3}$$

$$W_c = C_c w B_c^2 \tag{4}$$

value for trench load on a buried pipe. Here W_d or W_c is the vertical dead load of fill in lb/linear ft or kN/m on the buried pipe, w is the unit weight of fill material in lb/ft^3 or kN/m^3, B_d is the width of the trench in ft or m, measured 1 ft (0.3 m) above the top of the buried pipe, B_c is the outside pipe diameter in ft or m, and C_d or C_c are load coefficients for pipe (**Tables 4** and **5**). As the width of the trench increases, a point is reached where there is no longer a trench and the pipe is in the embankment condition. In this case the smaller value obtained for W_d or W_c controls. In addition to dead loads, vehicular and impact loads must be considered. As the live load is applied to the ground above the pipe, the load is distributed; its intensity decreases with increased depth of cover. Depth of cover is a major concern in sewerage systems to maintain pipeline integrity.

TABLE 5. Values of load coefficients C_c for embankment condition*

Ratio of cover depth to pipe diameter	C_c
1.0	1.2
2.0	2.8
3.0	4.7
4.0	6.7
6.0	11.0
8.0	16.0

* Rigid pipe, unyielding base, noncohesive backfill.
SOURCE: R. K. Linsley and J. B. Franzini, *Water Resources Engineering,* 2d ed., McGraw-Hill, 1979.

The sum of all loads, multiplied by a safety factor usually taken as 1.5, divided by the rated three-edge crushing strength of the selected pipe, establishes the load factor or type of bedding condition required (Table 1). The combination of pipe material, crushing strength, and bedding is an economic consideration in the design of sewerage systems. *See* LOADS, DYNAMIC; LOADS, TRANSVERSE; STRENGTH OF MATERIALS.

Maintenance. Clogging from accumulations of debris, tree roots, or illegally dumped materials that prevents proper flow of liquid through appurtenances and pipes is the major problem encountered in sewerage systems. Cleaning equipment to remove these obstructions includes cutting or boring tools to cut or rip through roots or balled-up debris and special tools known as pigs to move deposits of sand, grit, or other materials mechanically and hydraulically from the line to a manhole, where orange-peel buckets lift out the deposited materials.

In modern systems, breaks, cracks, spalling of pipe, poor joints, or other defects are inspected through the use of underwater television cameras. Specialized methods of concreting, patching, or even inserting a liner within the pipe have been developed to reduce infiltration, maintain the usefulness of pipes, and reduce the amount of surface excavation needed to make repairs. *See* UNDERWATER TELEVISION.

Safety. A sewerage system is a confined space where organic matter is undergoing organic decomposition. Hazardous and toxic gases generally exist; in addition, because these areas are belowground, there may be a lack of oxygen. Before personnel can safely enter a manhole, vault, or chamber, the space must be ventilated mechanically for a period of time and then tested for gases.

Entry into the sewerage system is through manholes that have narrow entry hatches or manholes with steps cast into the wall. Care must be exercised that at least two people work together, that the entering person wears a safety harness, and that the steps are safe. As a minimum, OSHA recommendations and regulations should be followed. *See* INDUSTRIAL HEALTH AND SAFETY; SEWAGE. Gerald Palevsky

Bibliography. American Concrete Pipe Association, *Concrete Pipe Design Manual*, 16th ed., 2004; American Iron and Steel Institute, *Modern Sewer Design*, 4th ed., 1999; M. J. Hammer, *Water and Wastewater Technology*, 5th ed., 2003; R. K. Linsley and J. B. Franzini, *Water Resources Engineering*, 4th ed., 1991; G. Tchobanoglous, F. L. Burton, and H. D. Stensel (eds.), *Wastewater Engineering: Treatment and Reuse*, 4th ed., 2002.

Sewage solids

The accumulated, semiliquid material consisting of suspended, colloidal, and dissolved organic and inorganic matter separated from wastewater during treatment. Sludges are developed as contained pollutants, and contaminants are separated by mechanical, hydraulic, biological, or chemical processes. The various classes of solids that are removed and collected must be disposed of in a safe, nuisance-free manner without hazard to health or the environment. Collection, handling, transporting, and disposal of removed solids are difficult and costly, since they are offensive and putrescible, with 92–99.5% water content. Sewage solids must be treated by thickening, chemical conditioning, mechanical dewatering, thermal action, biological stabilization, or digestion to convert putrescible organic matter to relatively inert end products, remove water, and reduce weight and volume.

Types. Sewage solids are classified as screenings, scum, grit, septage, or sewage sludges (see **table**).

Screenings. Large solids, carried by incoming wastewater, are captured mechanically on screens or racks with openings of various sizes. These protective units remove floating debris, including wood, clothing, cans, rags, paper, rubber and plastic goods, and stringy material that could damage equipment or create problems in plant maintenance and operation. Fine screens, or microstrainers, may be employed at plant outfalls to prevent discharge of esthetically objectionable floating solid matter. Modern, self-cleaning, inclined, rotary-drum or centrifugal screens are being installed to augment or replace grit chambers and primary settling tanks.

Screening volume varies with type of waste flow, season, weather, degree of urban development, and screen opening. Increased use of mechanical cutters and shredding devices has reduced volume of removed screenings. Coarse screenings range from 1.0 to 4.0 $ft^3/10^6$ gal (7.5 to 30 $m^3/10^6$ m^3).

Scum. This is defined as the floating fraction of sewage solids, with specific gravity under 1.0, that, under quiescent conditions, rises to the surface of the wastewater. Primary tank skimmings contain oils, fats, soaps, rubber and plastic hygienic products, cigarette filter tips, paper, and similar materials.

Grit. Heavy suspended solids consisting of sand, cinders, coffee grounds, seeds, small metal objects, and other generally inorganic particles carried in wastewater inflow are collectively known as grit. The amount of grit varies with type of sewer, season, weather, intensity of runoff, condition of streets and sewers, and use of household garbage disposal units. In new, separate sanitary sewers, with proper street sweeping, grit volumes in sanitary sewage are generally less than 4 $ft^3/10^6$ gal (30 $m^3/10^6$ m^3).

Septage. This consists of partially digested material pumped from on-site sanitary wastewater disposal systems. It contains a mixture of grit, scum, and suspended solids, adding to treatment plant sludge. *See* LEACHING; SEPTIC TANK.

Sewage sludge. Sedimentation, subsidence, or clarification of suspended matter from turbid water by gravity occurs when the velocity of that liquid has been reduced below its transport capacity. Transport capacity is controlled by varying flow-through velocity, which causes differential hydraulic subsidence. Thus heavy inorganic solids, grit, can be settled

Sources of solids and sludge from a conventional wastewater treatment facility*

Unit operation or process	Type of solids or sludge	Remarks
Screening	Coarse solids	Coarse solids are often comminuted and returned to the wastewater for removal in subsequent treatment facilities.
Grit removal	Grit and scum	Scum-removal facilities are often omitted in grit-removal facilities.
Preaeration	Scum	In some plants, scum-removal facilities are not provided in preaeration tanks.
Primary sedimentation	Primary sludge and scum	The quantities of both sludge and scum depend on the nature of the collection system and whether industrial wastes are discharged to the system.
Aeration tank	Suspended solids	Suspended solids are produced from the conversion of biological oxygen demand (BOD). If wasting is from the aeration tank, flotation thickening is normally used to thicken the waste-activated sludge.
Secondary sedimentation	Secondary sludge and scum	Provision for scum removal on secondary settling tanks is now a requirement of the U.S. Environmental Protection Agency.
Sludge-processing facilities	Sludge and ashes	The characteristics and moisture content of the sludge and ashes depend on the operations and processes that are used.

*After Metcalf & Eddy, Inc., G. Tchobanoglous, and F. Burton, *Wastewater Engineering: Treatment, Disposal, and Reuse,* 3d ed., McGraw-Hill, 1990.

in one chamber while light agglomerated organic particles, sludges, settle in others.

Sludge derives its name from the unit process from which it settles out. Primary sludge, or raw sludge, develops as solids in incoming wastewater settle hydraulically. Raw sludge, containing up to 5% solids by weight, is gray, greasy, viscous, unsightly, contains visible fecal solids and scraps of household wastes, and has a disagreeable odor. Use of household garbage grinders can increase primary sludge weight and volumes by 25–40%.

Chemical precipitation may be used to increase the removal of primary solids, to adjust pH for further treatment or to remove nutrients as phosphorus. Chemical dosage is high, producing large volumes of sludge. *See* PRECIPITATION (CHEMISTRY).

Secondary treatment uses aerobic biota to enmesh, coagulate, oxidize, nitrify, and mineralize the unsettled colloidal and dissolved organic matter that remains after primary treatment. Aerobic processes such as trickling filters, rotating biologic disks, or activated sludge utilize microorganisms to reduce organic and nutrient matter, and form sludges of settled biologic floc, which are light tan to dark brown in color, spongy, generally uniform in texture and form, having an earthy odor with solids content up to 1.5% by weight. Anaerobic decomposition will cause these sludges to turn septic rapidly, become dark to black and develop malodorous gases. *See* WATER TREATMENT.

Sludge thickening. This process removes water, increases the concentration of solids, reduces weight and volume, and prepares sludges for further treatment and handling. Flotation uses fine air bubbles to entrain and carry solids to the surface for skimming. Gravity thickening allows self-compaction of accumulated sludge, while gentle stirring causes water to separate. Use of a centrifuge compresses sludge against a porous surface that retains solids but allows liquid to pass through. Supernatant, the liquid drawn off from a sludge, is turbid, contains fine solid particles, has a high organic content, is odorous, and is returned to the inflow for retreatment.

Stabilization. When microorganisms multiply rapidly, anoxically, on organic and volatile substances contained within sludges, putrefaction and offensive odors develop. Sludges are stabilized biologically, chemically, or thermally to make them less putrescible and odorous, reduce pathogenic content, reduce weight and volume, and prepare them for final disposal. When stabilized biologically, either aerobic or anaerobic microorganisms reduce organics, volatile matter, and nutrients by converting these to gases, water, and cell tissue. Use of chemicals kills microorganisms, causing biota cells to shrivel and coagulate, releasing and separating water. Thermal processes sterilize (pasteurize) the biota, causing cell destruction, coagulation, and release of water. Chemical and thermal processes generally do not change organic or nutrient levels.

Aerobic stabilization. This allows aerobic microorganisms, during prolonged periods of aeration, to oxidize organic matter within sludge to carbon dioxide, water, ammonia, and nitrates. A humuslike, relatively odor-free and biologically stable mass that has some useful soil conditioning values is developed in some 20 days of controlled operation. The aerobic stabilization process is relatively easy to operate, and it produces a fairly clear supernatant liquor. Drawbacks are high power costs, difficulty with final settling, dewatering, and drying of the sludge, and absence of produced methane gas as a useful by-product.

Anaerobic digestion. Digestion, the most common method of handling sludges in municipal plants, occurs in closed tanks. Anaerobic digestion comprises a series of complex, interrelated biochemical reactions, performed by groups of microorganisms functioning in the absence of free oxygen. Under controlled temperature and pH, the anaerobic microorganisms produce methane gas as they decompose the sludge to a drainable, relatively inert, humic mass.

The digestion process can be envisioned as occurring in discrete, independent steps. In the first stage, acid fermentation, pH drops rapidly as colloidal solids, lipids, dissolved carbonaceous matter,

and sugars are rapidly converted to organic acids, with evolution of carbon dioxide and hydrogen sulfide. In the second stage, acid regression, pH rises slowly as organic acids formed earlier, and some proteins, are digested to acetate and ammonia compounds. In the third stage, alkaline fermentation or alkaline digestion, proteins, organic acids, and amino compounds are attacked. The pH rises to around neutral (pH 7), with large volumes of gases, predominantly methane, produced. In a properly functioning anaerobic digester the raw sludge is changed from a putrescible mass with over 70% volatile solids and high water content to a more mineralized, easily draining and drying humic mass having an inoffensive tarry odor with less than 50% volatile solids.

The sizing of digestion tanks is a function of raw sludge volumes added, digestion and water separation rates, and the volume of digested sludge drawoff. Sludge digestion occurs at temperatures from 42 to 140°F (5.6 to 60°C), with increased temperatures increasing bacterial action and reducing digestion time. Digestion time is expressed by the days necessary to attain 90% of ultimate gas production (see **illus.**). Thickening sludge prior to its being fed into the digester reduces input sludge volumes, assists in maintaining tank operation at optimum digestion temperatures, and saves space and energy.

The presence of methane gas, as it forms and leaves the mixed sludge liquor, indicates organic stabilization. Gas production at optimum mesophilic temperatures is 12–18 ft^3/lb (0.75–1.12 m^3/kg) of volatile matter destroyed. Digester gas, 60–70% methane, has a heating value of 600–700 Btu/ft^3 (22–26 kJ/m^3). The digester gas is generally used at treatment plants as an economical fuel for powering engines and heating digester tanks. *See* METHANE.

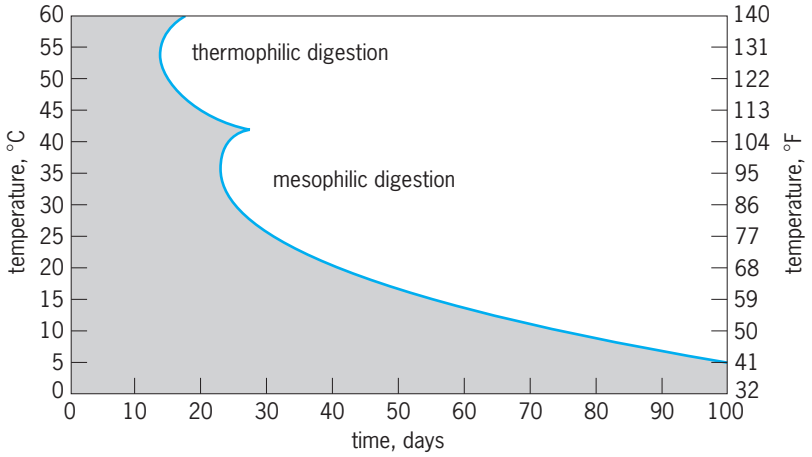
Chemical stabilization. This method uses large dosages of chlorine or lime to render sludges unsuitable for microorganisms to metabolize the organic substances, to grow or survive. The chemical stabilization process generates large volumes of sludge, often with problems of ultimate disposal.

Thermal processes. These use a variety of heating methods and sludge handling to destroy cell structure and reduce the quantity of moisture. The end solid is relatively dry, sterilized, generally free of offensive odors, and is frequently used as a soil amendment.

Disposal. Materials removed from the wastewater must be disposed of by ultimately returning them to the land, ocean, or atmosphere without damaging the receiving environment. Regulations have prohibited ocean dumping, and strict air emission rules have reduced the use of thermal systems. Solids are generally disposed of in landfills, buried, composted, or recycled as soil amendments.

Screenings, after shredding, may be returned to inflow for retreatment or are buried or incinerated. Washed grit is buried or used as fill or cover within landfills.

Digested liquid sludges may, in isolated areas, be spread or plowed under fields, but this is not a generally accepted practice. Wastewater solids may contain pathogenic organisms, but more importantly, higher concentrations of initial pollutants than were in the original wastewater. Disposal of removed solids must neither adversely affect the food chain nor endanger the environment. *See* FOOD WEB.



Time required for 90% digestion of sewage sludge at different temperatures, (*After K. Imhoff and G.M. Fair, Sewage Treatment, John Wiley & Sons, 1940, reprint 1946*)

Digested sludge as withdrawn is semiliquid; therefore, for ease of handling, storing, transporting, or landfilling and for increased efficiency in further thermal processing, it must be dewatered or dried. Where adequate area exists, open drying beds can be employed to reduce water content; otherwise centrifuges, belt presses, vacuum filters, and thermal or other processes are used. When added to soil, the dried humuslike digested sludge increases the water-holding capacity of the soil, improves texture and tilth, reduces erosion, and provides some nutrient benefit. Digested sludge, an incomplete fertilizer, requires additions of nitrogen, phosphorus, potassium, and even trace elements.

Composting, a natural thermophilic biological process, further reduces pathogens and volatile solids in digested sludge, forming a stable humuslike soil-conditioning material that can be recycled.

Heat drying uses external energy and reduced pressure to reduce moisture content to under 12%. Since drying by heat pasteurizes (sterilizes) the sludge and stops further bacterial action, dried digested sludges can easily be processed mechanically, stored, handled, and applied to land.

Thermal processes, including wet oxidation, pyrolysis, and incineration, require sludges with high content of organic solids that can produce significant heat value. Prestabilization of the sludge with reduction in volatile sludge content is not usually employed. In wet oxidation sludge is mixed with compressed air under high heat to oxidize organic matter. Pyrolysis, a thermal cracking process, converts sludge in reduced oxygen atmospheres to gases, water, and char. In both wet oxidation and pyrolysis, withdrawn supernatant liquors are characterized by high biochemical oxygen demand and solids and are retreated. Incineration is complete combustion and, as practiced in large plants, provides a reduction in volume. Environmental considerations require that

air pollution standards be met. Residual ash is usually disposed of in landfills. *See* AIR POLLUTION; HAZARDOUS WASTE; PYROLYSIS; SEWAGE; SEWAGE TREATMENT.

Gerald Palevsky

Bibliography. Environmental Protection Agency, *Sludge Treatment and Disposal*, Process Des. Man. EPA 625/1-79-011, Center for Environmental Research Information, Cincinnati, 1979; A. F. Gaudy, Jr., and E. T. Gaudy, *Microbiology for Environmental Scientists and Engineers*, 1980; J. G. Henry and G. W. Heinke, *Environmental Science and Engineering*, 2d ed., 1996; Metcalf & Eddy, Inc., G. Tchobanoglous, and F. Burton, *Wastewater Engineering: Treatment, Disposal, and Reuse*, 3d ed., 1990; H. S. Peavy, D. R. Rowe and G. Tchobanoglous, *Environmental Engineering*, 1985; T. D. Reynolds, *Unit Operations and Processes in Environmental Engineering*, 2d ed., 1995; P. A. Vesilind, *Treatment and Disposal of Wastewater Sludges*, rev. ed., 1979.

Sewage treatment

Unit processes used to separate, modify, remove, and destroy objectionable, hazardous, and pathogenic substances carried by wastewater in solution or suspension in order to render the water fit and safe for intended uses. Treatment removes unwanted constituents without affecting or altering the water molecules themselves, so that wastewater containing contaminants can be converted to safe drinking water. Stringent water quality and effluent standards have been developed that require reduction of suspended solids (turbidity), biochemical oxygen demand (related to degradable organics), and coliform organisms (indicators of fecal pollution); control of pH as well as the concentration of certain organic chemicals and heavy metals; and use of bioassays to guarantee safety of treated discharges to the environment.

In all cases, the impurities, contaminants, and solids removed from all wastewater treatment processes must ultimately be collected, handled, and disposed of safely, without damage to humans or the environment.

Treatment processes are chosen on the basis of composition, characteristics, and concentration of materials present in solution or suspension. The processes are classified as pretreatment, preliminary, primary, secondary, or tertiary treatment, depending on type, sequence, and method of removal of the harmful and unacceptable constituents. Pretreatment processes equalize flows and loadings, and precondition wastewaters to neutralize or remove toxics and industrial wastes that could adversely affect sewers or inhibit operations of publicly owned treatment works. Preliminary treatment processes are employed to protect plant mechanical equipment; remove extraneous matter such as grit, trash, and debris; reduce odors; and render incoming sewage more amenable to subsequent treatment and handling. Primary treatment employs mechanical and physical unit processes to separate and remove floatables and suspended solids and to prepare wastewater for biological treatment. Secondary treatment utilizes aerobic microorganisms in biological reactors to feed on dissolved and colloidal organic matter. As these microorganisms reduce biochemical oxygen demand and turbidity (suspended solids), they grow, multiply, and form an organic floc, which must be captured and removed in final settling tanks (**Fig. 1**). Tertiary treatment, or advanced treatment, removes specific residual substances, trace organic materials, nutrients, and other constituents that are not removed by biological processes.

On-site sewage treatment for individual homes or small institutions uses septic tanks, which provide separation of solids in a closed, buried unit. Effluent is discharged to subsurface absorption systems. *See* SEPTIC TANK; STREAM POLLUTION; UNIT PROCESSES; WATER TREATMENT.

Pretreatment. Wastewater, depending on its source, varies in quantity, strength, and composition. Pretreatment, particularly of sewage containing industrial wastes, neutralizes, alters, or removes nonbiologically degradable, toxic, hazardous, and highly corrosive materials before they enter the municipal sewer system. Chemical neutralization brings pH values to the range 6.0–8.0 and provides buffering capacity. Coagulation, flotation, and surface skimming of oils, greases, or other light constituents are enhanced by adding air or chlorine.

In the equalization process, widely varying flows are stored and then discharged to the treatment facility in a more uniform manner over a longer time period. Mixing and storing different flow streams neutralizes, dilutes, and dampens both quantities and concentrations, reducing shock loads on receiving systems. Equalization tanks are aerated to mix the sewage, inhibit septicity, increase gas exchange, and prevent sedimentation.

Preliminary treatment. Wastewaters contain sand, cinders, ceramic shards, glass, and other granular inorganic matter that is abrasive, causes wear on pumps and equipment, clogs lines, and occupies tank space. By controlling velocities in grit chambers (**Fig. 2**) at about 1 ft/s (0.3 m/s), dense granular matter [larger than 0.008 in. (0.2 mm)] settles, while materials with low specific gravity remain in suspension within the flow. Settled solids, being removed from the bottom by grit elevators or detritors, are agitated within the incoming flow. Agitation abrades and separates clinging organic matter from the inert solids, and the flowing water carries these removed putrescible substances away, leaving a washed grit that is ready for landfill disposal. *See* SEWAGE SOLIDS.

Unwanted materials like plastic objects, throwaway sanitary hygiene products, diapers, and towels, and items ranging from string to bed sheets or toothpicks to tree limbs become part of the sewage flow. Screens are provided to protect equipment from impact damage, from stringy material winding around and fouling rotating shafts, or from plastics and rags that clog valves. Bar racks and screens remove such large solids and flotables. Bar racks composed of rigid

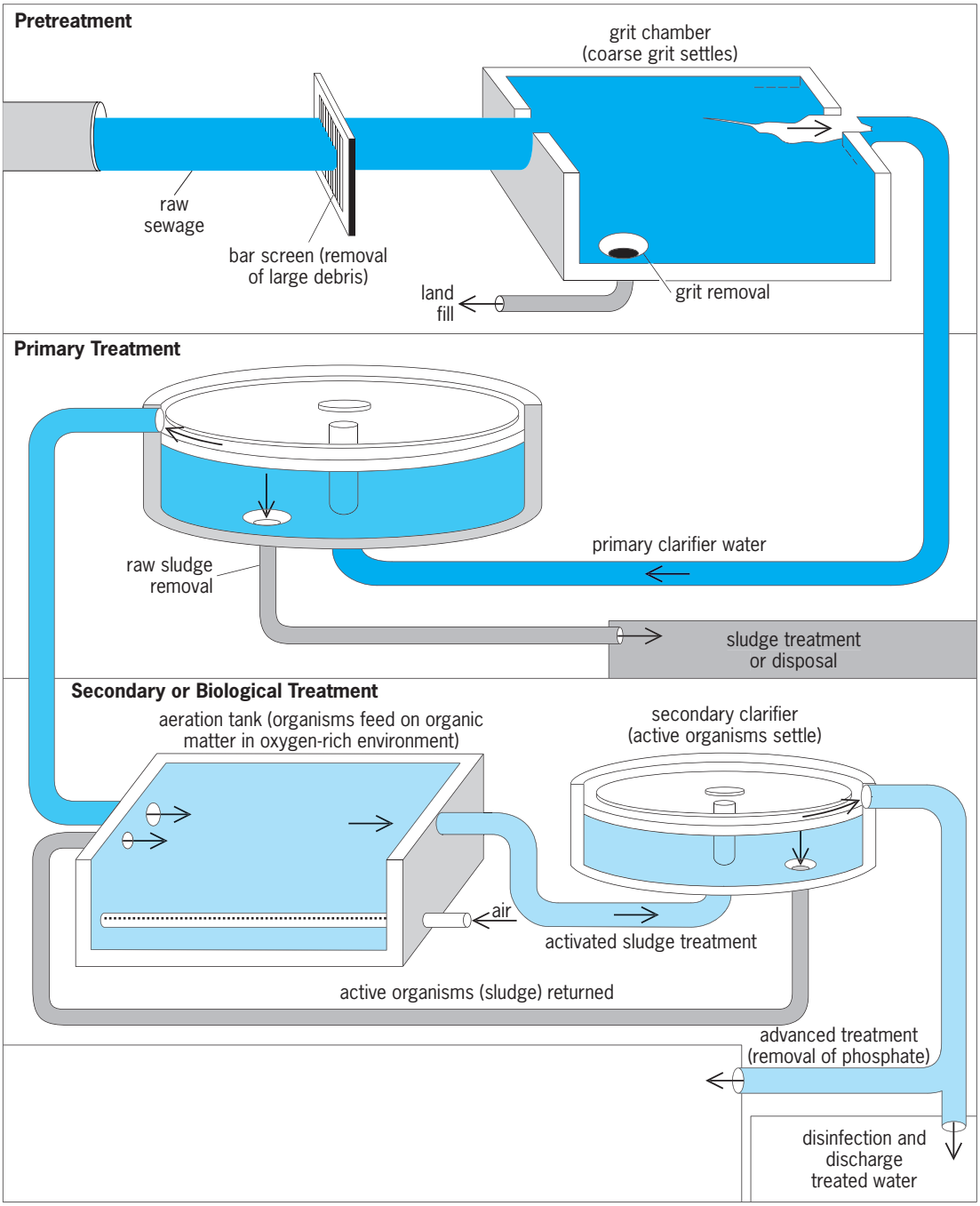


Fig. 1. Schematic diagram of liquid wastewater treatment through the secondary process. (*After B. J. Nebel, Environmental Science: The Way the World Works, 2d ed., Prentice-Hall, 1981*)

parallel bars, spaced 0.5–3 in. (12.5–75 mm) apart, held in place by support cross members, are installed within a controlled flow channel to catch this debris. Large, coarse material that is carried in the wastewater impacts the bar rack and is retained on the upstream face, from which it is mechanically removed and collected by traveling rakes.

Screens are formed of wire mesh, screen cloth, or perforated plates, with openings 0.5 in. (12.5 mm) or less. Coarse screens, with openings greater than 0.2 in. (5 mm) are used as pretreatment. Medium screens, 0.01–0.06 in. (0.8–1.5 mm), may be used in lieu of primary treatment. Fine screens, up to 0.003 in. (0.075 mm) openings, are used as final screening before plant discharge.

Disk, drum, band, and other styles of preliminary treatment screens are movable, with only a portion rotating within the flow. As the screen rotates, clean surfaces enter the wastewater, and previously immersed areas are lifted out and cleaned of solids caught on the surface.

Removal, handling, and disposal of screenings from bar racks and coarse screens are generally difficult maintenance problems, with associated odors and esthetically unpleasant conditions. To deal with this situation and to develop a smaller, more uniform

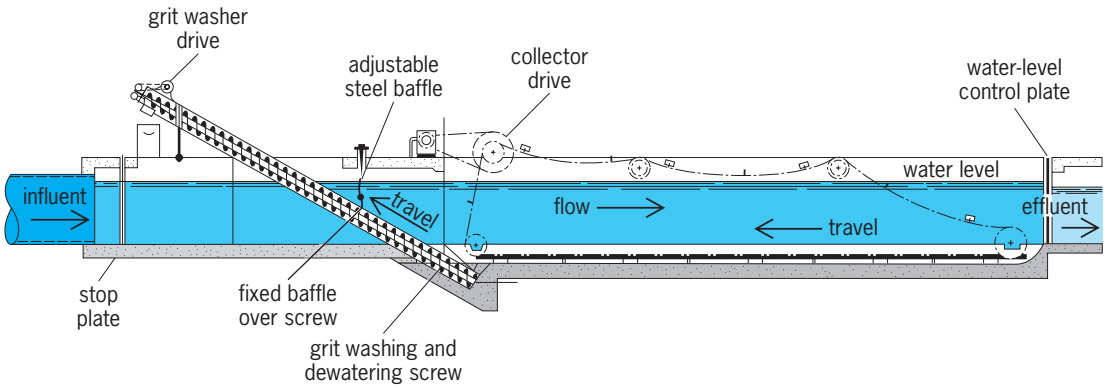


Fig. 2. Diagram showing components of a grit chamber. (*After Joint Committee of Water Pollution Control Federation and American Society of Civil Engineers, Wastewater Treatment Plant Design, MOP 8, 1977*)

sized solid, comminution (shredding) devices are used. Solids, trapped and held by the force and flow of liquid on a slotted plate or bars, are cut up or shredded by moving cutters or teeth. The solids, reduced in size, pass through the openings and are carried downstream to further treatment.

Primary treatment. Primary sedimentation, or clarification, is the first process where removal of substantial quantities of suspended solids and materials causing biochemical oxygen demand in wastewater flow occurs. Sedimentation tanks, rectangular or circular, operate on a continuous flow-through basis, yet they maintain relative quiescence to permit suspended solids to settle if the specific gravity is greater than that of water, or to float if the specific gravity is less. Mechanical devices remove the accumulated suspended solids (sludge) from the tank bottom, while floatable materials (skimmings) are taken off the surface. The clarified liquor, known as primary effluent, is discharged over the tank's effluent weirs.

Discrete particles of uniform size, shape, and specific gravity settle according to Stokes' law. Theoretically, if the velocity of a downward-settling particle is greater than the horizontal velocity of the carrying liquid, the particle will be captured and retained within the sedimentation tank. However, empirical values are used because sewage solids vary in shape, size, specific gravity, and concentration, and because the rate and path of settling are affected by temperature, age, density, and wind currents as well as eddies developed by mechanical scrapers and turbulence at influent and outlet structures.

Removal of suspended solids and biochemical oxygen demand is based on surface overflow rate and detention time. The surface overflow rate is expressed in gal/ft^2 (m^3/m^2) of tank surface area per day. Detention time varies from 30 min for minimal removal prior to some types of biological treatment, to 60–90 min (typical values within conventional plants), and longer times if chemical coagulation or nutrient removal is employed (**Fig. 3**).

A solids contact clarifier diverts the incoming wastewater through a layer or blanket of previously

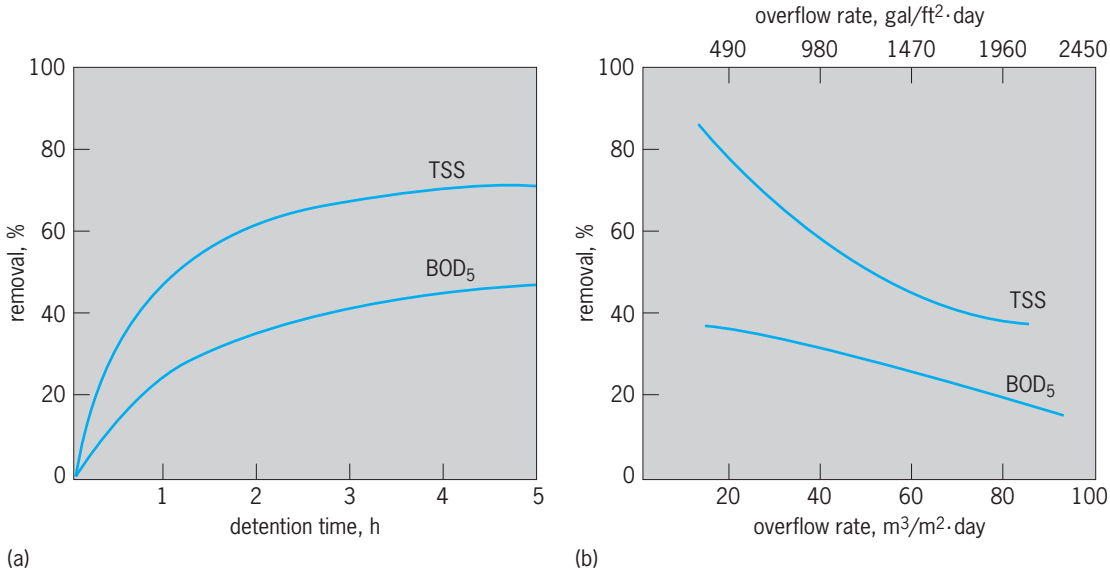


Fig. 3. Removal of suspended solids and biochemical oxygen demand in primary clarifiers at various (*a*) detention times and (*b*) overflow rates. (*After S. R. Qasim, Wastewater Treatment Plants, Planning, Design and Operations, CBS College Publishing, Holt, Rinehart and Winston, 1985*)

settled solids. Suspended solids within incoming sewage contacts the sludge blanket, is agglomerated, and remains trapped within the sludge blanket as the liquid rises and overflows the outlet weirs (**Fig. 4***a*). Although efficient for chemical and special suspensions, and frequently used in water treatment, solids contact clarifiers are used infrequently for biological floc, because extended detention times, particularly when it is warm, leads to septicity, gasification, and odors.

Based on discrete particles, suspended solids removal is related to available surface area and is independent of tank depth. The use of inclined tubes or plates (lamella) within sedimentation tanks increases the theoretical tank surface area and reduces both the settling time of the particles and the depth at which they come to rest. With highly inclined plates or tubes, there is increased settling efficiency, and solids contact benefits may also be derived, but some problems with biological decomposition remain (Fig. 4*b* and *c*).

Use of additives such as lime, alum, chlorine, clays, or polyelectrolytes increases suspended solids removal from wastewater, but produces more sludge that impacts subsequent handling of solids. The added costs for chemicals, feeders, and handling of solids make the use of chemicals less attractive.

Various configurations are employed to reduce velocities at inlet and outlet structures and prevent adverse mixing, currents, short circuiting, or scour. Weir overflow rates are limited to 15,000 gal per day per lineal foot of weir face (186 $m^3/m \cdot d$).

Static or inclined self-cleaning screens, centrifugal screens, and rotary-drum screens are being used to upgrade overloaded primary units and to replace preliminary and primary units. Removal of suspended solids and biochemical oxygen demand by screens is generally not as complete as by conventional units, but overall costs are lower. Subsequent plant units must be sized to compensate for these differences in removal. *See* PRECIPITATION (CHEMISTRY); SEDIMENTATION (INDUSTRY).

Secondary treatment. Biological organisms, predominantly aerobic bacteria, convert and metabolize dissolved and colloidal matter remaining in wastewater to new cellular material, carbon dioxide, and water. These biological processes use organisms that form either a fixed film (attached growth) or are motile (suspended growth); they convert degradable organics to different, larger forms without significantly reducing total biochemical oxygen demand. Safe discharge to the environment and compliance with effluent standards require that the biota and newly formed cellular material be removed. This is accomplished by provision of secondary settling tanks that function as an integral part of the biological process.

Given adequate detention time, dissolved oxygen, controlled pH, appropriate temperature, absence of any toxic materials or shock loads, these biological processes can reduce biochemical oxygen demand to under 15 mg/liter and suspended solids to under 20 mg/liter. Not all wastewater organics can be fully

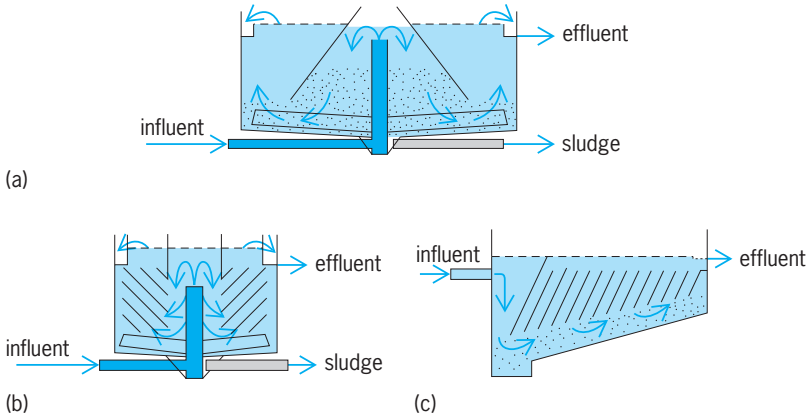


Fig. 4. Diagrams showing solids contact and inclined lamella settlers. (*a*) Circular settling tank using solids contact with sludge blanket. (*b*) Inclined lamella within a circular settling tank. (*c*) Inclined lamella within a rectangular sedimentation tank. (*After S. R. Qasim, Wastewater Treatment Plants, Planning, Design and Operation, CBS College Publishing, Holt, Rinehart and Winston, 1985*)

or adequately degraded in the time available and under the constraints found in publicly owned systems.

Fixed-film systems. Trickling filters (**Fig. 5**) contain beds of filter media—inert materials, with large surface area and voids, that support a slime growth, that is, an attached zoogleal mass composed of algae, fungi, protozoa, rotifera, nematoda, and most importantly, aerobic bacteria. Wastewater is applied to the surface of the filter media, which are made up of rock, slag, or synthetically formed shapes of manufactured materials, of different depths. The organic materials in this intermittently applied wastewater, flowing over and through the aerobic gelatinous matrix adhering to the support surfaces, undergo biologic coagulation, precipitation oxidation, and clarification as the flow moves downward to the underdrains. Organisms, in the presence of oxygen, convert the organic substances in the wastewater by various physical, chemical and enzymatic processes to energy for life, growth, and replication, and form carbon dioxide, nitrates, and sulfates as by-products.

This reactive zoogloeal mass converts organic matter and trapped particles to a humic mass as long

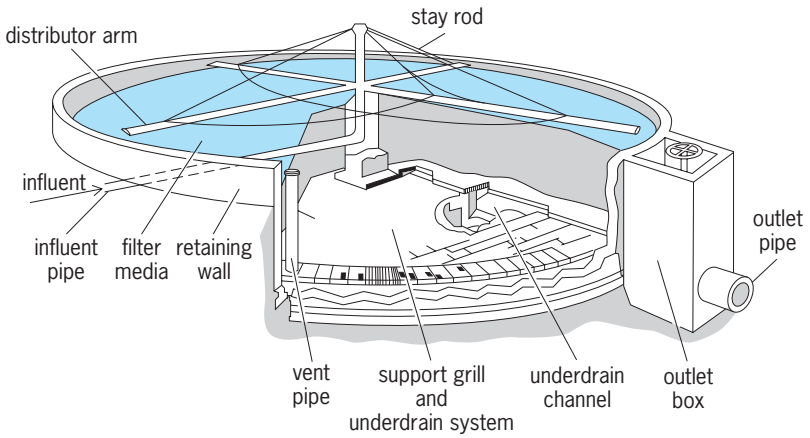


Fig. 5. Diagram of the components of a typical trickling filter. (*After Water Pollution Control Federation, Operation of Wastewater Treatment Plants, MOP 11, 1976*)

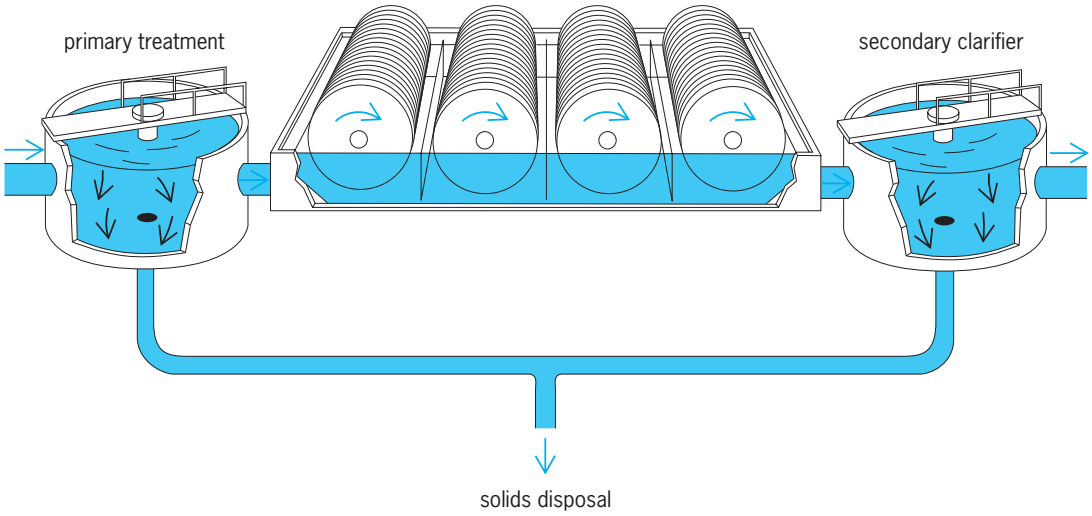


Fig. 6. Schematic diagram of a rotating biological contactor. (*After Water Pollution Control Federation, Operation of Wastewater Treatment Plants, MOP 11, 1976*)

as there is adequate oxygen transfer. If the film becomes too thick, or otherwise so impacted or compacted that oxygen cannot diffuse, portions become anoxic, and the film loses its capability to adhere to the support surface. Applied wastewater physically scours away nonattached portions, rejuvenating the filter by exposing areas for new colonization. Dislodged humic mass, known as the slough-off, is washed through the filter's large voids and conveyed through the underdrains to the secondary settling tank for ultimate capture and disposal.

Trickling filters are classified as standard or low-rate, intermediate, high-rate, and roughing filters according to hydraulic and organic loadings and rate of recirculation. Two-stage operation, the placing of two filters in series, is used for high organic loads. Rotating biological contactors (**Fig. 6**), also known as biodisks, use closely spaced plastic disks mounted on a rotating shaft. An aerobic zoogleal mass attaches itself and grows on the disks as they are intermittently exposed to organic materials and air as they rotate, with approximately 40% of their area immersed in the wastewater flow. The relative motion between the rotating disks and the flowing wastewater in the tank scours the attached film and carries the slough-off to the secondary settling tank.

Because daily wastewater flows into a publicly owned treatment works are highly variable both in amount of flow and in organic loading, they cause fluctuations in the liquid applications to the trickling filters. Rapid swings in biological density within the attached matrix develop, reducing efficiency. To provide flow and loading uniformity, a portion of the wastewater is recirculated after passing through the filter. This recirculation maintains continuous growth and sloughing, improves operation, increases efficiency, and reduces bed volume. Recirculation rate is the ratio of recycle flow to average daily plant inflow.

Total liquid flow, generally applied through rotating arms over the surface of the bed, intermittently doses each portion and is known as the hydraulic application rate. The average daily dry weight of new organic matter applied to the total bed volume is the organic loading rate. Typical application rates and design data are shown in **Table 1**.

Activated sludge. This is an aerobic suspended-growth process (**Fig. 7**) in which biodegradable organics in wastewater are intimately mixed with a concentrated mass of biota and oxygen within an aeration tank. New microorganisms grow and flocculate as the biotic mass adsorbs, oxidizes, and reduces the organic wastes. As the mixed liquor leaves the aeration tank following several hours of aeration, the biotic mass with the newly formed floc is separated within final settling tanks. A portion of the settled floc, the activated sludge, is returned to the aeration tank to maintain the required concentration of biota, while excess sludge is removed for solids handling and ultimate disposal. Air, required for utilization of dissolved oxygen in metabolism and respiration, is also provided to maintain mixing and prevent sedimentation in the aeration tank.

Modifications of the activated-sludge process relate to how, where, when, and for how long returned activated sludge and oxygen are introduced into the aeration tank and maintained in contact with the mixed liquor. These modifications of conventional activated sludge are known as tapered aeration, step aeration, extended aeration, contact stabilization, complete mix, and pure oxygen. Mechanical or compressed air systems are used to introduce and

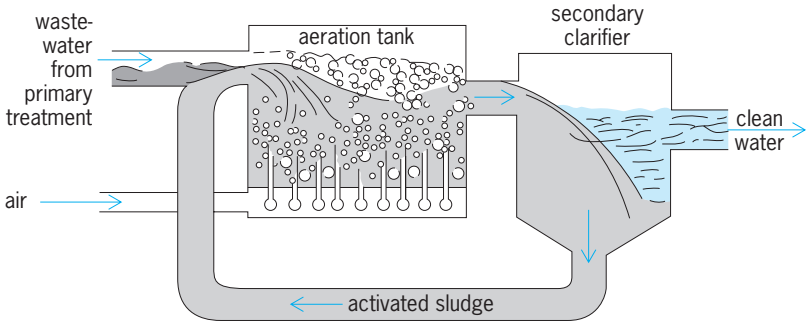


Fig. 7. Schematic diagram of activated sludge treatment. (*After B. J. Nebel, Environmental Science: The Way the World Works, 2d ed., Prentice-Hall, 1981*)

TABLE 1. Typical design information for trickling filters*

Item	Low rate	Intermediate rate	High rate	Superhigh rate	Roughing	Two-stage
Filter medium	Rock, slag	Rock, slag	Rock	Plastic	Plastic, redwood	Rock, plastic
Hydraulic loading, gal/ft^2 (liters/m^2) min	0.02–0.06 (1.2–3.6)	0.06–0.16(3.6–9.4)	0.16–0.64(9.438)	0.2–1.20 (12–70)	0.8–3.2 (47–187)	0.16–0.64 (9.4–38)
Mgal/acre (10^6 liters/ hectare) day	1–4 (9–36)	4–10 (36–90)	10–40 (90–360)	15–90 (135–810)	50–200 (450–1800)	10–40 (90–360)
BOD$_5$† loading, lb/10^{3}ft^3 (kg/m^3) day	5–25 (0.08–0.40)	15–30 (0.24–0.48)	30–60 (0.48–0.96)	30–100 (0.24–1.6)	100–500 (1.6–8)	60–120 (0.96–1.9)
Depth, ft (m)	6–8 (1.8–2.4)	6–8 (1.8–2.4)	3–6 (0.9–1.8)	10–40 (3–12)	15–40 (4.5–12)	6–8 (1.8–2.4)
Recirculation ratio	0	0–1	1–2	1–2	1–4	0.5–2
Filter flies	Many	Some	Few	Few or none	Few or none	Few or none
Sloughing	Intermittent	Intermittent	Continuous	Continuous	Continuous	Continuous
BOD$_5$† removal efficiency, %	80–90	50–70	65–85	65–80	40–65	85–95
Effluent	Well nitrified	Partially nitrified	Little nitrification	Little nitrification	No nitrification	Well nitrified

*After Metcalf & Eddy, Inc., G. Tchobanoglous, and F. L. Burton, *Wastewater Engineering: Treatment, Disposal, and Reuse*, 3d ed., 1991, and S. R. Qasim, *Wastewater Plants: Planning, Design, and Operation*, 1985.
†BOD$_5$ = Measure of strength of waste in miligrams per liter.

maintain dissolved oxygen and mixing. Mechanical systems entrain atmospheric air through brushes, impellers, propellers, or turbines, while compressed air is introduced near the tank bottom through porous diffuser plates, spirally wound tubes, spargers, nozzles, or injection jets (**Fig. 8**).

Depending upon the process used, parameters such as empirical ratios of biodegradable organics (food) to concentration of microorganisms, aeration contact time, concentration of solids in the mixed liquor, and loading criteria for aeration tanks are established to achieve the desired degree of reduction of biochemical oxygen demand. Suspended growth systems, with relatively short detention times, are more sensitive to hydraulic, organic, and toxic shock loads and require greater operational skill and monitoring than fixed-film processes. However, suspended-growth systems occupy much less space and can produce effluents with lower soluble biochemical oxygen demand and suspended solids (**Table 2**).

For the suspended-growth system to operate optimally, the solids in the mixed liquor leaving the aeration tank must be completely separated, settled, and returned to the aeration tank quickly. Final settling tanks, similar in general configurations to primary tanks, generally have lower values of surface overflow rate and are deeper, since both separation and sludge densification of the light, fluffy biofloc must be accomplished.

For industrial wastes and small plants, aerated lagoons, oxidation ponds, and waste-stabilization lagoons are employed for secondary treatment or for final polishing after conventional treatment. Oxidation ponds, with large surface areas, depths less than 5 ft (1.5 m), with loading rates not to exceed 50 lb/acre (56 kg/ha) of biochemical oxygen demand, rely on natural reaeration, diffusion, and photosynthesis. Bacterial and algal actions reduce putrescible organics, converting them to stable algal cells. Although some solids settle and undergo anaerobic decomposition, oxidation ponds can support fish life. Evaporation, screened or filtered settled outflow, maintains liquid levels.

Aerated lagoons are facilities that are intermediate between oxidation ponds and activated sludge, since no sludge is recycled. They are frequently used to pretreat sewage with high organic loads or

TABLE 2. Common design parameters for activated sludge systems

Process modification	Sludge retention time (θ_c), d	F/M,[a] d^{-1}	Aerator loading,[b] kg/m^3·d	MLSS,[c] mg/liter	Aeration[d] period, h	Recirculation ratio (Q_r/Q)
Conventional	5–15	0.2–0.4	0.3–0.6	1500–3000	4–8	0.25–0.5
Tapered aeration	5–15	0.2–0.4	0.3–0.6	1500–3000	4–8	0.25–0.5
Step aeration	5–15	0.2–0.4	0.6–1.0	2000–3500	3–5	0.25–0.75
Complete mix aeration	5–15	0.2–0.6	0.8–2.0	3000–6000	3–5	0.25–1.00
Extended aeration	20–30	0.05–0.15	0.1–0.4	3000–6000	18–36	0.5–2.0
Contact stabilization	5–15	0.2–0.6	1.0–1.2	1000–4000[e] 4000–10000[f]	0.5–1.0[e] 3.0–6.0[f]	0.5–1.0
Pure oxygen	8–20	0.25–1.0	1.6–3.3	6000–8000	2–5	0.25–0.5

[a]The F/M (ratio of food to microorganisms) is kilograms of BOD$_5$ applied per day per kilogram of MLVSS in the aeration basin [lb BOD$_5$/(lb MLVSS · d)]. MLSS = mixed-liquor suspended solids in milligrams per liter. MLVSS = mixed-liquor volatile suspended solids, that is, that portion of the solids in the aeration tank that are organic and hence biodegradable. Q_r/Q = ratio of return sludge to average daily plant inflow.
[b]Aerator loading is kilograms of BOD$_5$ applied per day per cubic meter of aeration tank capacity [lb/(10^3 ft^3 · d)].
[c]Generally the ratio of MLVSS to MLSS is 0.75–0.85.
[d]Aeration period is the time in hours that the mixed liquor is in the aeration tank.
[e]Contact tank.
[f]Reaeration or stabilization tank.

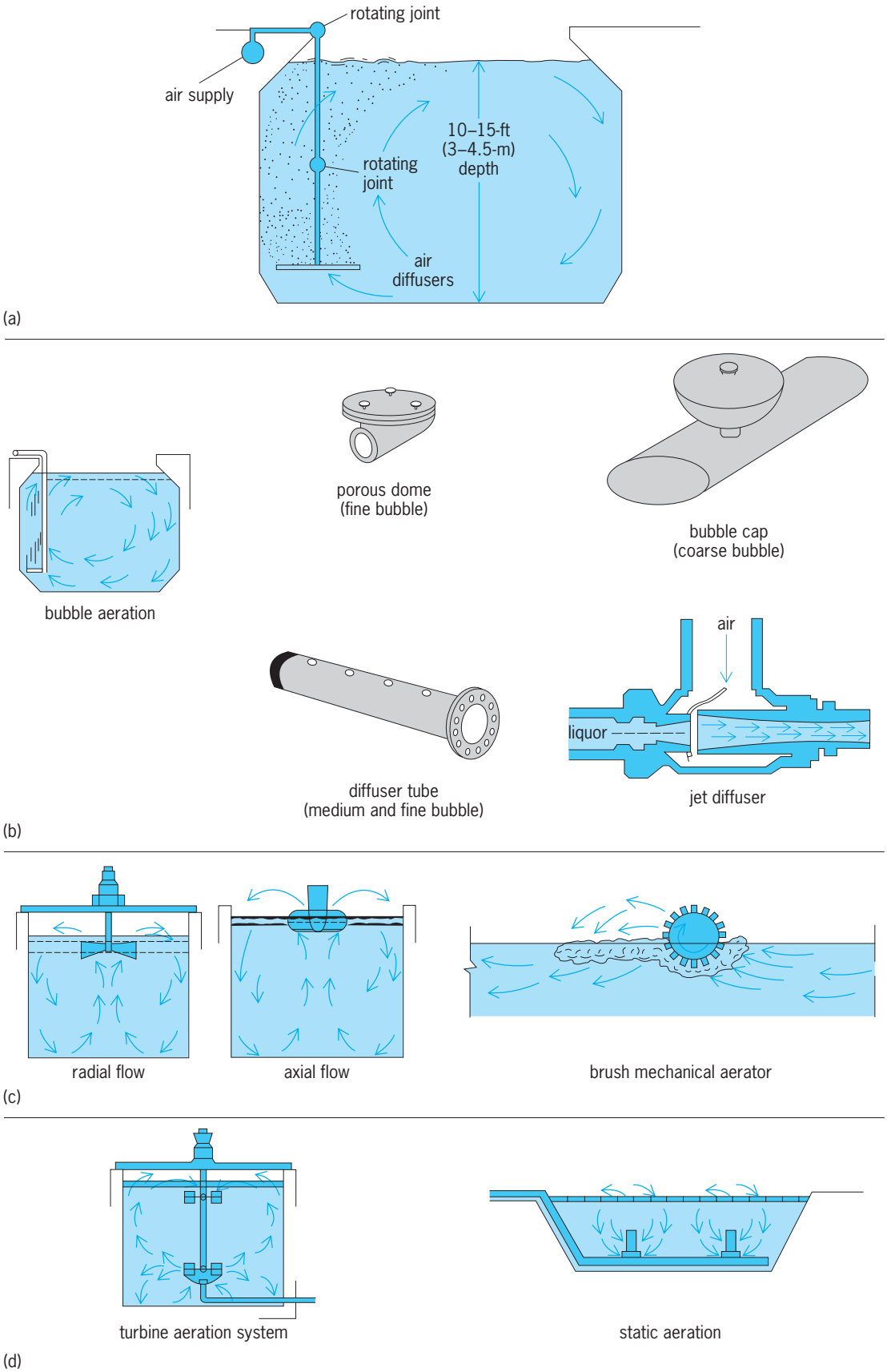


Fig. 8. Aeration systems and devices. (*a*) Cross section of a typical spiral-flow aeration tank showing the spiral-flow pattern created by aeration along one side (*after W. Viessman, Jr., and M. J. Hammer, Water Supply and Pollution Control, 4th ed., Harper and Row, 1985*). (*b*) Diffused aeration units. (*c*) Surface aeration systems. (*d*) Below-surface air release. (*b*, *c*, *d after W. W. Eckenfelder, Jr., Industrial Water Pollution Control, McGraw-Hill, 1989*)

industrial wastewaters prior to discharge to publicly owned treatment works. Basins are 8–18 ft (2.4–5.5 m) deep, and complete mixing is carried out by using mechanical devices, with detention times of several days. Some solids may settle and undergo anaerobic decomposition. When used as secondary treatment, basins are frequently followed by facultative lagoons, where liquid is retained for up to 20 or more days. The upper portion of a facultative lagoon is maintained aerobic, while the bottom, benthic zone, where solids accumulate, undergoes anaerobic decomposition. Released gases, rising particles, and bacteria are enmeshed, absorbed, destroyed, and converted to algal cells in the upper portion. Effluent may be subject to final mechanical screening, although frequently the effluent evaporates, percolates, or is used for land applications.

Final polishing. Rotating-drum microscreens, with stainless steel or plastic cloth having openings 30 micrometers in diameter, remove solids carried over from biological treatment and provide an effluent that is relatively free of turbidity. Filters are used for polishing of the final effluent as well as for removal of chemical precipitates from advanced treatment systems. Filtration, which is a complex process of screening, adsorption, and sedimentation, is defined and classified by the media, physical force, or application method employed. Flow may be induced by gravity, pressure, or vacuum; may be applied intermittently or continuously; and may be upflow, downward, or across sand, mixed media, granular carbon, membranes, or diatomaceous earth. Cleaning of filters, usually by backwashing, results in a new wastewater that must be captured, handled, and ultimately disposed of.

Disinfection. This process differs from sterilization in that it kills pathogenic organisms by physical or chemical means without total destruction of all the organisms present. Disinfecting agents should, at low concentrations, be highly selective and toxic to targeted pathogens. They must be stable, noncorrosive, economical, safe to humans and the environment, unaffected by pH or temperature, and long-lasting. None of the available agents or processes meets all or even most of these criteria. Disinfecting agents such as heat, phenols, alcohols, strong acids or alkalies, and halogens cause coagulation and denaturing of cell protein. Soaps and detergents injure cell membranes, change surface tension, or affect permeability. Oxidizing agents such as halogens, hydrogen peroxide, ozone, and heavy metals inhibit or denature enzymes and remove the free sulfhydryl (—SH) group. Direct sunlight, ultraviolet rays, or ionizing radiation destroys pathogens. Thermal processes and moist heat cause protein coagulation. Even though there are some problems with chlorine in the formation of trihalomethanes, it remains the agent most frequently used for waste water disinfection. Even with greater attention to control and dichlorination, it is relatively inexpensive, readily applied and controlled, and can be monitored by a simple and quick test. *See* ANTIMICROBIAL AGENTS; ENVIRONMENTAL ENGINEERING; STERILIZATION; ULTRAVIOLET RADIATION (BIOLOGY).

Advanced wastewater treatment. This is a treatment process (tertiary or polishing) that goes beyond the biological stage. Advanced wastewater treatment produces high-quality effluent, with reduced targeted toxic chemicals, nutrients, and trace organics. Such systems are a step in the cycle of transforming sewage to drinking water; they are closely related to processes that are used for industrial water treatment and for providing safe drinking water. Most advanced wastewater treatment systems include denitrification and ammonia stripping, carbon adsorption of trace organics, and chemical precipitation.

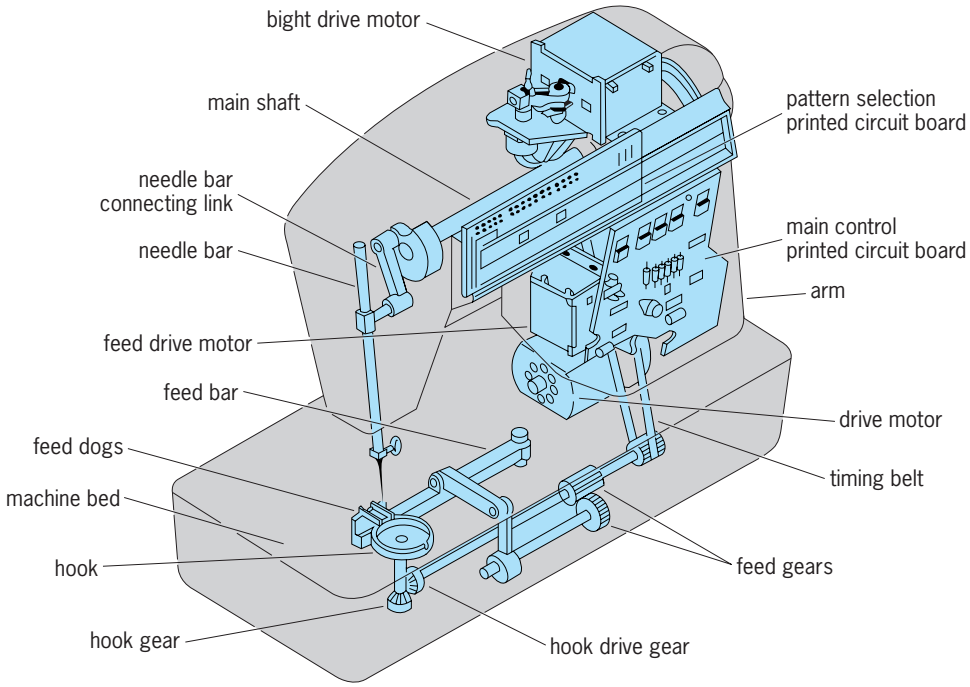
Nutrient removal of phosphorus and nitrogen is possible by chemical coagulation using lime or by biologic methods. The latter utilize systems having aerobic, anaerobic, and anoxic sequences to nitrify and then dentrify wastewaters. Ammonia can be removed by air stripping, breakpoint chlorination, or ion exchange. Carbon absorption traps refractory organics onto exposed carbon surfaces. Liquid can be passed over carbon in packed columns, or powdered carbon can be added to treated liquid and then filtered out on diatomaceous earth filters, on microstrainers, or by ultrafiltration. Evaporation, distillation, electrodialysis, ultrafiltration, reverse osmosis, freeze drying, freeze-thaw, floatation, and land application, with particular emphasis on the increased use of natural and constructed wetlands, are being studied and utilized as methods for advanced wastewater treatment to improve the quality of the treated discharged to reduce unwanted effects on the receiving environment. *See* ABSORPTION; DISTILLATION; EVAPORATION; SEWAGE; ULTRAFILTRATION; WETLANDS.

Gerald Palevsky

Bibliography. Metcalf & Eddy, Inc., G. Tchobanoglous, and F. L. Burton, *Wastewater Engineering: Treatment, Disposal, and Reuse*, 3d ed., 1991; S. R. Qasim, *Wastewater Treatment Plants, Planning, Design and Operation*, 1985; A. M. Thayer, Water treatment chemicals: Tighter rules drive demand, *Chem. Eng. News*, 68(13):17–34, 1990; W. Viessman and M. J. Hammer, *Water Supply and Pollution Control*, 4th ed., 1985; Water Environment Federation, *Design of Municipal Wastewater Treatment Plants*, Man. Prac. 8, 1993; Water Environment Federation, *Operation of Municipal Wastewater Treatment Plants*, Man. Prac. 11, 1991; S. R. Qasim, *Wastewater Treatment Plants, Planning, Design and Operation*, 2d ed., 1998; A. M. Thayer, Water treatment chemicals: Tighter rules drive demand, *Chem. Eng. News*, 68(13):17–34, 1990; W. Viessman and M. J. Hammer, *Water Supply and Pollution Control*, 6th ed., 1998; Water Environment Federation, *Design of Municipal Wastewater Treatment Plants*, Man. Prac. 8, 1993; Water Environment Federation, *Operation of Municipal Wastewater Treatment Plants*, Man. Prac. 11, 1991.

Sewing machine

A mechanism that stitches cloth, leather, book pages, and other material by means of a double-pointed needle or an eye-pointed needle. In ordinary



Components of a modern sewing machine. (*Singer Co.*)

two-threaded machines, a lock stitch is formed (see **illus**.). An eye-pointed reciprocating needle carries an upper thread through the layers of fabric, forming a loop beneath the material. A shuttle carrying a bobbin of under thread passes through the loop. Alternatively, a rotary hook takes the loop of upper thread and passes it around the bobbin of under thread. The needle withdraws, and a thread take-up lever pulls the stitch tight. The machine carries out these necessary motions and also feeds the material past the needle intermittently between each pass of the needle. A presser foot held against the material with a yielding spring adjusts itself automatically to variations in thickness of material and allows the operator to turn the material as it feeds through the machine. A cluster of cams, any one of which can be selected to guide the needle arm, makes possible a variety of stitch patterns. *See* CAM MECHANISM.

Frank H. Rockett

Sex determination

The genetic mechanisms by which sex is determined in all living organisms. The nature of the genetic basis of sex determination varies a great deal, however, among the various forms of life.

Forms of sexuality. There are two aspects of sexuality: the primary form involves the gametes, and the secondary aspect is gender.

Gametes. In its broadest usage the term "sex" refers to the processes that enable species to exchange materials between homologous chromosomes, that is, to effect recombination. Generally, recombination is essential to their mechanism for reproduction. For most organisms this involves, either exclusively or as one stage in the life cycle, the formation of special cells, known as gametes, by meiosis. *See* GAMETOGENESIS.

Most sexually reproducing species produce two different kinds of gametes (**Fig. 1**). The relatively large and sessile form, an ovum or egg, usually accumulates nutriments in its cytoplasm for the early development of the offspring. The relatively mobile form, a sperm (or pollen grain in many plants), contributes little beyond a haploid chromosome set. Thus the primary form of sex differentiation determines which kind of gamete will be produced. The formation of gametes usually involves the concomitant differentiation of specialized organs, the gonads, to produce each kind of gamete. The ova-producing gonad is usually known as an archegonium or ovary (in flowering plants it is part of a larger organ, the pistil or carpel); the gonad producing the more mobile gametes is usually known as a testis in animals and an antheridium or stamen in plants. *See* OVARY; OVUM; SPERM CELL; TESTIS.

Gender. In most animals and many plants, individuals become specialized to produce only one kind of gamete. These individuals usually differ not only in which kind of gonad they possess but also in a number of other morphological and physiological differences, or secondary sex characteristics. The latter may define a phenotypic sex when present, even if the typical gonad for that sex is absent or nonfunctional. The form that usually produces ova is known as female; the one that usually produces sperm or pollen is known as male. Since some sexual processes do not involve gametes, the more universal application of the term "gender" refers to any donor of genetic material as male and the recipient as female.

Humans. Human males and females tend to differ in the following secondary sex characteristics: cephalic

hairline, facial hair, larynx and voice, breasts, axillary hair, body hair, pubic hair, body configuration, phallic size, carrying angle of the arm, and knee relations. The female phenotype usually includes breast development and a characteristic disposition of fat over the hips and thighs. The pelvis tends to increase its lateral diameters, producing in adaptation a typical angulation of the forearm to the arm and the touching of the knees in the standard anatomical position. The pubic hair is kinky or curly and level at the upper margin, and the facial hair is inconspicuous. In contrast, the male pubic hair is straight and extends toward the umbilicus, while the facial hair increasingly becomes more conspicuous, as does that on the chest. After puberty the male's hairline tends to recede, the larynx to increase in size (the so-called Adam's apple), and the voice to deepen. The phallic organ (penis or clitoris) is usually much larger in the male, especially after puberty. Most of these characteristics are controlled by hormones, produced principally by the gonads (and therefore traceable to the primary sex determination). Their anomalous appearances in the two sexes are usually traceable to endocrine abnormalities. *See* ANDROGEN; ENDOCRINE MECHANISMS; ESTROGEN; HORMONE.

Plants. Sexual reproduction is not always accompanied by the kinds of differentiation described above. Thus, the major structures in the life cycles of some simple organisms such as the pond scum *Spirogyra*, a green alga, are filaments composed of haploid cells; these organisms sometimes reproduce sexually by passing the contents of a cell from one filament into a cell of an adjacent filament by means of a protoplasmic bridge or conjugation tube (**Fig. 2**). Since both conjugating cells are the same size and shape, they are known as isogametes. (The corresponding term for unequal-sized gametes is anisogamy; oogamy is anisogamy when the larger gamete is nonmotile.) Union of the two haploid isogametes produces a diploid zygospore that undergoes meiosis and germinates to form new filaments composed of haploid cells. The majority of plant species are monoecious, with both kinds of gonads on the same plant (**Fig. 3***a*). Plants such as the date palm, willow, poplar, and papaya that bear male and female gonads on separate plants are dioecious (Fig. 3*b*). They occur in about 60 of the 400 or so families of flowering plants, 20 of which are thought to contain exclusively dioecious species. *See* REPRODUCTION (PLANT).

Animals. Although the sexes are distinct in most animals, many species are hermaphroditic; that is, the same individual is capable of producing both eggs and sperm. This condition is particularly common among sessile or sluggish, slowly moving forms, including most flatworms, a large percentage of the worms that are parasitic in humans and other mammals, many members of the annelid phylum (such as earthworms and leeches), many mollusks (such as garden snails, oysters, scallops, and clams), and some fishes. Some hermaphroditic and monoecious species are homothallic; that is, the eggs and sperm of the same individual can combine successfully; but most are heterothallic, the gametes being capable only of cross-fertilization, often evolving special mechanisms to ensure its occurrence. In some cases the cross-fertilization can involve only certain members of the same species, a condition very similar to separation of the sexes even though the cross-fertilizing individuals are both hermaphroditic.

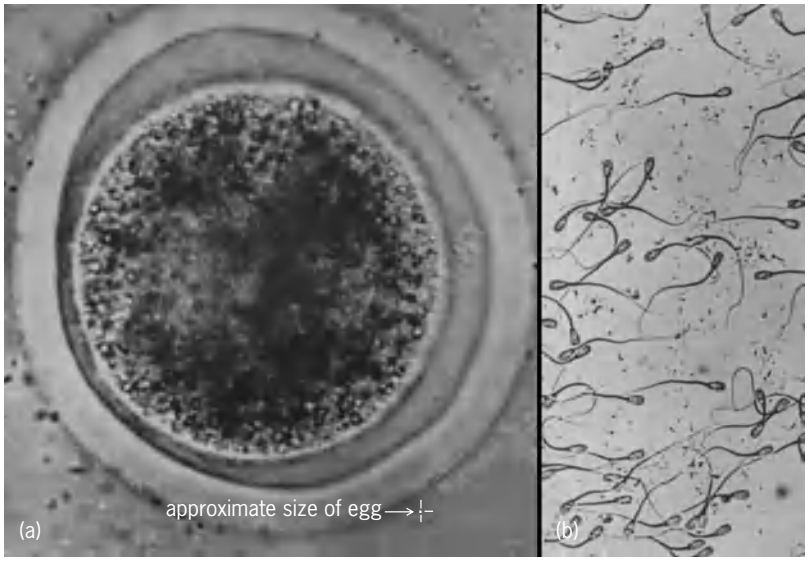


Fig. 1. Bovine gametes. (*a*) Egg. (*b*) Sperm. (*From R. C. Cook, A one-celled cow, J. Hered., 23(5):193–199, 1932*)

Chromosomal sex differentiation. Sex differentiations are often accompanied by consistent chromosomal dimorphisms, leading to the presumption that the chromosomal differences are related to, and possibly responsible for, the sex differences. Indeed, the chromosomes that are not alike in the two sexes were given the name sex chromosomes very soon after their discovery. Some workers use the term "heterosomes" to distinguish them from the autosomes,

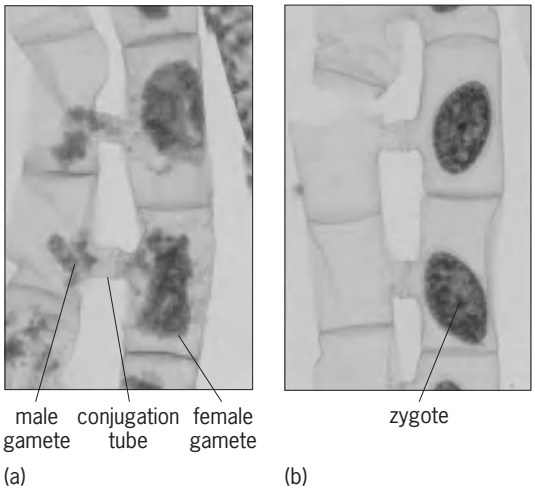


Fig. 2. *Spirogyra* (watersilk), showing the union of isogametes. (*a*) The condensed protoplasts in the cells on the left are functioning as male gametes that are migrating through the conjugation tubes to the stationary female gametes on the right. (*b*) Zygotes have been produced in the cells as a result of fusion of gametes. (*Adapted with permission from K. R. Stern, S. Jansky, and J. E. Bidlack, Introductory Plant Biology, 9th ed., McGraw-Hill, 2003*)

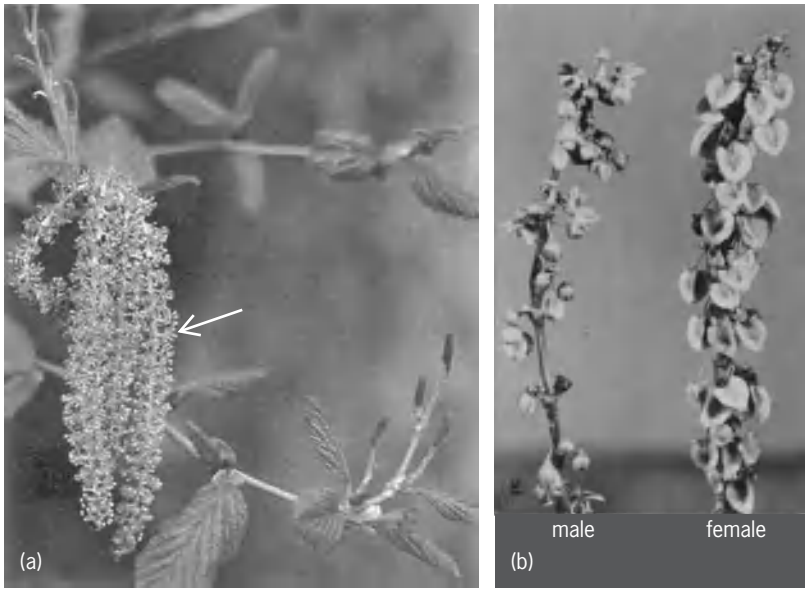


Fig. 3. Sexuality in plants. (*a*) Staminate and pistillate flowers of a birch (*Betula* sp.). Birches are monoecious; their staminate flowers hang down in long, yellowish tassels, while their pistillate flowers mature into clusters of small, brownish, conelike structures (*reprinted with permission from P. H. Raven and G. B. Johnson, Biology, 6th ed., McGraw-Hill, 2002*). (*b*) Dioecious plant *Rumex hastatulus*, with stamens and pistil located on separate plants (*courtesy of B. W. Smith*).

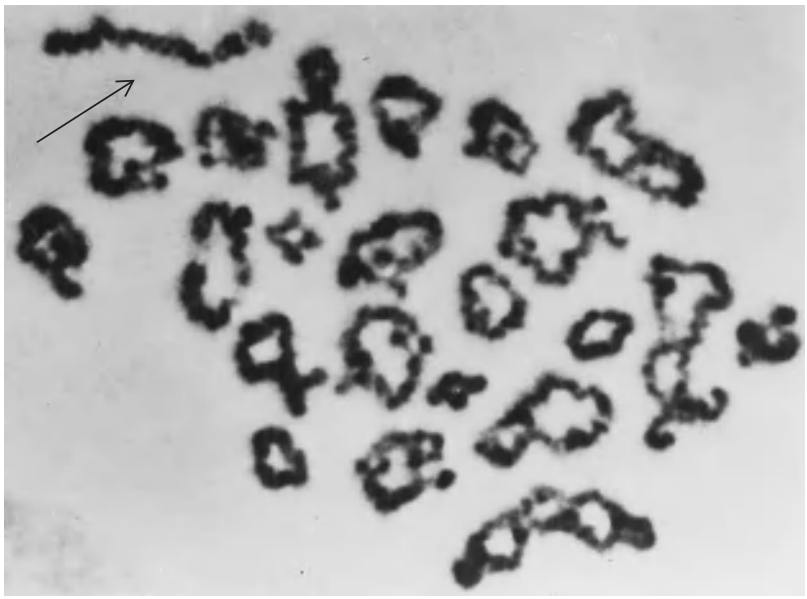


Fig. 4. Evidence of end-to-end pairing between X and Y chromosomes (arrow) during human spermatogenesis during diakinesis, the last stage of the first meiotic prophase in the male. Chromosome pairing is an indication of the presence of regions of homology. The figure shows the end-to-end pairing between X and Y chromosomes, reflecting the very short region of homology between them. (*From M. Levitan, Textbook of Human Genetics, 3d ed., Oxford University Press, 1988*)

which are the chromosome pairs that are morphologically identical to each other in both sexes.

In most species, one of the sex chromosomes, the X chromosome, normally occurs as a pair in one gender but only singly in the other. The gender with two X chromosomes is known as the homogametic sex, because each gamete normally receives an X chromosome after meiosis. The gender with only one X chromosome generally also has a morphologically different sex chromosome, the Y chromosome. The X and Y chromosomes usually pair to some extent at meiosis (**Fig. 4**), with the result that the XY is the heterogametic sex, with half its gametes containing an X and half containing a Y.

Although the earliest demonstrations of sex chromosomes were based on differences in gross morphology, geneticists soon noted that the fundamental dimorphism of X and Y chromosomes lay in their genic contents: X chromosomes of the species share homologous loci, just as do pairs of autosomes, whereas the Y chromosome usually has few, if any, loci that are also represented on the X. Thus X and Y chromosomes are sometimes very similar in shape or size but are almost always very different in genetic materials.

In the larger bisexual organisms, females are normally XX and males are XY. This is true in mammals, including humans (**Fig. 5**); in many insects, including most species in the orders Coleoptera and Diptera as well as many species of the orders Heteroptera, Dermaptera, and Neuroptera; and in many dioecious plants, including gingkos, poplars, and willows. Some organisms have a similar system of homogametic, XX, females and heterogametic males, but the males lack a Y chromosome, and their diploid number is odd, being one less than that of females. Such an XX:XO arrangement (O indicating the absence of a Y or X chromosome) occurs in many insects of the orders Orthoptera, Homoptera, and Heteroptera. Y chromosomes are commonly absent also in spiders and nematodes.

The reverse system (that is, the males are homogametic and the females heterogametic) occurs in

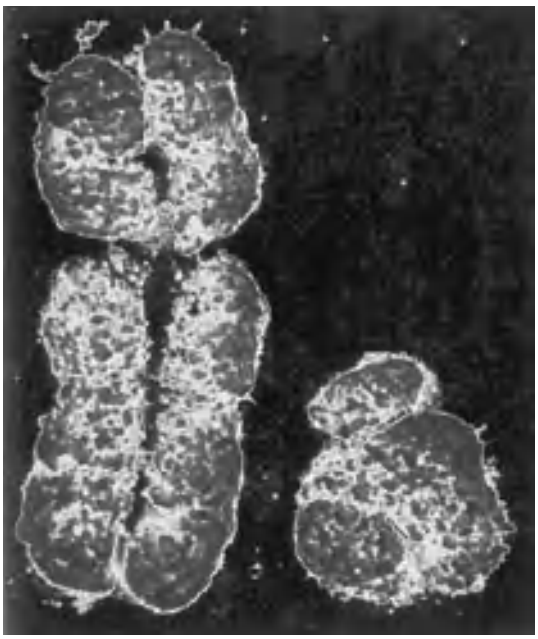


Fig. 5. The X and Y chromosomes. In humans, females are the homogametic sex (XX) and males are the heterogametic sex (XY). (*Reprinted with permission from R. Lewis, Human Genetics: Concepts and Applications, 5th ed., McGraw-Hill, 2003*)

the insect order Lepidoptera (moths and butterflies), all birds, the blood fluke (*Schistosoma* douthitti) and, among plants, the strawberry. Some moth females are XO instead of XY. The XX males and XY females are less exclusively found in the insect orders Diptera and Odonata (dragonflies), crustaceans, fishes, amphibians, and reptiles. Instead of XX and XY, the sex chromosomes in these organisms are sometimes termed ZZ and ZW, respectively. In some fishes, female heterogamety is the rule in "domesticated" forms, but wild members of the same species exhibit male heterogamety. Many more complex systems of sex chromosomes exist, particularly species with more than one kind of X or more than one kind of Y. These systems may be classified formally as X_1X_2Y, $X_1X_2X_3Y$, X_1X_2O, XY_1Y_2, and so forth, according to the number of nonhomologous (or only partly homologous) chromosomes in the heterogametic sex. The spiders and nematodes belong to this group. Since the species with these systems are usually closely related to species with a simple sex chromosomal difference (XX:XY or XX:XO), the more complex systems are thought to have evolved by means of chromosomal interchanges between sex chromosomes and autosomes that in effect converted autosomes, or parts of them, into sex chromosomes. *See* CHROMOSOME.

Genetic basis of sex differentiation. Even organisms having the same form of chromosomal dimorphism often differ in the genetic basis for sex determination. Thus, although both humans and *Drosophila melanogaster* have the XX female, XY male system, a fly with two or more X's and a Y is a female (usually fertile), whereas humans that are XXY, XXXY, XXXXY, and so on are phenotypic males (though usually infertile). In contrast, an XO *Drosophila* is a sterile male, whereas an XO human is a phenotypic female, but also sterile and with gonadal dysgenesis.

Y-dominant system. The major factor in sex differentiation in humans is a locus on the short arm of the Y chromosome designated SRY or SrY (for sex-determining region of the Y). This comparatively small gene contains no introns and encodes for a protein with only 204 amino acids. The protein appears to be a deoxyribonucleic acid (DNA)-binding type that causes somatic cells of the developing gonad to become Sertoli cells that secrete a hormone, Müllerian inhibiting substance (MIS), that eliminates the Müllerian duct system (the part that would produce major female reproductive organs). The gonad is now a testis, and certain cells in it become the Leydig cells that produce testosterone, which causes the primordial Wolffian duct system of the embryo to develop the major male reproductive organs (**Fig. 6**). If no MIS is produced, further development of the Müllerian duct structures occurs, and in the absence of testosterone the Wolffian ducts disappear, producing the normal female structures. Embryos lacking SRY or having mutated forms of it normally become females even if they are XY.

The presence of SRY does not ensure the male phenotype. XY's with testicular fertilization, for example, are phenotypic females, lacking ovaries and a uterus because a mutant X-chromosome gene results in absent or defective androgen receptors. However, phenotypic males that are apparently XX have been found to contain SRY that had been translocated elsewhere in their genome. In a number of instances, SRY came to reside in the X chromosome derived from the father as a result of crossing-over in the "pseudoautosomal" region in which the small amount of pairing between the X and Y chromosomes occurs.

This system of sex determination is called Y-dominant. It appears to be characteristic of almost all mammals, even marsupials, among others. In Y-dominant salamanders of genus *Axolotl*, the pertinent genes are female-determining, as the males are the homogametic sex. *See* MUTATION.

While SRY is the primary gene, many other genes, both autosomal and X-chromosomal, are involved in the course of developing the two sexes in mammals, as manifested by disorders in gender development or sex reversal when they mutate. Many genes remain to be fully elucidated, and more probably remain to be discovered. Some, referred to as *SOX* genes, are autosomal and X-linked loci that are SRY-related in that they a share with *SRY* a 79-amino-acid sequence

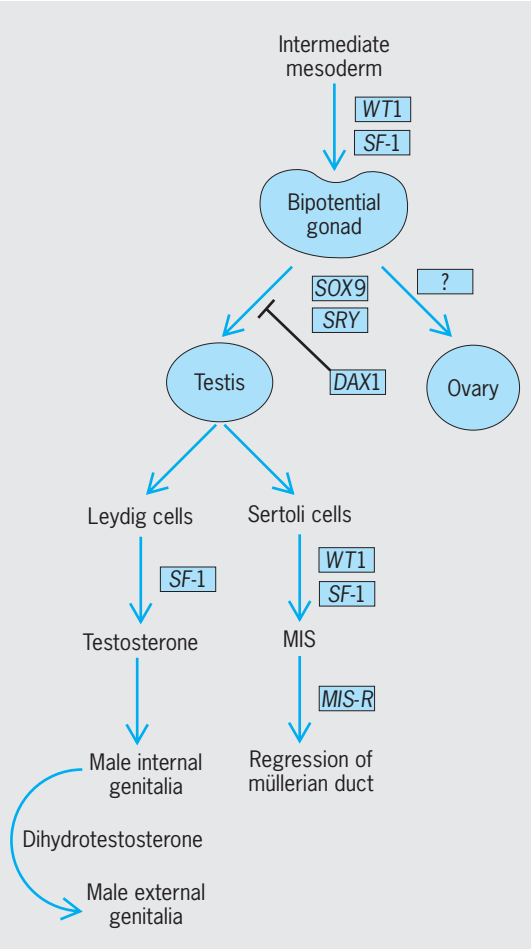


Fig. 6. (*Reprinted with permission from D. Federman, Three Facets of Sexual Differentiation, NEJM, 350:323–324, 2004.*)

(the HMG box); they are believed to be involved in gonadal function, spermatogenesis, and development of the nervous system (Fig. 6). Another gene, *WT*, whose mutated form causes Wilms' tumor of the kidneys, is expressed in Sertoli cells and thought to be normally involved with the action of *SRY* (Fig. 6), as homozygotes for complete absence of the normal allele fail to form gonads in mice. An X-chromosome gene, *Dax1*, appears to be an antagonist of *SRY*; that is, it can interfere with testis formation when present in double dose in XY individuals, and it may normally be important in ovary development (Fig. 6). Female development also involves an autosomal gene called *Wnt-4*, one of a group of similar genes that have critical signaling actions during embryogenesis in animals as diverse as the worm *Caenorhabditis elegans, Drosophila*, mice, and humans.

Balance theory. The Y-dominant mechanism would not fit the *Drosophila* data, the XX/XO situations, or most of the more complex sex chromosome complements. These are best explained by the so-called balance theory. This hypothesis holds that the sexual phenotype in species with male heterogamety depends on a balance between many female-determining genes on the X-chromosome and male-determining genes on the autosomes, often referred to as the X:A ratio. It is essentially a ratio of the relative quantities of proteins determined by the X-chromosome and autosomal genes involved. When at least two X's are present together with a normal complement of autosomes, then X:A = 1, and the individual would develop as a female. In the presence of only a single X, X:A = 0.5; however, the autosomal genes prevail, and the result is a male.

In *Drosophila melanogaster* the so-called master switch in this mechanism is a gene called sex-lethal (*Sxl*). Very early in development, around the blastomere stage, if X:A = 1, *Sxl* is activated, and the result is a female; if X:A = 0.5, *Sxl* is silent, and the result is a male. Activated *Sxl* produces a protein which controls the degree of transcription of other genes, such as the normal alleles of *dsx* (doublesex) and *msl* (male-specific lethal), in development of the gonad and female characteristics, and in dosage compensation. Much remains to be worked out about the exact nature and action of all the genes involved.

When X:A balance determination occurs in female heterogamety, the female-determining genes would be the ones on the autosomes, and vice versa. The balance method is used also by another well studied species, the nematode worm, *Caenorhabditis elegans*. The master switch is again an X-linked gene, *xol-1*, but, unlike in *Drosophila*, it is activated when the *male* X:A ratio is 0.5, and repressed when the ratio X:A = 1.

Theoretically, the balance hypothesis would be compatible with a hypothesis of single genes on the X chromosomes or the autosomes being responsible. Indeed, a number of autosomal mutations are known in *Drosophila* that transform chromosomally female individuals into normal, albeit sterile, males or into intersexes. Experiments have shown, however, that the number of female-determining genes on the X chromosome is large and that they exist on all regions of the chromosome. The balance hypothesis implies that the Y chromosome of many organisms contributes little, if anything, to sex determination. However, in certain moth species all the female-determining genes are thought to be on the Y chromosome (moth females being XY) and the male-determining genes are on the X, but the primary mechanism is still one of balance. *See* GENE; GENE ACTION.

Because the sex determination methods of *D. melanogaster* and *C. elegans* have been the best studied at the genic level, it was assumed that the balance method was more widespread than has proved to be the case. Although *Sxl* has been found to be generally present even in Diptera outside the genus *Drosophila* (Muscids, for example), it has been determined that it does not act as a master switch in these organisms, leading to the realization that "Y-dominance" or some modification of it, rather than X:A balance, is probably the predominant form of sex determination.

Other systems. In many of the social insects, such as the honeybees, ants, and wasps, diploid individuals become females, while haploid ones, which develop parthenogenetically from unfertilized eggs, become males. Sex chromosomes are not identifiable as such, but one of the chromosomes contains a locus with multiple alleles that is critical to sex determination. Heterozygosity at this locus results in a viable female, whereas homozygosity results in males that usually (or in some species, always) die early or are sterile; only those with a single allele develop as viable males.

A single locus that determines sex has also been demonstrated in garden asparagus. Usually dioecious, occasionally viable seeds are produced by self-fertilization in staminate plants that also have rudimentary pistils. The offspring of these seeds produce staminate and pistillate plants in mendelian ratios that are typical for segregation in a heterozygote for a single pair of alleles; the gene for staminate is dominant to the one for pistillate.

In bacteria that can undergo recombination, the process resembles conjugation in the green alga. Meiosis does not occur, however, so that the exchange takes place directly between the transferred genetic material and the chromosome of the recipient cell. The capacity to donate genetic material in this process (maleness) derives from the presence in some cells of an F (fertility) factor, that is, a self-duplicating circular piece of deoxyribonucleic acid (DNA). The F factor is an episome: that is, it may exist free in a cell or be integrated into the chromosome. Strains where it is integrated are known as Hfr (high frequency of recombination) because they undergo conjugation about a thousand times as often as strains in which the F factor is free-standing. The presence of an F factor brings about a change of morphology not unlike a secondary sex characteristic; the resultant male produces appendages, called sex

pili or fimbriae, which attach to a female cell (one lacking an F factor) and provide the cytoplasmic bridge for transfer of DNA. *See* BACTERIAL GENETICS.

Environmental influence. In some organisms, sex determination seems to depend on the environment rather than on genes or chromosomes. In the sea worm *Dinophilus*, large eggs become females and small eggs become males. In another sea worm, *Bonnelia*, sex seems to depend on where the free-swimming larvae land: those that land on the sea bottom develop into females, each with a long proboscis; those that land on this proboscis develop into tiny males that live as parasites in the genital ducts of the female. In the horsetail plant, *Equisetum*, sex depends on growth conditions: if conditions are good, the plants become female, and if conditions are poor, they are male. Similarly, in many fishes and reptiles the temperature at which the eggs develop determines the resultant sex ratio. In alligators, for example, warm temperatures result in an excess of males, whereas cool temperatures produce an excess of females. In a number of turtle species the opposite is true, with females predominating in the warmer environment.

The internal environment also must play an important role in sex determination much the same way that such factors determine how some cells of the pluripotential zygote become nervous structures whereas others become the linings of the gut. This is particularly likely in hermaphroditic animals and monoecious plants, and the many organisms, such as the algae, where the male and female tissues seem to have the same genes and chromosomes.

X inactivation. Development of the Y-dominant pattern of sex determination in which the Y chromosome is almost entirely devoid of genes that are counterparts (homologs) of those on the X chromosome posed a serious developmental problem: Since the genome is effectively balanced when there is a single X chromosome with a diploid set of autosomes, as occurs in XY and XO individuals, one would expect that having a second X chromosome would be detrimental to the homogametic sex because of the resultant imbalance. In other words, it is a problem of dosage compensation.

The mammalian solution of the problem was discovered by the British geneticist Mary Lyon, and is referred to as the Lyon hypothesis. From her work with mice, she concluded that one of the X chromosomes in somatic tissues of the XX female is inactivated early in embryogenesis, so that in effect the genes on only one X chromosome are active, just as in the heterogametic sex. Both the active and inactive X's are faithfully reproduced in subsequent mitoses.

The inactivated X is the late-replicating one. Since inactivation is, in most cases, a random process, in some cell lineages the paternal, while in others the maternal, X chromosomes are the ones affected. Hence, mammalian females that are heterozygous for X-linked loci are usually mosaics for them. The exception is the locus for steroid sulfatase, which is never inactivated.

X inactivation appears to occur in all mammals, even the monotremes (such as Platypus), though there is some variability in the amount of X chromosome and the type of cells that escape inactivation. In humans with supernumerary X's (for example, XXXY males or XXX females), all X chromosomes in excess of one are usually inactivated, correlating with the number of heterochromatic sex chromatin (Barr) bodies seen in their somatic cells.

Inactivation begins at an inactivation center in band 13 on the long arm of the chromosome (Xq13) and spreads upstream and downstream from it, extending to normally autosomal cells that have been traslocated to the portion of the X chromosome bearing the inactivation center (but not to the X chromosome genes on the portion of the translocated X that lacks Xq13). The part of Xq13 involved has been narrowed down to a gene called Xist (for inactive specific transcript), which governs the production of an RNA (but no protein) that remains in the nucleus and binds to the inactive X. Apparently it acts to block transcription from the inactive X. The corresponding gene on the active X appears to be silenced by methylation. *See* HUMAN GENETICS. M. Levitan

Bibliography. J. S. Bogan and D. C. Page, Ovary? testis?: A mammalian dilemma, *Cell*, 76:603–607, 1994; J. J. Bull and E. L. Charnov, Enigmatic reptilian sex ratios, *Evolution*, 43:1561–1566, 1989; K. M. Cadigan and R. Nusse, *Wnt* signaling: A common theme in animal development, *Genes Develop.*, 11:3286–3305, 1997; T. W. Cline, Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state, *Genetics*, 107:231–277, 1984; T. W. Cline and B. J. Meyer, Vive la difference: Males vs. females in flies vs. worms, *Annu. Rev. Genet.*, 30:637–702, 1996; J. Godwin, J. A. Luckenbach, and R. J. Borski, Ecology meets endocrinology: environmental sex determination in fishes, *Evcl. Develop.*, 5:40–49, 2003; L. O. F. Penalva and L. Sanchez, RNA binding protein Sex-lethal (*Sxl*) and control of *Drosophila* sex determination and dosage compensation, *Microbiol. Mol. Biol. Rev.*, 67:343–359, 2003; G. Saccone, A. Pane, and L. C. Polito, Sex determination in flies, fruitflies, and butterflies, *Genetica*, 116:15–23, 2002; D. C. A. Shearman, The evolution of sex determination systems in dipteran insects other than *Drosophila, Genetica*, 116:25–43, 2002; J. Weiss et al., *Sox3* is required for gonadal function, but not sex determination, in males and females, *Mol. Cell. Biol.*, 23: 8084–8091, 2003.

Sex-linked inheritance

The inheritance of a trait (phenotype) that is determined by a gene located on one of the sex chromosomes. These traits, including many diseases, have characteristic patterns of familial transmission; for a century these patterns were used to localize the genes for these traits to a specific chromosome. As the genomes of many organisms have now been

completely sequenced, reliance on the specific pattern of inheritance to map genes on sex chromosomes has waned.

Sex determination. The expectations of sex-linked inheritance in any species depend on how the chromosomes determine sex. For example, in humans, males are heterogametic, having one X chromosome and one Y chromosome, whereas females are homogametic, having two X chromosomes. In human males, the entire X chromosome is active (although not all genes are active in every cell), whereas one of a female's X chromosomes is largely inactive. Random inactivation of one X chromosome occurs during the early stages of female embryogenesis, and every cell that descends from a particular embryonic cell has the same X chromosome inactivated. The result is dosage compensation for X-linked genes between the sexes (the Lyon hypothesis). A specific gene on the long arm of the X chromosome, called *XIST* at band q13, is a major controller of X inactivation. This pattern of sex determination occurs in most vertebrates, but in birds and many insects and fish the male is the homogametic sex. *See* SEX DETERMINATION.

Sex-linked traits. In general terms, traits determined by genes on sex chromosomes are not different from traits determined by autosomal genes. Enzymes, structural proteins, receptor proteins, and so forth are encoded by genes on both autosomes and sex chromosomes. Sex-linked traits are distinguishable by their mode of transmission through successive generations of a family. The term sex linkage was introduced by T. H. Morgan in 1914, based on his Nobel prize–winning investigations of *Drosophila*. His experiments with the X chromosome of fruit flies provided the first proof of linkage of any type, the first linkage maps, convincing evidence that heritable traits are encoded on chromosomes, and the concept that x-rays produce heritable changes (mutations) by disrupting the chromosome.

Sex-linked inheritance has been studied extensively in humans, for whom it is preferable to speak in terms of X-linked or Y-linked inheritance.

X-linked traits. Red-green color blindness was the first human trait proven to be due to a gene on a specific chromosome. This was accomplished by E. B. Wilson in 1911 when he correctly interpreted family studies showing X-linked recessive inheritance. The characteristics of this pattern of inheritance are readily evident (**Fig. 1***a*). Males are more noticeably or severely affected than females; in the case of red-green color blindness, women who have one copy of the mutant gene (that is, are heterozygous or carriers) are not at all affected. Among offspring of carrier mothers, on average one-half of their sons are affected, whereas one-half of their daughters are carriers. Affected fathers cannot pass their mutant X chromosome to their sons, but do pass it to all of their daughters, who thereby are carriers. A number of other well-known human conditions such as XY gonadal dysgenesis behave in this manner, including the two forms of hemophilia, Duchenne muscular dystrophy, the fragile X syndrome that is a common cause of mental retardation, and glucose-6-phosphate dehydrogenase deficiency, which predisposes to hemolytic anemia. *See* ANEMIA; COLOR VISION; HEMOPHILIA; MUSCULAR DYSTROPHY.

Human color vision is based on expression of blue, green, and red pigments in specific cells of the retina. The gene for the blue pigment is located on chromosome 7, and the genes for the red and green pigments are closely linked at the end of the long arm of the X chromosome (**Fig. 2**), exactly where classic family studies of red-green color blindness had predicted them to be. In people with normal color vision, there is one copy of the red pigment gene followed by a variable number of green pigment genes in tandem array. Various degrees of red-green color blindness result from disruption of one or more of these genes, primarily by the process of unequal recombination. *See* RECOMBINATION (GENETICS).

Y-linked traits. A number of Y-linked conditions (such as Sertoli-cell only syndrome) account for a substantial percentage of infertility in males. Y-linked traits never occur in females and are passed to every son of an affected father (Fig. 1*b*). Variations in DNA sequence along the Y chromosome, which are often completely common and harmless (polymorphisms), are also passed only from fathers to sons. Such polymorphisms have been extremely useful to anthropologists in studying migration patterns and male lineages.

Advances in mapping. Refined cytogenetic and molecular techniques have supplemented family

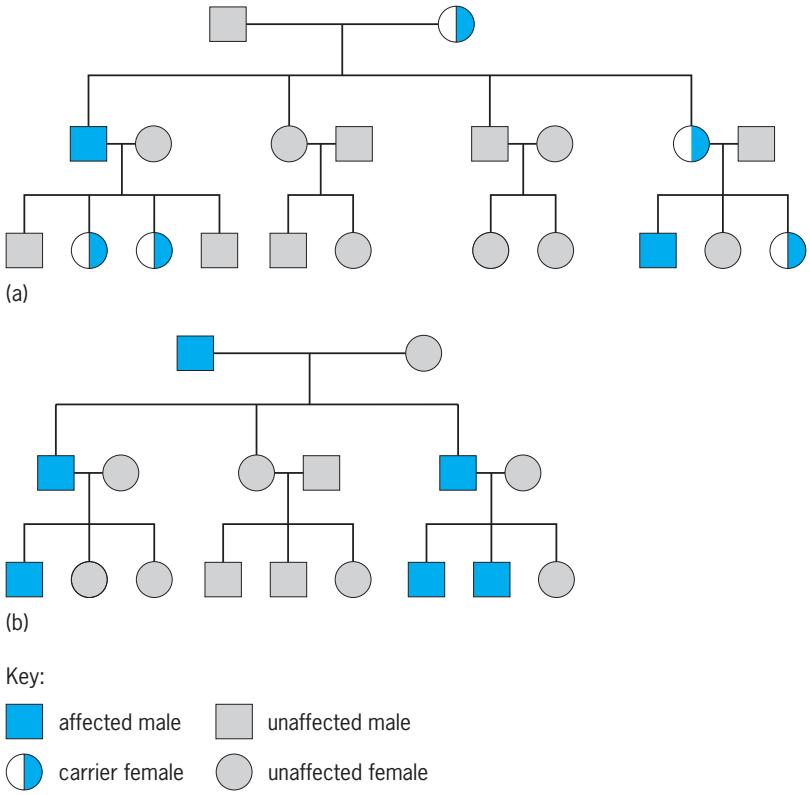


Fig. 1. Patterns of sex-linked inheritance in humans. (*a*) X-linked recessive inheritance. (*b*) Y-linked inheritance.

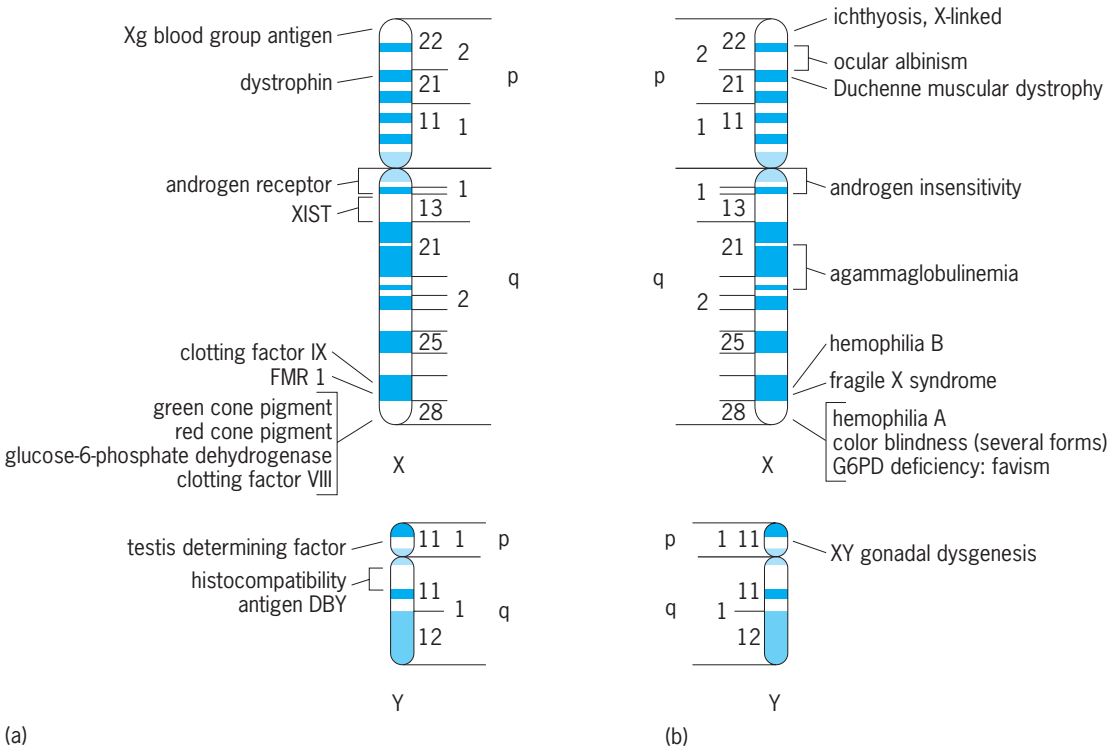


Fig. 2. Genetic maps of the human sex chromosomes. (*a*) Genes and (*b*) diseases mapped to specific chromosome regions.

studies as a method for characterizing sex-linked inheritance and for mapping genes to sex chromosomes in many species. Over 500 human traits and diseases seem to be encoded by genes on the X chromosome, and nearly 500 genes have been mapped. Among mammals, genes on the X chromosome are highly conserved. Thus, identifying an X-linked trait in mice is strong evidence that a similar trait, and underlying gene, exists on the human X chromosome. The human Y chromosome harbors about 50 genes. *See* GENETICS; HUMAN GENETICS.

Reed E. Pyeritz

Bibliography. J. Brennan and B. Capel, One tissue, two fates: Molecular genetic events that underlie testis versus ovary development, *Nat. Rev. Genet.*, 5:509–521, 2004; M. Kohn et al., Wide genome comparisons reveal the origins of the human X chromosome, *Trends Genet.*, 20:598–603, 2004; C. Krausz et al., Y chromosome polymorphisms in medicine, *Ann. Medi.*, 36:573–583, 2004; L. Ronfani and M. E. Bianchi, Molecular mechanisms in male determination and germ cell differentiation, *Cell Mol. Life Sci.*, 61:1907–25, 2004.

Sextant

A navigation instrument used for measuring angles, primarily altitudes of celestial bodies. Originally, the sextant had an arc of 60°, or one-sixth of a circle, from which the instrument derived its name. Because of the double-reflecting principle used, such an instrument could measure angles as large as 120°.

In modern practice, the name sextant is commonly applied to all instruments of this type regardless of the length of the arc, which is seldom exactly 60°. Occasionally, the terms octant, quintant, and quadrant are applied to instruments having arcs of 45°, 72°, and 90°, respectively.

The optical principles of the sextant are similar to those of the prismatic astrolabe. *See* PRISMATIC ASTROLABE.

Development. John Hadley, an Englishman, and Thomas Godfrey, an American, are credited with inventing the marine sextant independently. The sextant has remained virtually unchanged since each man designed similar instruments in 1730. Various, less accurate, more cumbersome angle-measuring devices preceded the sextant, including the common quadrant, astrolabe, cross-staff, backstaff or sea

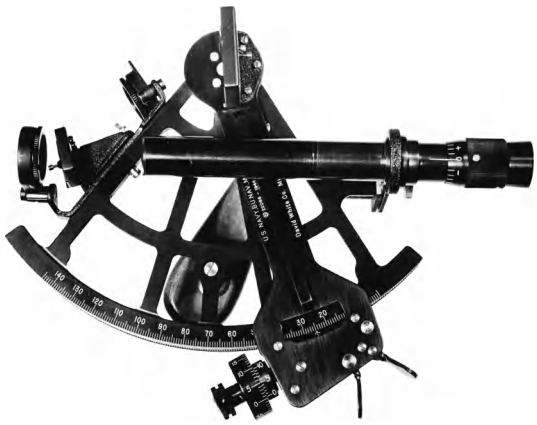


Fig. 1. Marine sextant.

Fig. 2. Periscopic sextant for use in aircraft.

quadrant, and nocturnal. Tycho Brahe (1546–1601) invented several instruments with arcs of 60°, which he called sextants. In 1700, Isaac Newton designed an instrument incorporating the double-reflecting principle, but his invention was not made public until after sextants had been constructed by Hadley and Godfrey.

Classes. Modern instruments may be grouped into two classes, marine sextants and air sextants.

Marine sextant. This is an instrument designed for use by mariners. It utilizes the visible sea horizon as the horizontal reference. It is equipped with a removable telescope to magnify the image of the horizon, and with shade glasses or filters to reduce glare (**Fig. 1**). The marine sextant has been widely used as a symbol of navigation.

Air sextant. An instrument designed for use in aircraft is called an air sextant. Such sextants vary widely in design, but all modern instruments have built-in artificial horizons. A bubble is most commonly used for this purpose, but pendulums and gyroscopes have also been used. Most modern air sextants are periscopic to permit observation of celestial bodies without need of an astrodome in the aircraft (**Fig. 2**).

Sextant altitude corrections. Altitudes measured with the sextant are subject to certain corrections. These may be classified as those related to (1) inaccuracies in reading, (2) inaccuracies in horizontal reference, (3) bending of the ray of light from the body, (4) adjustment to equivalent reading of the center of the body, and (5) adjustment to the equivalent reading at the center of the Earth. *See* CELESTIAL NAVIGATION.

Alton B Moody

Bibliography. N. Bowditch, *American Practical Navigator*, U.S. Defense Mapping Agency Hydrographic Topographic Center Publ. 9, 1977; E. S. Maloney, *Dutton's Navigation and Piloting*, 14th ed., 1985.

Sexual dimorphism

Any difference, morphological or behavioral, between males and females of the same species. In many animals, the sex of an individual can be determined at a glance. For example, roosters have bright plumage, a large comb, and an elaborate tail, all of which are lacking in hens. Sexual dimorphism arises as a result of the different reproductive functions of the two sexes and is a consequence of both natural selection and sexual selection. Primary differences, such as the structure of the reproductive organs, are driven by natural selection and are key to the individual's function as a mother or father. Other differences such as the peacock's (*Pavo cristatus*) enormous tail are driven by sexual selection and increase the individual's success in acquiring mates. *See* ORGANIC EVOLUTION.

Reproduction. A less obvious sexual dimorphism is the difference in size of male and female gametes. In animals, females produce eggs (ova) and males, sperm. In plants, females produce ovules and males, pollen. Fusion of egg and sperm is required for production of offspring. In nearly all cases, the sperm (or pollen) are substantially smaller and more numerous than the ova. Eggs are large because they contain nutrients essential for development of the embryo. However, the sole purpose of sperm is to fertilize the egg. Sperm do not contain any nutrients and can therefore be small. For the same investment of nutrients, a male can produce more sperm than a female can produce eggs. Human males, for example, produce about 300 million motile sperm per ejaculate, whereas females normally produce only one egg (30,000 times larger than a single sperm) per month. *See* GAMETOGENESIS.

In order to perform their sexual function, males have testes in which sperm are produced. In animals with internal fertilization, males often have an intromittent organ (penis) to transfer sperm to the female. Females often have a more complex reproductive system. As well as ovaries producing eggs, female mammals have a uterus, placenta, and mammary glands. Female insects have sperm storage organs and an ovipositor for laying eggs. *See* FERTILIZATION; INSECTA; REPRODUCTION (ANIMAL); REPRODUCTIVE SYSTEM.

Body size. In nearly all animal groups (apart from mammals and birds), females are larger than males because bigger females tend to produce more eggs. Therefore large female size is favored. For example, in the Atlantic silverside fish (*Menidia menidia*), large mothers lay more eggs than do small mothers, and the same is true for the Edith's checkerspot butterfly (*Euphydryas editha*). Males of many species, such as weevils, prefer to mate with bigger females. However, size does not appear to limit males' ability to produce sperm. The male blanket octopus (*Tremoctopus violaceus*) weighs only 0.25 g and is dwarfed by the female which may be up to 10,000 times heavier. *See* OCTOPUS.

In contrast, among mammals and birds males are generally the larger sex. For example, male great bustards (*Otis tarda*, a rare ground-living bird) are more than twice as heavy as females, and male elephant seals (*Mirounga angustirostris*) weigh eight times as much as the females. The reason that male elephant seals are so large is that they fight for access

to females, with only the largest male getting to mate with all the females on the beach. *See* PINNIPEDS.

In a few species, males may be either large or small but rarely of intermediate size. The males of the Pacific coho salmon (*Onchorhynchus kisutch*) come in two sizes. Large hook-nosed males fight over females, whereas inconspicuous satellite males sneak in to fertilize the eggs, undetected by their large rivals. *See* SALMONIFORMES.

Differences in body size and shape can be caused by factors other than reproductive success. Sexual dimorphism can arise as a consequence of competition between the sexes over resources, or because the sexes use different resources. For example, in many species of snake, males and females use different habitats and eat different food, which has led to differences in their head shape and feeding structures. Spider females often cannibalize their partners during mating. Being small is advantageous for the male, as he can sneak up and mate with the female without prior detection, and his small size may also make him less tempting as a meal.

Weapons. Some males use weapons to compete with each other for access to mates. Deer and antelope use their antlers to fight over females, as do stag beetles (*Lucanus cervus*) and narwhals (*Monodon monoceros*). Elephants fight with their tusks, and pheasants fight using leg spurs. Females of these species do not fight over mates and therefore lack these weapons.

Ornamentation in animals. Males and females also show dimorphic structures that are not involved in combat but are used in mate choice. Female reproductive capacity is limited by resources (both egg production and pregnancy are energetically costly), whereas male reproductive capacity depends largely on how many females he mates with. Males tend to compete with each other for access to as many females as possible, while females look for the best candidate among different males.

Female preference for showy males can lead to exaggerated male ornaments. For example, the peacock has an impressively large tail which he uses in an elaborate dance to impress the drab peahen. Bright plumage and showy tails, specifically long tails, are widespread ornaments in male birds. Females tend to prefer the male with the most elaborate tail because only males of high genetic quality are capable of producing showy tails. Therefore, by mating with a male with a showy tail, the female will have high-quality offspring. Tail length may be an honest indication of the male's ability to resist parasites which will be passed on to the offspring. For example, male barn swallows (*Hirundo rustica*) with long tails carry fewer parasitic mites than do short-tailed males. Females prefer these long-tailed males, and their offspring inherit the ability to resist mite infection.

Female choosiness can lead to increasingly exaggerated male ornaments, even when the ornamentation interferes with his survival. The longer the peacock's tail, the more it hampers his flight and maneuverability. Further exaggeration within the species is halted when the tail finally becomes too costly to the bearer, and the peacock with the overexaggerated tail does not survive long enough to mate and pass the genes encoding this trail to the next generation.

Females may also prefer bright males even when ornamentation conveys no information about quality. Female stickleback fish are drab silver, whereas males have a red belly that varies in brightness. Females prefer mates with a bright red belly, hence bright males mate more often than dull males. Sons inherit the red color from their fathers, and daughters inherit a preference for red-bellied mates from their mothers.

While sexual selection tends to favor bright, showy males, natural selection often promotes cryptic coloration in females. For example, the male mallard duck (*Anas platyrhynchos*) has iridescent plumage to woo females. In contrast, the female is drab and inconspicuous, which helps her to avoid predators while incubating the eggs.

Ornamentation in plants. Plants also differ in showiness. Many plants bear both male and female flowers (simultaneous hermaphrodites), but male flowers are sometimes larger and more conspicuous. For example, the female catkins of willow are dull gray compared with the bright yellow male catkins, because male flowers compete with each other to attract pollinators. In plant species with separate sexes (dioecious), males tend to produce more flowers than females. For example, males of the American holly (*Ilex opaca*) produce seven times as many flowers as females in order to increase their chances of pollen transfer to females. *See* FLOWER; POLLINATION.

Vocalization and chemical signals. Animals and plants show marked sexual dimorphism in other traits. Calling, singing, pheromones, and scent marking can all be explained by competition between males and by female mate choice. *See* ANIMAL COMMUNICATION.

Vocalization. Just as with visual ornaments, male vocal calls are used in competition with other males and to attract females. The male red-winged blackbird (*Agelaius phoeniceus*) sings to defend his territory from rival males as well as to attract a female. Male bullfrogs call only during the breeding season, and their impressive croaks reveal their size to females. The silent females prefer the call of large, high-quality males. Male crickets also call to attract mates, and the longer a male calls, the more females he will attract. Unfortunately for the male, calling also announces his presence to predators and parasitoids (such as bats and flies), exemplifying how mating success and individual survival can sometimes conflict. *See* TERRITORIALITY.

Chemical signals. Chemical signals are commonly used for attracting mates in insects, attracting pollinators in plants, and territorial marking in mammals. In many nocturnal insects, males use pheromones rather than visual ornaments to attract mates. Male and female moths produce different pheromones. Females emit long-range scents to attract males. At close range, males court the female using a

different scent and females choose the sweetest-smelling male. *See* PHEROMONE.

In mammals such as domestic cats and mice, males use pheromones to stake out their territory. Males can detect the status of a rival from his scent. When choosing between mates, female mice prefer the odor of dominant males.

There is also some evidence that females prefer males with different (complementary) major histocompatibility complex (MHC) genes and that this preference is based on scent. The MHC is a large cluster of genes essential to the immune system of vertebrates. MHC-based mate choice may maximize parasite resistance in offspring and/or minimize the risks of mating with kin. Even human females can discriminate between potential partners by their smell, preferring mates with a different MHC variant to their own. Remarkably, in humans the same odor receptor genes are present both on sperm and in the nose. *See* HISTOCOMPATIBILITY.

Plants also use scent to attract pollinators and reward them with nectar. In some species, only male flowers provide nectar. In the Central American tree *Jacaratia dolichaula*, female flowers smell just as sweet as male flowers but do not provide any nectar rewards.

Behavior. Associated with morphological sexual dimorphism are several behavioral differences between males and females. Many of these are related to locating a mate, competition between males, and female choosiness.

In most animal and plant species, males invest more time and energy into finding mates. Female angler fish live in total darkness at a depth of 2000 m (6600 ft) or more. The male finds his spouse using his huge, sensitive nose. He invests enormous resources into locating a female, having the largest nose relative to his size of any vertebrate. In many insects, only the males have wings so that they can fly in search of females. Males of many species are ready to mate before females. For example, many male insects reach adulthood before females, enabling them to monopolize and mate with females as soon as they become sexually mature. Males of most migrating birds species arrive and set up territories at the breeding ground before females. Male plants tend to flower earlier in the season and more often than females. This aids pollen flow toward the scarcer, less attractive female flowers. In the shrimplike *Gammarus*, males guard females by carrying them around and waiting for them to lay their eggs. Males expend a lot of energy in carrying females for several days. As a result, unusually for invertebrates, *Gammarus* males are larger than females. *See* GAMMARIDEA.

Males may form large aggregations (leks) where they display their sexual ornamentation to onlooking females. During the breeding season, sage grouse (*Centrocercus urophasianus*) form leks of 15–20 males in which they strut and display their plumage to females. Females often visit several males before selecting a mate. Frogs also form large leks, with several hundred males calling throughout the night. The more vocal the frog, the more attractive he is to females.

Male displays are costly to produce and depend on overall condition, which is determined by many genes. For example, in dung beetles (*Onthophagus taurus*) there is genetic variation between males in courtship rate. Females prefer males with a high courtship rate, as this reflects their overall condition.

Some males even vigorously display to females when other males are not present. The male bird of paradise performs an elaborate and complicated dance to impress a female, and the male bowerbird builds an elegant structure, the bower, which he adorns with feathers and flowers to impress the female. The male bowerbird may even destroy rival males' bowers and steal their decorations to enhance the beauty of his own creation. *See* BEHAVIORAL ECOLOGY; REPRODUCTIVE BEHAVIOR.

Parental care. Animals also show sexual dimorphism relating to their roles as parents. Many parents continue to provide for their young after birth, with the female performing the bulk of the care in most species. In marsupials such as kangaroos, females have a brood pouch in which the young develop and feed on the mother's milk. Similarly, female cichlid fish protect their growing young by brooding them in their mouths. Earwig females tend the nest, keeping their young free of fungi, and scorpion mothers carry their young on their back. Female mammals suckle their young, whereas males cannot because they lack mammary glands. However, some mammals (such as gibbons and prairie voles) and many birds share parental duties, with both males and females feeding and protecting the young. *See* MATERNAL BEHAVIOR.

Sex role reversal. In some species, males invest more time and resources in their offspring than females. This reversal in sex roles is reflected in reversed sexual dimorphism. Both male sticklebacks and South American daddy longlegs (*Zygopachylus albomarginis*) construct a nest for their young, and in pipefish, males instead of females brood the young. Female pipefish are the showy sex, and males prefer to mate with females with the brightest coloration. Seahorse (*Hippocampus*) males have a special brood pouch and even a placenta to provide the growing young with nutrients. The American jacana (*Jacana spinosa*) is rare among birds because the male exclusively cares for the young and females fight with each other over males. Consequently, in contrast with most birds, female jacanas have a larger body size than males, and even possess weapons in the form of wing spurs, which are used in female-female combat.

Many male insects provide food items (nuptial gifts) to the female, which she invests in egg production. In some katydid species, females fight for access to males and their gifts, and males prefer to mate with the largest female because she will lay more eggs. In some empidid dance flies, males seduce females with a nuptial gift and only females are

ornamented. Females deceive males by inflating air sacs on their abdomen to appear larger and, therefore, more attractive.

Sex determination and sex change. For many animals, sex is determined at conception by the sex chromosomes that they inherit from each parent. In humans and most other mammals, daughters inherit two X chromosomes, one from their mother and one from their father. Sons inherit an X chromosome from the mother and a Y chromosome from the father. Conversely, male birds and butterflies inherit a Z chromosome from their mother and a Z from their father, while females receive a Z from their father and a W from their mother. These sex-chromosome differences underlie all other aspects of sexual dimorphism between males and females.

One potential problem is that males and females have different numbers of X (or Z) chromosomes. Sex chromosome–linked genes will therefore be expressed at different levels in the two sexes (for example, twice as much in XX females), which can cause developmental disruption. This problem is solved through dosage compensation, a regulating mechanism ensuring that genes are equally expressed in males and females.

Many animal species have a default sex regulated by the sex chromosomes. For example, in humans the default sex is female, and male development is initiated by testosterone production during embryonic development. In contrast, the default sex in the fruit fly (*Drosophila melanogaster*) is male. The sexual phenotype is regulated by a genetic switch that controls other genes by turning them on or off; when this switch is turned on, the fly develops as a female.

More rarely, sex is determined by environmental conditions experienced during development (environmental sex determination). For example, in the Atlantic silverside fish, eggs laid early in the season become females, and eggs laid later, males. This is because individuals laid early have more time to grow, and large size is favored in females since they can produce more eggs. The peanut worm (*Bonellia viridis*) shows extreme sexual size dimorphism that is associated with environmental sex determination. Larvae that settle on the muddy bottom become females and burrow in the mud, feeding through a long proboscis. A larva that settles on or near a female becomes a dwarf male one-thousandth of her size and lives in her proboscis until he matures and makes a one-way journey to fertilize her eggs. Temperature even regulates sexual phenotype in some animals that have sex chromosomes, including alligators and brush turkeys (*Alectura lathami*).

Not all animals have fixed reproductive roles. Some start life as one sex and then change to the opposite sex later (sequential hermaphrodites). The clownfish (*Amphiprion akallopisos*), which lives in anemones on coral reefs, begins life as a small male. As the fish grows, it becomes a female. This sex change is favorable since size is unimportant for male mating success whereas bigger females lay more eggs. In contrast, the blue-headed wrasse (*Thalassoma bifasciatum*) male defends a territory, and only large males are successful territory holders. These fish start life as females and become males when they are large enough to defend a territory and compete successfully for mates. *See* SEX DETERMINATION.

Males and females tend to share most of their genes, yet many genes are sexually antagonistic as they benefit one sex while being costly for the other. For example, bright coloration is advantageous for male guppies in obtaining mates. In females, conspicuousness would be costly due to high predation risk, and does not confer any benefit in terms of mate attraction. One solution to the problem of sexually antagonistic genes is sex-limited expression, in which the genes are expressed only in the sex in which the trait is beneficial. One way of achieving this is to place such genes on the sex chromosomes. For example, in guppies most genes for body coloration are present only on the male sex chromosome (Y) and, therefore, are never expressed in females. *See* SEX-LINKED INHERITANCE.

Sexual dimorphism is seen throughout the animal and plant kingdoms, with males and females differing in traits ranging from gamete size to ornaments, weapons, and behavior. The driving force behind this dimorphism is the different investment made by males and females in mate attraction and offspring production.

Alison M. Dunn; Nina Wedell

Bibliography. M. Andersson, *Sexual Selection*, Princeton University Press, 1994; J. J. Bull, *Evolution of Sex Determining Mechanisms*, Benjamin-Cummings, California, 1983; T. H. Clutton-Brock, *The Evolution of Parental Care*, Princeton University Press, 1991; J. R. Krebs and N. B. Davies, *An Introduction to Behavioural Ecology*, 3d ed., Blackwell Science, Oxford, 1993; R. H. Peters, *The Ecological Implications of Body Size*, Cambridge University Press, 1986.

Sexually transmitted diseases

Infections that are acquired and transmitted by sexual contact. Although virtually any infection may be transmitted during intimate contact, the term sexually transmitted disease is restricted to conditions that are largely dependent on sexual contact for their transmission and propagation in a population. The term venereal disease is literally synonymous with sexually transmitted disease, but traditionally is associated with only five long-recognized diseases (syphilis, gonorrhea, chancroid, lymphogranuloma venereum, and donovanosis) and sometimes is considered to have pejorative connotations. Sexually transmitted diseases occasionally are acquired nonsexually (for example, by newborn infants from their mothers, or by clinical or laboratory personnel handling pathogenic organisms or infected secretions), but in adults they are virtually

never acquired by contact with contaminated intermediaries such as towels, toilet seats, or bathing facilities. However, some sexually transmitted infections (such as human immunodeficiency virus infection, viral hepatitis, and cytomegalovirus infection) are transmitted primarily by sexual contact in some settings and by nonsexual means in others.

Incidence and epidemiology. The incidences of most sexually transmitted diseases increased in industrialized countries beginning in the early 1960s. During the same period, sexually transmitted infections and their complications also increased in most developing nations, where they constitute some of the major causes of morbidity and contribute importantly to socioeconomic deprivation. In the United States, the reported incidence of gonorrhea began to decline slowly after 1975, due in part to national control measures instituted early in the decade. Although this trend accelerated in the 1980s, the incidence fell more rapidly in some segments of the population than others. The incidence of syphilis also fell, but began to rise rapidly once again after 1985. The decline in syphilis in the early to mid-1980s was due largely to falling rates in homosexually active men, a population that underwent major changes in sexual behavior in response to the epidemic of acquired immunodeficiency syndrome (AIDS). The rapid rise in the late 1980s occurred primarily in urban heterosexual populations with high rates of substance abuse, poverty, and prostitution. The rise in reported rates of *Chlamydia trachomatis* infection since the mid-1980s is due primarily to increasing case detection and reporting. The true incidence of chlamydial infection probably was relatively stable in the 1980s and early 1990s, at incidence rates five- to tenfold higher than those of gonorrhea.

Several behavioral, demographic, and biologic factors have fueled the epidemic of sexually transmitted diseases in industrialized countries. The "sexual revolution" that began in the 1960s and extended into the 1980s was associated with an increasing frequency of sexual activity with larger numbers of partners, especially for young women, and with increasing reliance on the birth control pill for contraception, at the expense of barrier methods that provided partial protection against sexually transmitted diseases. The maturing of the postwar "baby boom" generation resulted in unprecedented numbers of persons aged 18 to 35 years, the time of maximal sexual activity. Changing patterns of marriage and divorce, increasing urbanization, and other social factors abetted these demographic and behavioral changes and contributed to the epidemic. In developing nations, the social disruption associated with conversion from tribal-rural to industrialized-urban societies, economic deprivation, high prevalences of prostitution, and poor diagnostic and treatment facilities have contributed to the epidemic.

Biological factors influencing the sexually transmitted disease epidemic are related both to the sexually transmitted pathogens themselves and to changes in the human hosts. For example, increasing antimicrobial resistance among some organisms, notably *Neisseria gonorrhoeae* and *Haemophilis ducreyi*, has reduced the reliability of traditional inexpensive treatments, and waxing and waning of specific strains of *N. gonorrhoeae* has resulted in differing clinical manifestations and patterns of transmission of gonorrhea. The average duration of infection is another biological factor (partly influenced by behavioral and societal factors, such as access to health care) that markedly influences the epidemiology of sexually transmitted diseases. Infections with most sexually transmitted viruses, such as herpes simplex virus or human papillomavirus, are incurable, persist indefinitely, and remain transmissible for many years, often for life. Therefore, low rates of sexual partner change are sufficient to sustain infection, and these infections are highly prevalent in all segments of the population. In contrast, the mean interval from acquisition to treatment or to spontaneous cure of gonorrhea or chancroid is measured in days to weeks, so that frequent new sexual partnerships are required to sustain infection in the community. Thus, the classical bacterial sexually transmitted diseases are largely limited to small subsets of the population with especially high rates of partner change. The clinical manifestations of *C. trachomatis* infections are subtle, so that the mean duration of infectivity is longer than that of gonorrhea or chancroid; however, it is shorter than that of the viral infections. Thus, intermediate rates of partner change are required to sustain the prevalence of infection, and chlamydial infections are more widely distributed in society than gonorrhea but less so than genital herpes or human papillomavirus infection.

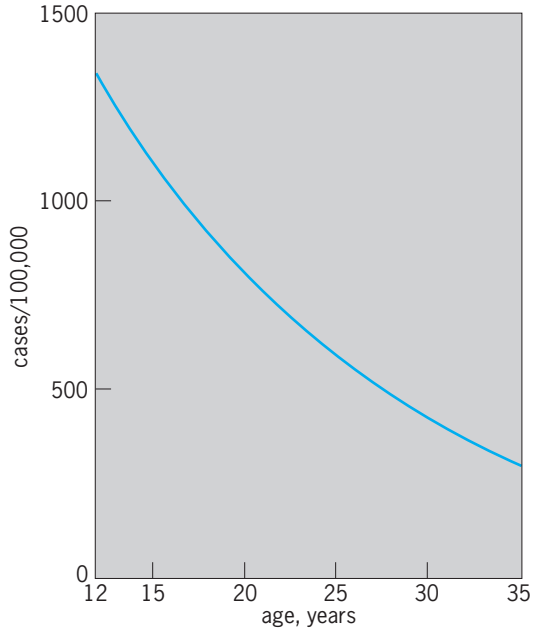
Some populations may be more susceptible to some sexually transmitted diseases than in the past. For example, at one time most young adults had partially protective antibodies to the herpes simplex viruses, but the majority of persons in North America now lack such immunity when they reach sexual maturity. This reduced immunity may partly explain the emergence of genital herpes as one of the most common sexually transmitted infections. By contrast, genital herpes is less common in developing countries, where oral herpes infection in childhood is almost universal. A similar phenomenon may have contributed to the increasing incidence of sexually transmitted infections with cytomegalovirus, hepatitis B virus, human papillomavirus, and molluscum contagiosum.

Although sexually transmitted diseases are common in all populations, in North America and Europe they are more prevalent in urban than in rural areas; in blacks, Hispanics, and native Americans than in whites and persons of Asian ethnicity; and in lower than in higher socioeconomic and educational strata. In the United States, the incidence ratio of gonorrhea in Blacks compared with that in Whites rose from 10:1 in the early 1980s to 25:1 in 1989. A similar trend was observed for syphilis. Minority populations are especially likely to seek care

at public sexually transmitted disease clinics, where case ascertainment and reporting are greater than in the private health care sector. However, this probably explains only a small proportion of the difference in incidence between racial groups. These contrasts reflect cultural differences in sexual behavior, in real or perceived access to health care, and in the behavioral response to early or mild symptoms of infection. Such behaviors in turn are related to social disruption, poor education, anonymity, and family instability. In addition, the abuse of alcohol and illicit drugs has long influenced the risk of sexually transmitted infections by affecting the selection of sexual partners and because substance abuse is closely linked with prostitution.

Although the rates of sexually transmitted diseases are higher for men than women, for most infections women are more likely than men to suffer complications. Moreover, in their early stages many sexually transmitted diseases more commonly are asymptomatic or cause nonspecific symptoms and signs in women than in men, fostering delays in seeking medical care, underdiagnosis, and inadequate treatment. Sexually active adolescent girls have dramatically higher rates of most sexually transmitted diseases and their complications than older women (see **illus.**), due to enhanced behavioral risks and perhaps to age-related physiologic differences.

There are important links between AIDS and the other sexually transmitted diseases. The sexual transmission of the human immunodeficiency viruses is enhanced in the presence of genital ulcers, such as those caused by chancroid, syphilis, and genital herpes. Preliminary data suggest that chlamydial infection, gonorrhea, and trichomoniasis may have similar effects. It is probable that any inflammatory condition of the genitals enhances the efficiency of HIV transmission, and that the wide variation in the epidemiology of HIV infection in heterosexual population around the world is largely due to differences in the frequency of other sexually transmitted infections. The clinical course and response to treatment of some sexually transmitted diseases (for example, syphilis and genital herpes) are adversely affected by human immunodeficiency virus infection.



Age-related incidence of hospitalized cases of acute pelvic inflammatory disease in sexually experienced women in the United States. (*From K. K. Holmes et al., eds., Sexually Transmitted Diseases, McGraw-Hill, 1990*)

Sexually transmitted diseases are not uncommon in children. In newborns they are acquired before birth or from an infected birth canal during delivery. From the newborn period to 1 year of age, sexually transmitted infections usually are acquired by nonsexual contact with an infected mother or by sharing a bed with an infected adult, often in a setting of poor hygiene. From age 1 year to puberty, most cases are the result of sexual abuse.

Classification. The sexually transmitted diseases may be classified in the traditional fashion, according to the causative pathogenic organisms, as follows:

Bacteria
- *Chlamydia trachomatis*
- *Neisseria gonorrhoeae*
- *Treponema pallidum*
- *Mycoplasma genitalium*
- *Mycoplasma hominis*
- *Ureaplasma urealyticum*
- *Haemophilis ducreyi*
- *Calymmatobacterium granulomatis*
- *Salmonella* species
- *Shigella* species
- *Campylobacter* species

Viruses
- Human immunodeficiency viruses (types 1 and 2)
- Herpes simplex viruses (types 1 and 2)
- Hepatitis viruses B, C, D
- Cytomegalovirus
- Human papillomaviruses
- Molluscum contagiosum virus
- Kaposi sarcoma virus

Protozoa
- *Trichomonas vaginalis*
- *Entamoeba histolytica*
- *Giardia lamblia*
- *Cryptosporidium* and related species

Ectoparasites
- *Phthirus pubis* (pubic louse)
- *Sarcoptes scabiei* (scabies mite)

Sexually transmitted diseases may also be classified according to clinical syndromes and complications that are caused by one or more pathogens as follows:

1. Acquired immunodeficiency syndrome (AIDS) and related conditions

2. Pelvic inflammatory disease
3. Female infertility
4. Ectopic pregnancy
5. Fetal and neonatal infections (conjunctivitis, pneumonia, pharyngeal infection, encephalitis, cognitive impairment, physical anomalies, deafness, immunodeficiency, death)
6. Complications of pregnancy (spontaneous abortion, premature labor, premature rupture of fetal membranes, chorioamnionitis, postpartum endometritis)
7. Neoplasia (cervical dysplasia and carcinoma, Kaposi's sarcoma, hepatocellular carcinoma, squamous cell carcinomas of the anus, vulva, and penis)
8. Human papillomavirus and genital warts
9. Genital ulcer–inguinal lymphadenopathy syndromes
10. Lower genital tract infection in women (cervicitis, urethritis, vaginal infection)
11. Viral hepatitis and cirrhosis no12
12. Urethritis in men
13. Late syphilis
14. Epididymitis
15. Gastrointestinal infections (proctitis, enteritis, proctocolitis, enterocolitis)
16. Acute arthritis no17
17. Mononucleosis syndromes no18
18. Molluscum contagiosum
19. Ectoparasite infestation (scabies, pubic lice)

Most of these syndromes may be caused by more than one organism, often in conjunction with nonsexually transmitted pathogens. They are listed in the approximate order of their public health impact, and it is noteworthy that the first half of the list is dominated by conditions that predominantly affect women. The remainder of this article considers the most important sexually transmitted diseases according to both the etiologic and syndromic classifications.

Bacteria. Bacteria cause the best-known and most frequently recognized sexually transmitted diseases, as well as several other sexually transmitted infections.

Neisseria gonorrhoeae. *Neisseria gonorrhoeae*, or the gonococcus, is the cause of gonorrhea, one of the longest-recognized sexually transmitted diseases. The incidence of gonorrhea in most industrialized countries peaked in the early to mid-1970s, then began to decline. Nevertheless, the disease persists at epidemic levels in most areas; in the United States, an estimated 600,000–800,000 cases occur annually. Increasing resistance of *N. gonorrhoeae* to penicillin and other antibiotics has hampered control efforts, especially in developing countries, where the expense of alternative treatments limits their availability. Cephalosporin antibiotics (for example, ceftriaxone or cefixime) or fluoroquinolones (for example, ofloxacin or ciprofloxacin) usually are used to treat gonorrhea in industrialized countries. *See* DRUG RESISTANCE.

Treponema pallidum. *Treponema pallidum* is the cause of syphilis, which has been recognized since antiquity. An inability to cultivate *T. pallidum* is an important impediment to understanding the pathogenesis of syphilis and to its prevention. Syphilis is a protean disease known historically for its ability to mimic many other conditions. Penicillin is highly effective against *T. pallidum* and remains the treatment of choice in all stages of the disease. However, even penicillin does not cure all individuals, especially those with underlying immunodeficiency due to AIDS or other conditions. The changing epidemiology of syphilis is a prime example of the influence that demographic, social, and behavioral factors can have on an infectious disease despite the ready availability of effective antibiotic therapy. *See* SYPHILIS.

Chlamydia trachomatis. *Chlamydia trachomatis* is a small bacterium that, like viruses, requires cell culture techniques for isolation. Some serovariants (L_1, L_2, and L_3) cause the classical sexually transmitted disease lymphogranuloma venereum (LGV), now uncommon in North America and Europe. Infection with the non-LGV serovariants of *C. trachomatis* is the most common bacterial sexually transmitted disease in industrialized countries. During the early 1990s, an estimated 3–4 million cases occurred annually in the United States. *Chlamydia trachomatis* is the most common cause of pelvic inflammatory disease and female infertility in industralized countries.

Several factors have inhibited control of chlamydial infections, including their insidious nature and the subtlety of early signs and symptoms; the lack of widespread screening programs for high-risk individuals; the lack of effective single-dose antibiotic therapy (in contrast to gonorrhea and syphilis); and inadequate knowledge among physicians, health authorities, and the public about the nature, extent, and implications of the chlamydia epidemic. The necessity for cell culture testing for chlamydial infection, an expensive and technically demanding procedure, contributed to these problems. The development of diagnostic tests based on genetic amplification methodology, such as the polymerase chain reaction and the ligase chain reaction, is an important advance that has greatly enhanced the diagnosis, prevention, and control of chlamydial infections. Chlamydial infections respond to treatment with the macrolide or tetracycline classes of antibiotics, such as azithromycin and doxycycline, respectively; penicillin is not effective.

Genital mycoplasmas. The genital mycoplasmas, *Ureaplasma urealyticum* and *Mycoplasma hominis*, are cell wall–free bacteria that reside in the urethra or vagina of most sexually active men and women; over 50% of persons with five or more lifetime sexual partners are colonized with either or both organisms. Sexual contact may not be the sole means of transmission, since some virgin women harbor these organisms. Although most infections are not associated with clinical disease, 20–30% of nongonococcal

urethritis cases in men may be caused by *U. urealyticum*. This organism also has been linked with low birth weight and perhaps premature delivery of infants born to infected mothers. *Mycoplasma hominis* may contribute to some cases of acute pelvic inflammatory disease and bacterial vaginosis. *See* MYCOPLASMAS.

Haemophilus ducreyi. *Haemophilus ducreyi* is the cause of chancroid, or soft chancre, one of the five classical sexually transmitted diseases. It is characterized by painful genital ulcerations, often accompanied by tender enlarged regional lymph nodes. Chancroid is uncommon in the United States, but the incidence rose rapidly from the early 1980s, when several hundred cases were reported annually, to 1989, when 5191 cases were reported; the incidence then declined once again, to fewer than 1,000 cases per year during the early 1990s. By contrast, chancroid is among the most common sexually transmitted diseases in many developing countries. In both industrialized and developing countries, chancroid is more closely associated with prostitution and illicit drug use than any other sexually transmitted infections. Several antibiotics are effective in treatment, including azithromycin, erythromycin, ceftriaxone, and ciprofloxacin.

Calymmatobacterium granulomatis. *Calymmatobacterium granulomatis* causes donovanosis, or granuloma inguinale, one of the five classical sexually transmitted diseases. It causes morphologically typical ulcerations of the skin on or adjacent to the genitals. Donovanosis is extremely rare in industrialized nations and uncommon in most of the world, except for a few localized areas such as Papua New Guinea and the Indian subcontinent.

Enteric bacteria. Several species of the intestinal pathogens *Shigella, Salmonella*, and *Campylobacter* commonly are transmitted by sexual practices that foster fecal-oral contamination, especially among homosexually active men.

Viruses. Although viruses are among the most common and serious sexually transmitted pathogens, recognition of the importance and public health impact of most viral sexually transmitted diseases is relatively recent.

Human immunodeficiency viruses. The human immunodeficiency viruses types 1 and 2 (HIV-1 and -2) cause AIDS. These viruses are transmitted almost exclusively by sexual contact, by direct exposure of the bloodstream to the virus, or by perinatal exposure of the fetus or newborn to maternal infection. Sexual transmission is enhanced in the presence of sexually transmitted diseases, especially those that cause genital ulceration. *See* ACQUIRED IMMUNE DEFICIENCY SYNDROME (AIDS).

Herpes simplex viruses. Genital herpes, caused by herpes simplex virus type 2 and less commonly by the type 1 virus, is among the most common and rapidly increasing sexually transmitted diseases. About 24% of the population of the United States becomes infected with herpes simplex virus type 2 by age 35. Most genital herpes infections are asymptomatic or cause trivial symptoms whose significance is misunderstood by the individual and physician alike. Nonetheless, genital herpes is a serious public health problem due to its potential to cause devastating infections in infants born to infected mothers and to the psychological stress that often accompanies infection. Acyclovir and related antiviral drugs can control the clinical manifestations of infection but are not curative. *See* HERPES.

Hepatitis viruses. Hepatitis B virus is transmitted by direct contact with infected blood or body secretions, as is hepatitus D virus (formerly delta agent), which causes disease only in the presence of hepatitis B infection. Hepatitis C virus, an important cause of viral hepatitis due to neither hepatitis A nor B, is bloodborne and probably is sexually transmitted as well, albeit less efficiently than hepatitis B virus. Routine screening of blood donors has markedly reduced the risk of transfusion-acquired hepatitis B, leaving sexual contact the most common mode of transmission. Sexual contact also plays a distinct but minor role in transmission of the hepatitis A virus. Immunization with hepatitis B vaccine is an effective but underutilized prevention measure. *See* HEPATITIS.

Cytomegalovirus. Cytomegalovirus is an extremely common herpes group virus that ultimately infects 50–90% of persons in industrialized nations and 90–100% of those in developing countries. Sexual contact is an important mode of transmission of cytomegalovirus in young adults in some industrialized societies, but nonsexual transmission is also common. Although cytomegalovirus infection usually is asymptomatic in healthy adults, it is nevertheless a major cause of congenital infection that may result in miscarriage, malformation, mental retardation, or deafness. The virus causes lifelong latent infections, and life-threatening disease due to reactivation of asymptomatic cytomegalovirus infection is common in persons with impaired immune function due to AIDS, cancer, and other conditions. *See* CYTOMEGALOVIRUS INFECTION; EPSTEIN-BARR VIRUS.

Human papillomaviruses. The human papillomaviruses, of which there are more than 50 distinct varieties, cause warts. Most types cause common warts of the hands, feet, or other parts of the body and are not sexually transmitted. Most people with genital human papillomavirus infection are asymptomatic and have no visible manifestation, but human papillomavirus infection may be the most common of all sexually transmitted infections. Human papillomavirus types 6 and 11 cause most cases of clinically apparent genital warts. By contrast, types 16, 18, and several others usually do not cause visible warts, but are closely linked with cancer of the cervix, anus, vulua, and penis. While tobacco use and other (largely unknown) factors influence the development of malignancy, the weight of evidence suggests that cervical cancer is a sexually transmitted disease of which human papillomavirus is the direct cause. Clinically apparent warts can be ablated by freezing or by chemical irritants, but no curative treatment exists.

Molluscum contagiosum. The molluscum contagiosum virus causes characteristic painless wartlike skin lesions that have no known serious sequelae. Seen most commonly in young children, who acquire the disease by nonsexual contact with other children, molluscum contagiosum in adults usually involves the genital region and is transmitted sexually.

Kaposi's sarcoma. The epidemiology of Kaposi's sarcoma, a malignancy that is frequent in homosexual men with AIDS but less common in other individuals with AIDS, suggested that it might be caused by a sexually transmitted pathogen other than human immunodeficiency virus.

Protozoa. *Trichomonas vaginalis* is a flagellated protozoon that causes vaginal inflammation and discharge, but asymptomatic vaginal colonization also is common. Sexual contact is the primary mode of acquisition. Most infections of the male genital tract are asymptomatic, but urethritis may occur. The infection is readily treated with metronidazole or related antibiotics. Complications of trichomoniasis are rare. *Giardia lamblia, Entamoeba histolytica*, and *Cryptosporidium* are intestinal protozoa that are most commonly acquired by ingestion of the organisms' cysts in contaminated water. As for intestinal bacteria, however, they commonly are sexually transmitted, especially among homosexually active men. *See* PROTOZOA.

Ectoparasites. *Phthirus pubis*, the pubic (crab) louse, is slowly mobile, survives only for a few hours away from a human host, and usually requires direct genital apposition for transmission. *Sarcoptes scabiei*, the itch mite, is also transmitted by direct skin-to-skin contact. Sex is an important but not an exclusive mechanism of spread of both intestations.

Clinical syndromes. Sexually transmitted diseases result in a variety of clinical syndromes. The major syndromes are reviewed below except for AIDS and syphilis, which are discussed in separate articles.

Pelvic inflammatory disease. Acute pelvic inflammatory disease, or salpingitis, is the result of infection of the Fallopian tubes (salpinges) with bacteria that ascend from the cervix or vagina, in the absence of recent childbirth or gynecological surgery. Salpingitis is often accompanied by uterine inflammation (endometritis), peritonitis, or—in its most advanced form—formation of abscesses involving the Fallopian tubes, ovaries, or peritoneal cavity. Pelvic inflammatory disease and its complications constitute one of the most important health problems of women. It is estimated that 85–90% of cases are sexually acquired. Up to 1,000,000 cases are estimated to occur annually in the United States. In North America and Europe, 10–20% of acute pelvic inflammatory disease episodes (depending on the bacterial etiology and clinical severity) result in infertility due to bilateral scarring and obstruction of the Fallopian tubes; the rate undoubtedly is higher in socioeconomically disadvantaged women who lack ready access to health care. In addition to infertility, the major complications of the disease are ectopic (tubal) pregnancy, chronic pelvic pain, and recurrence of salpingitis.

Among sexually experienced women, the risk of pelvic inflammatory disease is inversely correlated with age (see illus.). Use of certain intrauterine devices for contraception results in a two- to ninefold enhanced risk, and accounts for many nonsexually acquired cases of pelvic inflammatory disease. In contrast, the use of hormonal contraception reduces the risk by about 50%. A prior history of pelvic inflammatory disease markedly increases the risk. Moreover, recurrent episodes enhance the risk of long-term complications, and up to 75% of women with three or more episodes become infertile. Vaginal douching is a strong independent risk factor for pelvic inflammatory disease. As a result, douching is strongly discouraged for either so-called feminine hygiene or treatment of vaginal infection.

The most common causes of pelvic inflammatory disease are gonorrhea and chlamydial infection, but the relative contribution of each varies widely. Historically, *Neisseria gonorrhoeae* was long considered the primary cause. However, gonorrhea accounts for only 15–40% of cases in much of North America and Europe, where 20–50% are associated with chlamydial infection. Many women with the syndrome are infected with both of these organisms. Other bacteria commonly contribute to salpingitis, either alone or in combination with *Chlamydia trachomatis* or *N. gonorrhoeae;* these include *Mycoplasma hominis* and several aerobic and anaerobic species of the normal bacterial microflora of the vagina.

The clinical hallmarks of pelvic inflammatory disease are low abdominal pain and tenderness of the uterus, Fallopian tubes, and ovaries, elicited during bimanual pelvic examination. These commonly are accompanied by vaginal discharge or abnormal uterine bleeding. Nausea, vomiting, or fever commonly occur in severe cases. However, mild and even asymptomatic cases are common and cause many cases of infertility and ectopic pregnancy. Clinical management of pelvic inflammatory disease includes treatment with antibiotics, and occasionally surgery to drain or remove abscesses. Because of the broad spectrum of bacteria involved, two or more antibiotics typically are used in combination, such as ofloxacin with metronidazole or ceftriaxone with doxycycline. Correction of reversible predisposing factors also is important (such as removal of an intrauterine device and examination and treatment of sexual partners).

Infertility. The incidence of female infertility has risen in parallel with the sexually transmitted disease epidemic, and it is estimated that 100,000 or more women in the United States become infertile each year as the result of sexually transmitted infections. In localized parts of Africa, up to half of the women of childbearing age are estimated to be infertile because of prior gonococcal pelvic inflammatory disease. In the industrialized countries, about half of all female infertility is the result of tubal scarring or obstruction. About 75% of those

with tubal infertility have antibodies to *C. trachomatis*, indicating past chlamydial infection, compared with 0–30% of fertile women and those with other forms of infertility. Chlamydial infection may be the most common preventable cause of female infertility in industrialized countries. Most of the women with chlamydia-associated tubal obstruction lack past histories of overt pelvic inflammatory disease or unexplained abdominal pain, indicating that asymptomatic chlamydial infection can cause sufficient salpingitis to result in infertility. The treatments for tubal infertility (surgical reconstruction and in vitro fertilization) are expensive and largely unsatisfactory. Although infertility in men can result from testicular inflammation (epididymitis) due to *N. gonorrhoeae* or *C. trachomatis*, sexually transmitted disease apparently is not a common cause of male infertility.

Ectopic pregnancy. Ectopic (tubal) pregnancy occurs when an embryo implants in the Fallopian tube rather than the uterus; growth of the embryo and placenta ultimately causes tubal rupture and internal bleeding, a potentially fatal medical emergency. Ectopic pregnancy, like tubal infertility, usually is the result of Fallopian tube scarring. The incidence of ectopic pregnancy rose in parallel with the sexually transmitted disease epidemic, and the risk of ectopic pregnancy is increased about tenfold after the occurrence of pelvic inflammatory disease. *See* PREGNANCY.

Fetal and perinatal infections. Most sexually transmitted pathogens are transmissible either to the unborn fetus or to the newborn during or immediately after birth to an infected mother, with effects that range from trivial to devastating. The most dangerous sexually transmitted diseases for the newborn are human immunodeficiency virus infection, herpes, and syphilis. Neonatal herpes is most common when the mother experiences her first infection with herpes simplex virus immediately prior to delivery, and commonly results in encephalitis, hepatitis, and multiple organ failure. The mortality is about 50%, and most surviving infants suffer severe neurological and developmental impairment. It is estimated that several thousand cases occur annually in the United States.

Reported cases of congenital syphilis in infants under 1 year of age in the United States rose from 107 in 1980 to almost 700 in 1989 more than in any year since the 1940s. Neonatal human immunodeficiency virus infection rose similarly during this period. Both occur most frequently in socioeconomically disadvantaged populations and are especially common in some developing countries.

Cytomegalovirus is a common cause of congenital or perinatal infection. Although most cases are benign, several thousand clinically significant cases occur annually in the United States. In its most severe form, which usually follows primary maternal infection during pregnancy, congenital cytomegalovirus infection results in severe neurodevelopmental abnormalities.

Chlamydial infection, acquired during birth to an infected mother, is the most common neonatal sexually transmitted disease in industrialized countries. One to five percent of all newborns in the United States were estimated to be infected in the 1980s, primarily with asymptomatic nasopharyngeal infections or ocular infections (neonatal inclusion conjunctivitis), and less commonly with pneumonia. Undetected perinatal chlamydial infections may contribute to middle ear inflammation (otitis media) and persistent cough due to bronchiolitis in young children.

Gonococcal conjunctival infection (ophthalmia neonatorum) now is rare in industrialized countries but still is common in some locales; delayed treatment often leads to blindness. Neonatal hepatitis B virus infection is a major health problem in most tropical areas and in some populations in industrialized countries. Transmission of genital human papillomavirus infection to the newborn can result in warts of the genitals, perianal regions, or throat (pharyngeal papillomatosis); the latter condition is particularly difficult to treat.

Complications of pregnancy. In addition to infections of the fetus or newborn, many sexually transmitted pathogens adversely affect the course of pregnancy. Most surveys in North America and Europe have shown 5–10% of pregnant women to have chlamydial infection. Similar prevalences of gonorrhea exist in socioeconomically disadvantaged pregnant women in much of the world. Gonorrhea and probably chlamydial infection increase the risk of premature rupture of the fetal membranes, premature delivery, postpartum uterine infections (endometritis), and perhaps miscarriage and stillbirth. Infection of pregnant women with the genital mycoplasmas also may be responsible for some cases of prematurity, low birth weight, and postpartum endometritis. Bacterial vaginosis also is associated with early rupture of the fetal membranes, prematurity, and other complications of labor and delivery.

Neoplasia. For several decades, the risk for squamous cell cancer of the cervix, the most common of all malignancies in men and women, was known to be linked with certain sexual behaviors, including young age at first intercourse and childlessness. Both of these factors now are known to be linked with increased numbers of sexual partners and therefore with sexually transmitted disease risk. It also was shown that the risk of cervical cancer in permanently monogamous women is correlated with their husbands' lifetime sexual partners. Despite their relative youth, women attending sexually transmitted disease clinics have a higher prevalence of cervical cancer and its precursors than any other group of women that has been surveyed.

Because of such epidemiologic observations, at various times gonorrhea, trichomoniasis, and syphilis were postulated as causes of cervical cancer, but they clearly play no direct roles. There are conflicting data regarding herpes simplex virus, but this agent probably has no primary role. In contrast, a growing body of data suggests that some strains

Lower genital tract infections in women

Syndrome	Sites affected	Usual causes
Cervicitis	Uterine cervix	*Chlamydia trachomatis* *Neisseria gonorrhoeae* Herpes simplex virus Other or unknown
Vaginal infection	Vagina, labia	*Candida albicans* and other yeasts *Trichomonas vaginalis* Herpes simplex virus *Gardnerella vaginalis* with vaginal anaerobic bacteria Foreign body Other
Urethritis-cystitis	Urethra, bladder	*Escherichia coli* and other coliform bacteria *Staphylococcus saprophyticus* *Chlamydia trachomatis* *Neisseria gonorrhoeae* Herpes simplex virus

of human papillomavirus directly cause precancerous changes (dysplasia) and overt cancer. All women with past histories of sexually transmitted disease are particularly encouraged to undergo a cervical cancer screening examination (Papanicolaou test) annually.

Other malignancies also have been linked with sexually transmitted diseases, though generally less strongly than has cervical cancer. Homosexual men are unusually susceptible to anal cancer, probably due to human papillomavirus infection. Chronic carriers of hepatitis B virus have an enhanced risk of liver cancer (hepatocellular carcinoma). Cancer of the skin of the vulva and penis also is related to human papillomavirus. Kaposi's sarcoma, believed to be a malignancy of lymphatic epithelial cells, may be due to a sexually transmitted herpes group virus. *See* CANCER (MEDICINE).

Genital warts. Warts of the genitals, anus, or rectum (condyloma acuminatum) are a common and rapidly growing problem among sexually active individuals in North America and Europe. Although genital warts were long considered a benign (albeit unpleasant and inconvenient) health problem, the association of some strains of human papillomavirus with cancer has elevated the perceived public health impact of this infection. Most affected tissues are not visibly abnormal, and in most patients the true extent of infection is substantially greater than indicated by overt warts. Partly for this reason, eradication of genital warts often is difficult, and no treatment has been shown to eradicate the virus or to affect the risk of later malignancy.

Genital ulcer/inguinal lymphadenopathy syndromes. Several sexually transmitted infections cause ulcers or other lesions of the skin or mucous membranes of the genitals, tenderness and swelling of the lymph nodes of the groin (inguinal lymphadenopathy), or both. Primary syphilis is a common cause throughout the world, but the frequencies of the other diseases have great geographic or demographic variation. Genital herpes is by far the most common cause of genital ulcer disease in most of Europe and North America, where chancroid is rare; the reverse is true in most tropical settings. Lymphogranuloma venereum and donovanosis are the other major causes; both are rare in the United States.

Lower genital tract infections in women. Lower genital tract infection in women refers to cervical infection (cervicitis), vaginal infection, and infection of the bladder and urethra (cystitis-urethritis). These syndromes cause various combinations of vaginal discharge, vulvar pain or itching, and urinary symptoms such as painful, frequent, or urgent urination. Most lower genital tract infections are clinically mild. However, cervicitis and some vaginal infections predispose to pelvic inflammatory disease, and cystitis-urethritis can lead to serious infections of kidneys (pyelonephritis). Each syndrome may be caused by sexually transmitted infections or by nonsexually transmitted agents (see **table**).

Cervicitis is the female counterpart to urethritis in men; it usually results in discharge from the vagina or painful urination (dysuria), but commonly is asymptomatic. *Chlamydia trachomatis* and *Neisseria gonorrhoeae* are the most common causes; the proportion caused by each varies geographically and with other demographic factors, along the lines described above for pelvic inflammatory disease. A few percent of cases are due to herpes simplex virus, usually in conjunction with overt genital herpes. The etiology cannot be defined in up to one-half of the cases; the role of sexual transmission in these cases is uncertain.

There are three major causes of vaginal infection. Vulvovaginitis due to various yeast-type fungi (usually *Candida albicans*) affects the majority of women at one time or another, usually causing scant discharge and prominent vulvar irritation. It is not sexually transmitted.

Bacterial vaginosis usually causes a nonirritating white or gray vaginal discharge, often with a distinctive fishy odor. It results from overgrowth of various normal vaginal bacteria, especially anaerobic organisms (which cause the distinctive odor), *Gardnerella vaginalis*, the genital mycoplasmas, and *Mobiluncus curtisii*, and by depletion of *Lactobacillus* species, especially strains that produce hydrogen peroxide.

However, the precise cause of bacterial vaginosis is unknown. The syndrome occurs most frequently in women at risk for sexually transmitted diseases, but a sexually acquired pathogen has not been demonstrated. Long considered a benign condition, bacterial vaginosis is known to predispose individuals to potentially serious uterine infections (endometritis) after childbirth or gynecologic surgery, to early rupture of the fetal membranes and premature labor, and perhaps to pelvic inflammatory disease. Antibiotics with activity against anaerobic bacteria, such as clindamycin or metronidazole, are effective either by mouth or by intravaginal administration. No male counterpart to bacterial vaginosis has been defined, and treatment of the individual's sexual partners does not improve, cure, or prevent recurrence of the syndrome.

Trichomonas vaginalis causes discharge that may have a fishy odor. Trichomoniasis usually is acquired sexually, and the male sexual partner, who usually is asymptomatically infected, should be treated. However, a woman can develop symptoms after carrying the organism for several years, so cases are common in women not currently at risk for a sexually transmitted infection.

Numerous other conditions can result in increased vaginal discharge or irritation. Common causes include chemical irritation or allergy (for example, from commercial douche preparations or vaginal contraceptives); foreign body (such as a forgotten tampon); estrogen-lack ("atrophic") vaginitis following menopause or surgical removal of the ovaries; and increased normal discharge.

Cystitis-urethritis in young women may be due to sexually transmitted disease, especially gonorrhea or chlamydial infection. However, these are less common than nonsexually acquired urinary tract infections, the most common causes of which are *Escherichia coli* and *Staphylococcus saprophyticus*.

Viral hepatitis. These infections and their complications have been especially common in homosexual and bisexual men who have many sexual partners, persons who share equipment used to inject illicit drugs, and their sexual partners. Chronic hepatitis B and C are common causes of cirrhosis and liver failure, and chronic hepatitis B is the primary cause of hepatic carcinoma, one of the most common cancers in developing countries. *See* HEPATITIS.

Urethritis in men. Urethritis (urethral infection) is the most common of all sexually transmitted disease syndromes in men. Urethritis usually is associated with visible discharge of exudate from the penis and often with painful urination or a sense of irritation or itching within the urethra. These symptoms may be mild, and asymptomatic sexually transmitted urethral infections are common. Gonorrhea historically has been the most common cause of urethritis, and remains so in much of the world. However, nongonococcal urethritis, which refers to all urethritis that is not due to gonorrhea, is substantially more common in most populations in North America and Europe. About 30–40% of cases are due to *Chlamydia trachomatis*, 20–30% to *Ureaplasma urealyticum*, and under 5% each to herpes simplex virus and *Trichomonas vaginalis*. Some cases may be due to a newly recognized genital mycoplasma, *Mycoplasma genitalium*. The causes of up to 50% of cases are uncertain, although epidemiologic evidence indicates that these infections also are sexually acquired. About 10–20% of men with urethral gonorrhea in industrialized countries have simultaneous urethral chlamydial infections. Urethritis itself usually is a mild condition; its major significance to public health is as a source of sexually transmitted disease in women.

Epididymitis. Acute epididymitis is infection of the sperm-collecting ducts (epididymis) adjacent to the testis; the testis itself also may be involved (epididymo-orchitis). Epididymitis causes testicular pain and swelling, and results from ascending infection from the urethra via the vas deferens. It is the male counterpart to pelvic inflammatory disease, but is far less common than that syndrome. In young, sexually active men, *N. gonorrhoeae* and *C. trachomatis* are the major causes; in most settings in Europe and North America, chlamydial infection is a much more common cause than gonorrhea. Nonsexually transmitted bacteria, such as *Escherichia coli*, usually are responsible for epididymitis in men over age 35 and those who recently have undergone genitourinary surgery or instrumentation. Bilateral epididymitis can result in sterility. *See* URINARY TRACT DISORDERS.

Gastrointestinal infections. Sexually transmitted infections of the gastrointestinal tract are major health problems among homosexually active men; some of them also are common in heterosexuals, especially women. The most common rectal infections (proctitis) are gonorrhea, chlamydial infection, herpes, and syphilis. Proctitis results from direct inoculation of infectious material into the rectum, either via rectal intercourse or, in many women, through contamination of the anal area by vaginal discharge. Most other sexually transmitted gastrointestinal infections affect the small intestine (enteritis) or the colon (colitis) and are acquired during sexual activity that facilitates fecal-oral contamination. The protozoa *Entamoeba histolytica* and *Giardia lamblia* and the bacteria *Shigella, Salmonella*, and *Campylobacter* are the predominant causes. On a worldwide scale, however, the large majority of such infections are acquired by nonsexual routes.

Arthritis. Gonorrhea occasionally causes an acute arthritis that if untreated can result in permanent joint damage. Chlamydial infection is the most common trigger of Reiter's syndrome, a type of chronic arthritis that occasionally causes severe disability. These two diseases are among the most common causes of acute arthritis in sexually active young adults in the United States. *See* ARTHRITIS.

Mononucleosis. Mononucleosis is an illness characterized by fever, malaise, enlargement of lymph nodes, and often pharyngitis. Mononucleosis due to cytomegalovirus often is sexually acquired. The

more common infectious mononucleosis, due to Epstein-Barr virus, has been called a "kissing disease," reflecting carriage of the virus in saliva and the potential for sexual transmission. *See* INFECTIOUS MONONUCLEOSIS.

Ectoparasite infestation. Scabies and pubic louse infestation are common and cause pruritic skin rashes that usually are nuisances rather than serious threats to health.

Prevention and control. Prevention strategies for communicable diseases are categorized as primary prevention, the prevention of infection as such; secondary prevention, the detection of infection and early treatment, before complications ensue; and tertiary prevention, the treatment of complications and amelioration of long-term consequences. Primary prevention of sexually transmitted diseases includes educating those at risk to select sexual partners at low risk of infection, to use condoms or other barrier contraceptives, and to maintain permanent, monogamous sexual relationships. Secondary prevention emphasizes screening persons at risk to detect subclinical infection, antibiotic treatment, and notification and treatment of the sexual partners of infected persons. Examples of tertiary prevention for sexually transmitted infections include treatment of pelvic inflammatory disease or of infants with neonatal herpes; in vitro fertilization or fallopian tube reconstructive surgery for infertile women following pelvic inflammatory disease; and the provision of institutional care for infants with birth defects due to maternal sexually transmitted infections. Primary and secondary prevention are far more cost-effective than tertiary prevention and are emphasized in most control programs.

The only certain methods for prevention of sexually transmitted disease are sexual abstinence and permanent mutual monogamy. Less stringent limitations on sexuality confer varied levels of protection; examples include avoiding multiple simultaneous partners and foregoing sexual contact with persons suspected to be promiscuous, and with individuals known to have genital or other symptoms suggestive of sexually transmitted disease. The consistent and proper use of either male or female condoms provides a high degree of protection against pathogens transmitted between mucous membranes (for example, *C. trachomatis*, *N. gonorrhoeae*, and the human immunodeficiency viruses). Condoms confer significant but less complete protection against pathogens transmitted between skin surfaces (for example, herpes simplex virus and *T. pallidum*), because not all sexually exposed surfaces are covered. Latex or polyurethane condoms are believed to provide better protection than condoms made of biological membranes. No sexually transmitted pathogen passes through intact latex or polyurethane condoms. Diaphragms and vaginal spermicides may help to prevent some infections in the event of exposure. Intrauterine devices should be avoided by women at high risk for sexually transmitted disease.

Persons at high risk for sexually transmitted diseases should seek periodic examinations to detect asymptomatic infections. Prompt medical attention for appropriate symptoms usually prevents the development of complications. Preventive treatment with an appropriate antibiotic before or immediately after exposure may be effective, but is not practical in most circumstances.

Communitywide control measures include provision of adequate clinical services for diagnosis and treatment of sexually transmitted infections, which usually implies a combination of publicly supported clinics and appropriate training of health care providers in the community. Disease detection by routine screening of individuals at risk is central to control of gonorrhea, syphilis, chlamydial infection, and human immunodeficiency virus infection. Measures to confidentially locate and assure treatment and counseling of the sexual partners of infected persons are important to interrupt chains of transmission. For hepatitis B, and perhaps for other sexually transmitted diseases in the future (such as genital herpes), vaccination programs are valuable. Directed health education, with attention to the preventive measures outlined above, is important. However, educational programs that use scare tactics or whose primary goal is proscription or limitation of sexual activity ("moral prophylaxis") historically have been ineffective. Training of medical students, physicians, and other health care providers in the increasingly complex field of sexually transmitted disease is a crucial measure. The reporting of cases to health authorities is necessary to facilitate the targeting of resources for disease control. Finally, basic and applied research provides the growing knowledge on which all other control measures are based. *See* PUBLIC HEALTH.

H. Hunter Handsfield

Bibliography. A. M. Brandt, *No Magic Bullet: A Social History of Venereal Disease in the United States Since 1890*, 2d ed., 1990; H. H. Handsfield, *Color Atlas and Synopsis of Sexually Transmitted Diseases*, 2d ed., 2000; K. K. Holmes et al. (eds.), *Sexually Transmitted Diseases*, 2d ed., 1989; E. O. Laumann et al., *The Social Organization of Sexuality: Sexual Practices in the United States*, 2000.

Sferics

Electromagnetic radiations produced primarily by lightning strokes from thunderstorms. It is estimated that globally there occur about 2000 thunderstorms at any one time, and that these give rise to about 100 lightning strokes every second. The radiations are short impulses that usually last a few milliseconds, with a frequency content ranging from the low audio well into the gigahertz range. Sferics (short for atmospherics) are easily detected with an ordinary amplitude-modulation (AM) radio tuned to a region between radio stations, especially if there are thunderstorms within a few hundred miles. These

sounds or noises have been identified and characterized with specific names, for example, hiss, pop, click, whistler, and dawn chorus. They fall into what is generally known as radio noise. *See* ATMOSPHERIC ELECTRICITY; ELECTROMAGNETIC RADIATION; LIGHTNING; THUNDERSTORM.

Types of sferics. The various types of sferics include terrestrial, magnetospheric, or Earth-ionospheric.

Terrestrial. Occasionally lightning is associated with other natural phenomena. Violent volcanic eruptions accompanied by phreatic explosions when the molten lava comes into contact with seawater have been observed to produce lightning. The birth of the Icelandic island of Surtsey was accompanied by spectacular lightning displays, as was the eruption on the nearby island of Heimaey. Both eruptions provided a copious source of sferics, as do the somewhat regular eruptions of the Japanese volcano Sakurajima on Kyushu island. Dust storms and dust devils have also been observed to produce sferics. *See* DUST STORM.

Anthropogenic noise is also present in the sferics population: automobile ignition, motor brushes, coronas from high-voltage transmission lines, and various high-current switching devices are among human contributions to radio noise. Most of these noise sources, however, are too small to compete with the amplitude and power of the lightning-generated impulse except in close proximity to the source. One of these noises that does compete with the lightning sferic is the electromagnetic pulse that accompanies the detonation of a hydrogen (fusion-type) bomb. *See* ELECTRICAL NOISE.

Magnetospheric. Lightning-generated sferics are sometimes coupled into the magnetosphere, where they are trapped and guided by the Earth's magnetic field. In this mode, the impulse travels in an ionized region. As a result, the frequencies present in the original impulse are separated by dispersion (the higher frequencies travel faster than the lower) and produce the phenomena known as whistlers. Amplification may occur, and reflection at the ends of the geomagnetic ducts (magnetic North and South poles) result in the return of the sferic to the source. The increased length of the path effectively stretches the original impulse into a train of decreasing audio frequencies sometimes several seconds in duration. Sferics often undergo several reflections in the duct. Studies of whistlers have led to an increased understanding of the magnetosphere and regions of enhanced ionization. *See* MAGNETOSPHERE.

Earth-ionospheric. By far the dominant and most readily observed sferics are the lightning-produced impulses that travel in the spherical cavity formed by the ionosphere and the Earth's surface. Lightning currents produce strong radiation in the very low-frequency band, 3–30 kHz, and in the extremely low-frequency band, 6 Hz–3 kHz. For frequencies above about 3 kHz, the propagation in the Earth–ionosphere cavity for distances up to about 600 mi (1000 km) can be treated as a ground wave and a series of multiple reflections between the Earth and the ionosphere. The farther the observing station from the source, the greater the number of reflections that may reach the receiver. An example of a very low-frequency sferic followed by a series of reflections is given in **Fig. 1**. The initial downward deflection (about one-third the size of the maximum negative deflection) is the radiated electric field from a lightning stroke lowering negative charge to the Earth. This vestige of the ground wave has been strongly attenuated in traveling over the land. The first four ionospheric reflections show much less attenuation. More than 15 reflections are discernible as the sferic propagates in the Earth–ionosphere cavity. The train of high-frequency noise preceding the negative stroke impulse is produced by the downward-moving stepped leader that initiated the lightning flash.

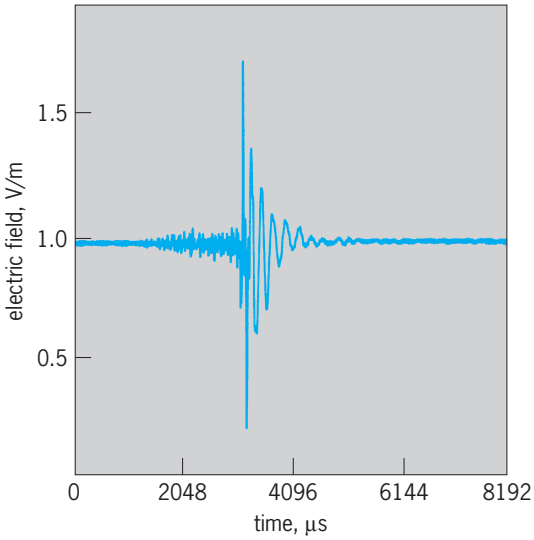


Fig. 1. Sferic that originated 1006 mi (1610 km) from the receiver.

An interesting variant of wave propagation in the Earth–ionosphere cavity is given for the mode $n = 0$. In this mode, known as the Earth–ionosphere waveguide mode, only frequencies less than about 2–3 kHz propagate, so that a lightning sferic with both a strong high- and low-frequency content reaches the receiver with two sets of signals: the normal very low-frequency component, $n = 1$ and greater, via a ground wave and ionospheric reflections, and a set with a slower group velocity (frequencies below 3 kHz) that propagate in the $n = 0$ waveguide mode. The sferic then exhibits a so-called slow tail (**Fig. 2**). The slow tail of this sferic indicates a significant frequency component below about 1 kHz in the source. It appears that strokes that lower positive charge to Earth exhibit a slow tail far more often than the common negative strokes. Positive strokes are a reasonably rare occurrence in summer thunderstorms: about 99% of all strokes to ground lower negative charge to Earth.

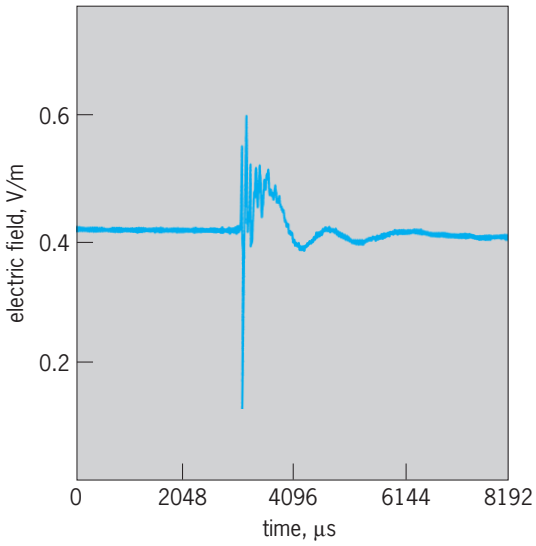


Fig. 2. Sferic exhibiting a slow tail; it originated in a positive lightning stroke 812 mi (1300 km) from the receiver.

The time between the first very low-frequency peak and the peak of the slow tail can be used to estimate the distance to the lightning flash. Similarly, the time between successive ionospheric reflections can also be used to estimate the distance to the source, provided the height of the ionosphere is known. Values for ionospheric height (which change with time of day) are approximately 40 mi (70 km) in daylight and 50 mi (90 km) at night. Although source distances can be estimated by the single-station methods described above, multiple-station measurements are usually preferred.

Lightning discharges are also observed to excite the Earth–ionosphere cavity at its fundamental and higher resonant frequencies, starting at about 8 Hz. These are known as the Schumann resonances. Studies of these resonances complement the studies of sferics at higher frequencies and have provided many interesting details related to specific properties of the ionosphere. *See* IONOSPHERE; RADIO-WAVE PROPAGATION.

Lightning detection. There are two principal methods used to locate lightning sources using sferics. One method uses a pair of loops or coils to detect a voltage induced by the magnetic field of the sferic. When properly calibrated, the coils produce equal voltages for a source located on the bisector of the angle between the loops. Usually, the coils are placed in vertical planes perpendicular to each other. The direction of the arriving signal is obtained from the ratio of the induced voltages in the coils. A value of the magnetic field strength related to the current impulse at the source is obtained by squaring and adding the voltages in the coils. The ambiguity in the direction of arrival of the sferic (from in front or from behind the coils) is resolved by making an independent measurement of the electric field of the sferic. Two or more stations provide two or more directions that intersect at the apparent source of the sferic. Accuracies of about 2° have been claimed for the crossed-loop systems.

The second system in use is based upon a measurement of the time of arrival of the sferic at three or more stations. This system requires excellent timing (microsecond accuracy) and also that there be specific characteristics of the sferic that are identifiable in the signal at each station, so that the arrival time of that particular feature can be estimated accurately. Both the time of arrival and the crossed-loop lightning-direction-finding systems are commercially available. They are used by the weather services and by private industry in many countries throughout the world. *See* STORM DETECTION.

Marx Brook

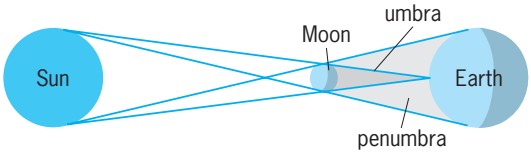
Bibliography. J. Galejs, *Terrestrial Propagation of Long Electromagnetic Waves*, 1972; R. A. Helliwell, *Whistlers and Related Ionospheric Phenomena*, 1965; M. A. Uman, *The Lightning Discharge*, 1987, reprint 2001; H. Volland (ed.), *Handbook of Atmospherics*, 1982; J. R. Wait, *Electromagnetic Waves in Stratified Media*, 1962.

Shadow

A region of darkness caused by the presence of an opaque object interposed between such a region and a source of light. A shadow can be totally dark only in that part called the umbra, in which all parts of the source are screened off. With a point source, the entire shadow consists of an umbra, since there can be no region in which only part of the source is eclipsed. If the source has an appreciable extent, however, there exists a transition surrounding the umbra, called the penumbra, which is illuminated by only part of the source. Depending on what fraction of the source is exposed, the illumination in the penumbra varies from zero at the edge of the full shadow to the maximum where the entire source is exposed. The edge of the umbra is not perfectly sharp, even with an ideal point source, because of the wave character of light. *See* DIFFRACTION.

For example, the shadow of the Moon shown in the **illustration** has an umbra which barely touches the Earth during a total eclipse of the Sun. The eclipse is only partial for an observer situated in the penumbra. Astronauts circling the Earth or the Moon observe this shadow pattern during each orbit. *See* ECLIPSE.

The term shadow is also used with other types of radiation, such as sound or x-rays. In the case of sound, pronounced shadows are formed only for high frequencies. The high frequency of visible light waves makes possible shadows that are nearly black. X-ray photographs are shadowgrams in which the bones and tissues appear by virtue of their different



Shadow of the Moon at the time of an eclipse of the Sun. Relative sizes are not to scale.

opacities or degrees of absorption. The bending of light rays by reflection or refraction may also produce shadow patterns. An important practical example is the schlieren method of observing flow patterns in wind tunnels. *See* RADIOGRAPHY; SCHLIEREN PHOTOGRAPHY. Francis A. Jenkins; William W. Watson

Shadowgraph

An optical method of rendering fluid flow patterns visible by using index-of-refraction differences in the flow. As a simple example, sunlight passing through the air above a warm toaster may be seen to form darker and brighter streaks on a surface a meter or two from the toaster. The method relies on the fact that rays of light bend toward regions of higher refractive index while passing through a transparent material. The fluid is usually illuminated by a parallel beam of light. **Figure 1** depicts the method as it might be applied to a fluid sample undergoing thermal convection between two parallel plates, with the lower plate being kept warmer than the upper one. As illustrated, the rays bend toward the cooler down-flowing regions, where the refractive index is higher, and away from the warmer up-flowing ones. After they have passed through the fluid layer, the rays tend to focus above the cooler regions and defocus above the warmer regions. If an image of the light beam is recorded not too far from the sample, brighter areas of the image will lie above regions of down flow, where the rays have been concentrated, and darker areas will lie above regions of up flow. If an image is recorded at a distance greater than the distance for which the rays actually focus, then it is possible to have a situation in which darker, rather than brighter, areas lie above down-flowing regions. In practice this is not a problem, because the image can be observed beginning near the sample and gradually moving away, until a suitable situation is achieved. Because the light passes completely through the sample, the bending effect for each ray is averaged over the sample thickness. *See* CONVECTION (HEAT); REFRACTION OF WAVES.

In convection experiments the refractive index varies because of thermal expansion of the fluid, but the method is not restricted regarding the mechanism responsible for disturbing the refractive index. Thus the same method may be used to visualize denser and less dense regions in a gas flowing in a wind tunnel, including Mach waves and shock waves, where the denser regions have a higher-than-average refractive index. *See* SHOCK WAVE; SUPERSONIC FLOW; WIND TUNNEL.

In a modern research apparatus, images are usually recorded by means of a charge-coupled-device (CCD) camera, digitized, and stored in a computer. Such a digitized image consists of an array of numbers, each number being proportional to the brightness at a particular point in the image. The image points (pixels) form a closely spaced rectangular grid. A reference image may be taken in the absence of any fluid flow, and the reference image may be divided point by point into images taken with the fluid moving. The divided image is then almost free of artifacts associated with nonuniformity of optical elements, such as windows, as well as nonuniformity of the intensity of the original beam (**Fig. 2**). *See* CHARGE-COUPLED DEVICES. David S. Cannell

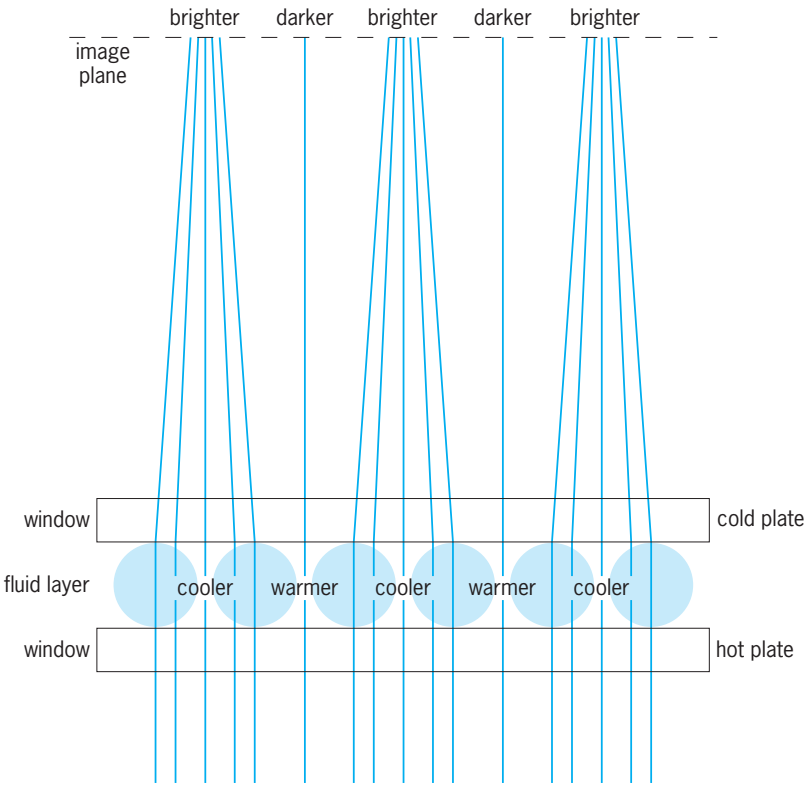


Fig. 1. Schematic of the shadowgraph method applied to reveal convection patterns in a fluid layer. A cross section of the apparatus perpendicular to the convection rolls is shown.

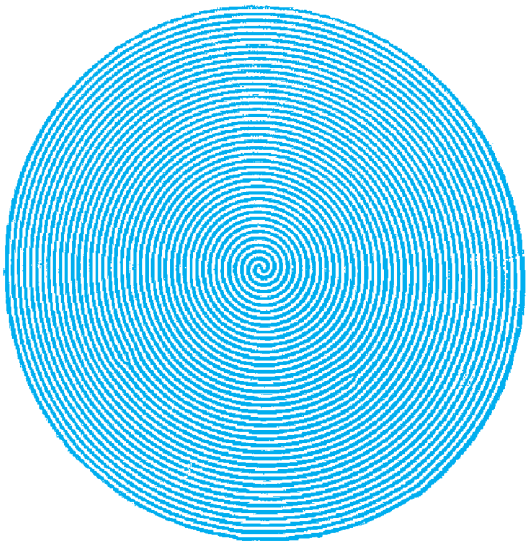


Fig. 2. Background-divided shadowgraph image of convection in high-pressure carbon dioxide gas. The convection rolls are in the form of a two-armed spiral lying within a set of concentric circular rolls. (*From E. Bodenschatz et al., Transitions between patterns in thermal convection, Phys. Rev. Lett., 67:3078–3081, 1991*)

Bibliography. J. R. de Bruyn et al., Apparatus for the study of Rayleigh-Bénard convection in gases under pressure, *Rev. Sci. Instrum.*, 67(6):2043–2067, 1996; R. J. Goldstein (ed.), *Fluid Mechanics Measurements*, 2d ed., Taylor & Francis, 1996; D. R. Jenkins, Interpretation of shadowgraph patterns in Rayleigh-Bénard convection, *J. Fluid Mech.*, 190:451–469, 1988; W. Merzkirch, *Flow Visualization*, 2d ed., Academic Press, Orlando, FL, 1987; G. S. Settles, *Schlieren and Shadowgraph Techniques*, Springer-Verlag, 2001.

Shaft balancing

The process (often referred to as rotor balancing) of redistributing the mass attached to a rotating body in order to reduce vibrations arising from centrifugal force.

Static versus dynamic balancing. A rotating shaft supported by coaxial bearings (for example, ball bearings) together with any attached mass (such as a turbine disk or motor armature) is called a rotor. If the center of mass of a rotor is not located exactly on the bearing axis, a centrifugal force will be transmitted via the bearings to the foundation. The horizontal and vertical components of this force are periodic shaking forces that can travel through the foundation to create serious vibration problems in neighboring components. The magnitude of the shaking force is $F = me\omega^2$, where m = rotor mass, ω = angular speed, and e = distance (called eccentricity) of the center of mass from the rotation axis. The product me, called the unbalance, depends only on the mass distribution, and is usually nonzero because of unavoidable manufacturing tolerances, thermal distortion, and so on. When the center of mass lies exactly on the axis of rotation ($e = 0$), no net shaking force occurs and the rotor is said to be in static balance. Static balance is achieved in practice by adding or subtracting balancing weights at any convenient radius until the rotor shows no tendency to turn about its axis, starting from any given initial orientation. Static balancing is adequate for relatively thin disks or short rotors (such as automobile wheels). However, a long rotor (for example, a turbogenerator) may be in static balance and still exert considerable forces upon individual bearing supports. Suppose that a long shaft is supported on two bearings, and that the unbalance m_1e_1 of the left-hand half of the shaft is equal in magnitude to the unbalance m_2e_2 of the right-hand side, but e_1 and e_2 have opposite signs. In this case, the rotor is in static balance; however, equal but oppositely directed centrifugal forces act on each of the two bearings, transmitting a so-called shaking couple to the foundation. This condition is called dynamic unbalance.

It may be shown that any rigid shaft may be dynamically balanced (that is, the net shaking force and the shaking couple can be simultaneously eliminated) by adding or subtracting a definite amount of mass at any convenient radius in each of two arbitrary transverse cross sections of the rotor. The so-called balancing planes selected for this purpose are usually located near the ends of the rotor, where suitable shoulders or balancing rings have been machined to permit the convenient addition of mass (lead weights, calibrated bolts, and so on) or the removal of mass (by drilling or grinding). Long rotors, running at high speeds, may undergo appreciable elastic deformations. For such flexible rotors it is necessary to utilize more than two balancing planes.

Balancing machines and procedures. The most common types of rotor balancing machines consist of bearings held in pedestals that support the rotor, which is spun at constant speed (as by a belt drive or a compressed air stream). Electromechanical transducers sense the unbalance forces (or associated vibrations) transmitted to the pedestals, and electric circuits automatically perform the calculations necessary to predict the location and amount of balancing weight to be added or subtracted in preselected balancing planes.

For very large rotors, or for rotors which drive several auxiliary devices, commercial balancing machines may not be convenient. The field balancing procedures for such installations may involve the use of accelerometers on the bearing housings, along with vibration meters and phase discriminators (possibly utilizing stroboscopy) to determine the proper location and amount of balance weight.

Committees of the International Standards Organization (ISO) and the American National Standards Institute (ANSI) have formulated recommendations for the allowable quality grade $G = e\omega$ for various classes of machines. A typical value of G (for fans, machine tools, and so on) is 0.25 in./s or 6.3 mm/s, but values as high as 63 in./s or 1600 mm/s (for crankshaft assemblies of large two-cycle engines) and as low as 0.02 in./s or 0.4 mm/s (for gyroscopes and precision grinders) have been recommended. *See* MECHANICAL VIBRATION; SHAFTING. Burton Paul

Bibliography. J. P. Den Hartog, *Mechanical Vibrations*, 1956, reprint 1985; F. Fujisaw et al., Experimental investigation of multi-span rotor balancing using least squares method, *J. Mech. Des., Trans. ASME*, 102:589–596, 1980; C. M. Harris, *Shock and Vibration Handbook*, 4th ed., 1995; B. Paul, *Kinematics and Dynamics of Planar Machinery*, 1979; S. S. Rao, *Mechanical Vibrations*, 3d ed., 1995.

Shafting

The machine element that supports a roller and wheel so that they can perform their basic functions of rotation. Shafting, made from round metal bars of various lengths and machined to dimension the surface, is used in a great variety of shapes and applications. Because shafts carry loads and transmit power, they are subject to the stresses and strains of operating machine parts. Standardized procedures have been evolved for determining the material characteristics and size requirements for safe and economical construction and operation.

Types. Most shafting is rigid and carries bending loads without appreciable deflection. Some shafting is highly flexible; it is used to transmit motion around corners.

Solid shafting. The normal form of shafting is a solid bar. Solid shafting is obtainable commercially in round bar stock up to 6 in. (15 cm) in diameter; it is produced by hot-rolling and cold-drawing or by machine-finishing with diameters in increments of $^1/_4$ in. (6 mm) or less. For larger sizes, special rolling procedures are required, and for extremely large shafts, billets are forged to the proper shape. Particularly in solid shafting, the shaft is stepped to allow greater strength in the middle portion with minimum diameter on the ends at the bearings. The steps allow shoulders for positioning the various parts pressed onto the shaft during the rotor assembly.

Hollow shafting. To minimize weights, solid shafting is bored out or drilled, or hollow pipes and tubing are used. Hollow shafts also allow internal support or permit other shafting to operate through the interior. The main shaft between the air compressor and the gas turbine in a jet aircraft engine is hollow to permit an internal speed reduction shaft with the minimum requirement of space and weight. A hollow shaft, to have the same strength in bending and torsion, has a larger diameter than a solid shaft, but its weight is less. The center of large shafts made from ingots are often bored out to remove imperfections and also to allow visual inspection for forging cracks.

Functions. Shafts used in special ways are given specific names, although fundamentally all applications involve transmission of torque.

Axle. The primary shafting connection between a wheel and a housing is an axle. It may simply be the extension of a round member from each side of the rear of a wagon, and on the end of each the hub of a wagon wheel rotates. Similarly, railroad car axles are large, round bars of steel spanning between the car wheels, supporting the car frame with bearings on the axle outside the wheels. Axles normally carry only transverse loads, as in the examples above, but occasionally, as in rear automobile housings, they also transmit torsion loads. *See* WHEEL AND AXLE.

Spindle. A short shaft is a spindle. It may be slender or tapered. A spindle is capable of rotation or of having a body rotate upon it. It is similar to an arbor or a mandrel, and usage defines the small drive shaft of a lathe as a live spindle. The term originated from the round tapering stick on a spinning wheel on which the thread is twisted.

Head. A short stub shaft mounted as part of a motor or engine or extending directly therefrom is a head shaft. An example is the power takeoff shaft on a tractor.

Countershaft. A secondary shaft that is driven by a main shaft and from which power is supplied to a machine part is called a countershaft. Often it is driven by gears, and thus rotates counter to the direction of the main shaft. Countershafts are used in gear transmissions to obtain speed and torque changes in transmitting power from one shaft to another.

Jackshaft. A countershaft, especially when used as an auxiliary shaft between two other shafts, is termed a jackshaft.

Line shafting. One or more pieces of shafting joined by couplings is used to transmit power from, for example, an engine to a remotely located machine. A single engine can drive many lines of shafting which, in turn, connect in multiple fashion to process equipment machines. Belts operate on pulleys to transmit the torque from one line to another and from the shafting to the machines. Clutches and couplings control the transfer of power from the shafting. *See* CLUTCH; COUPLING.

The delivery of power to the machines in a shop has generally been converted from line shafting to individual electric motor drives for each machine. Thus, in a modern processing plant, line shafting is obsolete. *See* BELT DRIVE; PULLEY; SHAFT BALANCING.

James J. Ryan

Shale

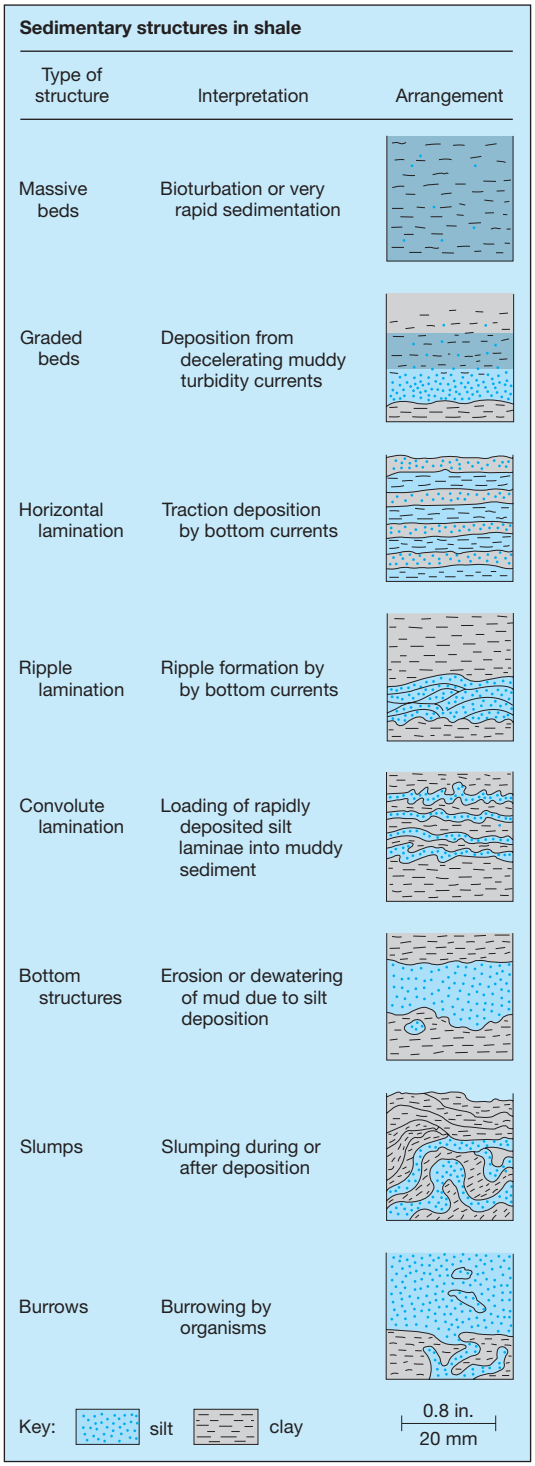
A class of fine-grained clastic sedimentary rocks with a mean grain size of less than 0.0625 mm (0.0025 in.), including siltstone, mudstone, and claystone. One-half to two-thirds of all sedimentary rocks are shales. *See* SEDIMENTARY ROCKS.

Origin and deposition. Shale is deposited as mud, that is, small particles of silt and clay. These particles are produced by weathering and erosion on land and are transported as a hydraulic suspension. The particles are deposited when fluid turbulence caused by currents or waves is no longer adequate to counteract the force of gravity, or if the water evaporates or infiltrates into the ground. There is sufficient turbulence in most bodies of water to keep clay-sized particles suspended, but clay particles often form larger aggregates which settle from suspension more rapidly than individual particles.

Aggregates may be formed by the inorganic process of flocculation, or by filter-feeding organisms which excrete fecal material made up of particle aggregates. Aggregates formed by flocculation tend to be arranged edge-to-face at oblique angles, whereas the particles in aggregates formed by organisms may form stacks of face-to-face clay particles called domains.

Individual silt particles and clay aggregates are often deposited as thin layers less than 10 mm (0.4 in.) thick called laminae. The arrangement of alternate silt and clay laminae in many shales indicates that the silt laminae are traction deposits of weak currents, whereas clay laminae are usually deposited from suspension (see **table**). *See* DEPOSITIONAL SYSTEMS AND ENVIRONMENTS.

Compaction and diagenesis. When deposited, the particle aggregates in clay laminae have highly disordered orientations, which results in a high percentage of void space (porosity) occupied by water. After deposition, the mud may be buried by deposition of more sediment. The weight of the overlying sediment causes compaction. During compaction, the



domains or floccules may be rearranged into a more nearly parallel alignment, which results in a loss of porosity, a reduction in average pore size, and the expulsion of water. The amount of porosity loss which occurs at a specific burial depth is dependent on a number of factors, including the original orientation of the particles, the amount of silt, and the rate of burial. The curve in **illus.** *a* shows the general trend of porosity loss for many shales; the most rapid loss of porosity occurs at shallow depths, and the rate of porosity loss decreases exponentially. *See* MARINE SEDIMENTS.

Diagenesis occurs when the elevated temperatures associated with deep burial cause some of the mineral and organic matter in shales to be transformed into more stable phases. Two chemical transformations in shales which have been studied in detail are the conversion of smectite to illite and the maturation of organic matter into hydrocarbons. The extent of diagenesis depends on the original composition of the sediment, the composition of the fluids which saturate the shale, the rate of deposition, and the thermal gradient. *See* DIAGENESIS.

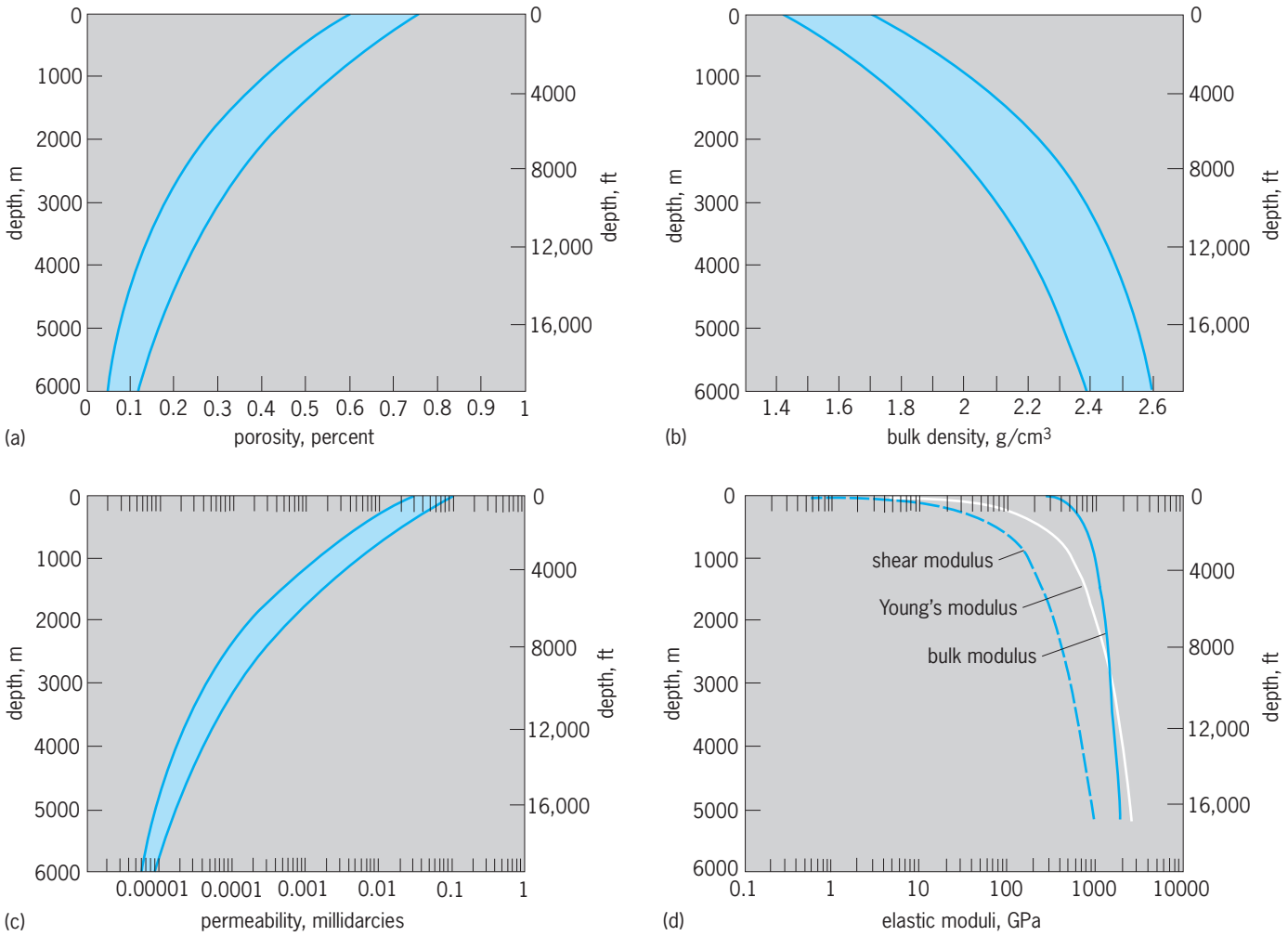
Mineralogical and chemical composition. Mineralogically, most shales are made up of clay minerals, silt-sized quartz and feldspar grains, carbonate cements, accessory minerals such as pyrite and apatite, and amorphous material such as volcanic glass, iron and aluminum oxides, silica, and organic matter.

The most common clay minerals in shales are smectite, illite, kaolinite, and chlorite. The type of clay particles deposited is dependent on the mineralogy, climate, and tectonics of the source area. Paleozoic shales rarely contain smectite or kaolinite, which is at least partially due to the diagenetic conversion of smectite to illite. However, smectite and kaolinite are weathering products commonly formed in soils. The absence of smectite and kaolinite in most Paleozoic rocks may also be related to less intense chemical weathering before land plants evolved in the late Paleozoic. Clay minerals also differ in particle size and in tendency to flocculate; clay mineralogy is thus also dependent upon the distance from the source and the depositional environment. *See* CLAY MINERALS; CHLORITE; ILLITE; KAOLINITE; MONTMORILLONITE; PALEOZOIC.

The amount and size of silt-sized quartz and feldspar in shales are related to the distance of transport and the depositional mechanism. Carbonate cements, amorphous material, and accessory minerals such as pyrite often precipitate in open pores or replace other minerals during burial as a result of diagenesis.

Organic material is normally oxidized shortly after deposition. Shales that are rich in organic material are normally black or dark in color and indicate zones of high productivity, rapid deposition and burial, or lack of oxygen in the depositional environment. After burial, some types of organic matter mature into hydrocarbons; this process is the major source of petroleum and natural gas deposits. *See* NATURAL GAS; PETROLEUM.

Physical and mechanical properties. Most physical and mechanical properties of shale are highly dependent upon changes which occur during burial as a result of compaction and diagenesis. For example, the bulk density of shale is a function of the average grain density, the fluid density, and the porosity. However, fluid and grain density variations are usually insignificant relative to the large amount of porosity lost during burial. Thus bulk density of shale tends to have an inverse relation to the porosity trend (illus. *b*); organic-rich shales generally are less dense,



Depth trends of shale properties: (*a*) porosity, (*b*) bulk density, (*c*) permeability, and (*d*) elastic moduli. Shaded areas show the range of values for most shales.

and shales which contain carbonate cement are more dense than average.

The small size of pores in shale relative to those in sandstone causes shale permeability to be much lower than sand permeability. Permeability in shale decreases exponentially as a function of porosity (illus. *c*); the exponential decrease of porosity with depth causes permeability to decrease four to five orders of magnitude with burial to depths of about 3000 m (10,000 ft). Although fracturing due to compaction stresses or to tectonic movements can create high-permeability deviations from this general trend, shales often form permeability barriers to fluid movement; this has important bearing on the occurrence of subsurface water and hydrocarbons. Ground-water aquifers are commonly confined by an underlying low-permeability shale bed or aquiclude, which prevents further downward movement of the water. Hydrocarbon reservoirs are often capped by low-permeability shale which forms an effective seal to prevent hydrocarbons from escaping. *See* AQUIFER.

The high water content of mud before it is compacted generally results in much less mechanical strength than in uncompacted sand. The realignment of particles, rapid porosity loss, and decrease in pore size cause the mechanical strength to increase rapidly with compaction (illus. *d*). However, even after compaction, the strength of shale is usually less than for most other rocks.

Classification. Shales are usually classified or described according to the amount of silt, the presence and type of lamination, mineralogy, chemical composition, and color. Variations in these properties are related to the type of environment in which the shale was deposited and to postdepositional changes caused by diagenesis and compaction. Fissility, the tendency for shales to split along smooth surfaces parallel to bedding planes, is sometimes confused with lamination. However, fissility is a weathering phenomenon, and although it may be related to lamination, fissility is no longer recommended for use in shale classification.

Most of the individual particles or aggregates which make up shale are too small to be resolved by the human eye, but without using a microscope or other laboratory equipment, shales can be described by determining the presence of silt, presence and type of lamination, presence of carbonate cement, and color. More complete description

of shales requires optical and electron microscopy, x-radiography, x-ray diffraction, and chemical analyses to accurately determine the details of particle size distribution, particle orientation, mineralogy, and chemical composition. *See* PETROFABRIC ANALYSIS; PETROLOGY.

Joseph R. Davis

Bibliography. K. Frye, *Mineral Science: An Introductory Survey*, 1992; C. Klein, *Manual of Mineralogy*, revised 22nd ed., 2001; D. Prothero and F. Schwab, *Sedimentary Geology: An Introduction to Sedimentary Rocks and Stratigraphy*, 1996; C. E. Weaver, *Clays, Muds, and Shales (Developments in Sedimentology)*, 1989.

Shape memory alloys

A group of metallic materials that can return to some previously defined shape or size when subjected to the appropriate thermal procedure. That is, shape memory alloys can be plastically deformed at some relatively low temperature and, upon exposure to some higher temperature, will return to their original shape. Materials that exhibit shape memory only upon heating are said to have a one-way shape memory, while those which also undergo a change in shape upon recooling have a two-way memory. Typical materials that exhibit the shape memory effect include a number of copper alloy systems and the alloys of gold-cadmium, nickel-aluminum, and iron-platinum.

A shape memory alloy may be further defined as one that yields a thermoelastic martensite, that is, a martensite phase that is crystallographically reversible. In this case, the alloy undergoes a martensitic transformation of a type that allows the alloy to be deformed by a twinning mechanism below the transformation temperature. A twinning mechanism is a herringbone structure exhibited by martensite during transformation. The deformation is then reversed when the twinned structure reverts upon heating to the parent phase.

Transformation characteristics. The martensitic transformation that occurs in shape memory alloys yields a thermoelastic martensite and develops from a high-temperature austenite phase with long-range order. The martensite typically occurs as alternately sheared platelets, which are seen as a herringbone structure when viewed metallographically. The transformation, although a first-order phase change, does not occur at a single temperature but over a range of temperatures that is characteristic for each alloy system.

There is a standard method of characterizing the transformation and naming each point in the cycle. Most of the transformation occurs over a relatively narrow temperature range, although the beginning and end of the transformation during heating or cooling actually extends over a much larger temperature range. The transformation also exhibits hysteresis in that the transformation on heating and on cooling does not overlap. This transformation hysteresis varies with the alloy system. *See* ALLOY.

Thermomechanical behavior and alloy properties. The mechanical properties of shape memory alloys vary greatly over the temperature range spanning their transformation. The only two alloy systems that have achieved any level of commercial exploitation are the nickel-titanium alloys and the copper-base alloys. Properties of the two systems are quite different. The nickel-titanium alloys have greater shape memory strain, tend to be much more thermally stable, have excellent corrosion resistance compared to the copper-base alloys, and have higher ductility. The copper-base alloys are much less expensive, can be melted and extruded in air with ease, and have a wider range of potential transformation temperatures. The two alloy systems thus have advantages and disadvantages that must be considered in a particular application.

Applications. The manufacturing process and the technology associated with both of the two commercial classes of shape memory alloys are quite different as are the performance characteristics. Therefore, in applications where a highly reliable product with a long fatigue life is desired, the nickel-titanium alloys are the exclusive materials of choice. Typical applications of this kind include electric switches and actuators. However, if high performance is not mandated and cost considerations are important, then the use of copper-zinc-aluminum shape memory alloys can be recommended. Typical applications of this kind include safety devices such as temperature fuses and fire alarms.

In order to maintain a satisfactory shape recovery during manufacturing, the deformation strain must be constrained from exceeding a critical value that is a function of a variety of processing parameters and the service environment. The critical strain depends upon the geometrical attributes of the product, the qualitative and quantitative nature of the loads, and the number of loading cycles that the product is designed to withstand.

Similarly, heating a shape memory alloy product to a temperature above some critical temperature is not recommended. The critical temperature for nickel-titanium is approximately 250°C (480°F), and for CuZnAl approximately 90°C (190°F). Extended exposure to thermal environments above these critical temperatures results in an impaired memory function regardless of the magnitude of the load.

Another technical consideration in the practical application of shape memory alloys pertains to fastening or joining these materials to conventional materials. This is a significant issue due to the fact that shape memory alloys undergo expansions and contractions not encountered in traditional materials. Therefore, if shape memory alloys are welded or soldered to other materials, they can easily fail at the joint when subjected to repeated loading. Alloys of nickel-titanium and copper-zinc-aluminum can also be brazed by using silver filler metals; however, the brazed region can fail because of cyclic loading. It is therefore desirable to devise some other mechanism for joining shape memory alloys to traditional

materials. Shape memory alloys cannot be plated or painted for similar reasons.

Free recovery. In this case, a component fabricated from a shape memory alloy is deformed while martensitic, and the only function required of the shape memory is that the component return to its previous shape upon heating. A prime application is the blood-clot filter in which a nickel-titanium wire is shaped to anchor itself in a vein and catch passing clots. The part is chilled so it can be collapsed and inserted into the vein; then body heat is sufficient to return the part to its functional shape.

Constrained recovery. The most successful example of this type of product is undoubtedly a hydraulic coupling. These fittings are manufactured as cylindrical sleeves slightly smaller than the metal tubing that they are to join. Their diameters are then expanded while martensitic and, upon warming to austenite, they shrink in diameter and strongly hold the tube ends. The tubes prevent the coupling from fully recovering its manufactured shape, and the stresses created as the coupling attempts to do so are great enough to create a joint that, in many ways, is superior to a weld.

Force actuators. In some applications, the shape memory component is designed to exert force over a considerable range of motion, often for many cycles. Such an application is the circuit-board edge connector. In this electrical connector system, the shape memory alloy component is used to force open a spring when the connector is heated. This allows force-free insertion or withdrawal of a circuit board in the connector. Upon cooling, the nickel-titanium actuator becomes weaker, and the spring easily deforms the actuator while it closes tightly on the circuit board and forms the connections.

An example based on the same principle is a fire safety valve, which incorporates a copper-zinc-aluminum actuator designed to shut off toxic or flammable gas flow when fire occurs.

Proportional control. It is possible to use only a part of the shape recovery to accurately position a mechanism by using only a selected position of the recovery, because the transformation occurs over a range of temperatures rather than at a single temperature. An example is a device for controlling the rate of fluid flow by carefully heating a shape memory alloy component just enough to close a valve the desired amount.

Other applications. A number of applications are based on the pseudoelastic (or superelastic) property of shape memory alloys. Some eyeglass frames use superelastic nickel-titanium alloy to absorb large deformations without damage. Guide wires for steering catheters into vessels in the body have been developed using wire fashioned of nickel-titanium alloy, which resists permanent deformation if bent severely. Arch wires for orthodontic correction also use this alloy.

Shape memory alloys have found application in the field of robotics. The two main types of actuators for robots using these alloys are biased and differential. Biasing uses a coil spring to generate the bias force that opposes the unidirectional force of the shape memory alloy. In the differential type, the spring is replaced with another shape memory alloy, and the opposing forces control the actuation. A microrobot was developed with five degrees of freedom corresponding to the capabilities of the human fingers, wrist, elbow, and shoulder. The variety of robotic maneuvers and operations are coordinated by activating the nickel-titanium coils in the fingers and the wrist in addition to contraction and expansion of straight nickel-titanium wires in the elbow and shoulders. Digital control techniques in which a current is modulated with pulse-width modulation are employed in all of the components in order to control their spatial positions and speeds of operation.

In medical applications, in addition to mechanical characteristics, highly reliable biological and chemical characteristics are very important. The material must not be vulnerable to degradation, decomposition, dissolution, or corrosion in the organism, and must be biocompatible.

Nickel-titanium shape memory alloys have also been employed in artificial joints such as in artificial hip joints. These alloys have also been used for bone plates, for marrow pins for healing bone fractures, and for connecting broken bones. *See* METAL, MECHANICAL PROPERTIES OF; PROSTHESIS. Mel Schwartz

Bibliography. T. W. Duerig, *Engineering Aspects of Shape Memory Alloys*, 1990; S. Eucken (ed.), *Progress in Shape Memory Alloys*, 1992; M. V. Gandi and B. S. Thompson, *Smart Materials and Structures*, 1992; D. E. Hodgson, *Using Shape Memory Alloys: Shape Memory Applications*, 1988.

Shear

A straining action wherein applied forces produce a sliding or skewing type of deformation. A shearing force acts parallel to a plane as distinguished from tensile or compressive forces, which act normal to a plane. Examples of force systems producing shearing action are forces transmitted from one plate to another by a rivet that tend to shear the rivet, forces in a beam that tend to displace adjacent segments by transverse shear, and forces acting on the cross section of a bar that tend to twist it by torsional shear (**Fig. 1**). Shear forces are usually accompanied by normal forces produced by tension, thrust, or bending.

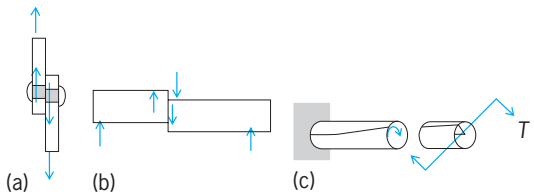


Fig. 1. Shearing actions. (*a*) Single shear on rivet. (*b*) Transverse shear in beam. (*c*) Torsion.

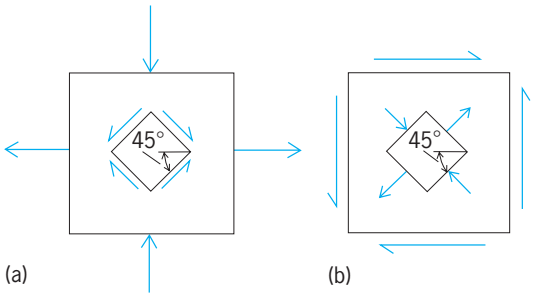


Fig. 2. State of pure shear. (*a*) Shear produces biaxial tension and compression. (*b*) Biaxial stresses induce shear.

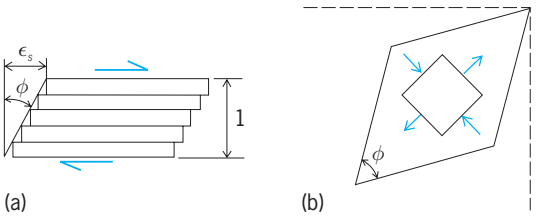


Fig. 3. Shear (*a*) strain and (*b*) distortion.

Shearing stress is the intensity of distributed force expressed as force per unit area. Shear stresses on mutually perpendicular planes at a point in a stressed body are equal. When no normal stresses exist, the state of stress is pure shear, which induces both normal and shear stresses on oblique planes. Elements oriented 45° to planes of pure shear are subjected to biaxial tension and compression equal to the shear stress (**Fig. 2***a*). Similarly, pure shear is induced by equal and opposite biaxial stresses (Fig. 2*b*). Pure shear is the maximum shear stress at the point. Under combined stress, the shear stress is found by principal stress analysis.

Shearing strain is the displacement ϵ_s of two parallel planes, unit distance apart, which accompanies shear stresses acting on these planes. Shearing distortion is visualized as sliding without separation or contraction of all planes parallel to the shear forces, like cards in a pack (**Fig. 3***a*). For planes unit distances apart, the relationship is shown in Eq. (1),

$$\epsilon_s = \tan\phi \approx \phi \text{ radians} \qquad (1)$$

where ϕ is the small angle of distortion.

Modulus of rigidity, designated E_s or G, is the shearing modulus of elasticity, which according to Hooke's law is the constant of proportionality between shearing stress S_s and shearing strain S_s/ϕ during elastic behavior.

After shearing distortion, the angles of a rectangular element are altered by the shear strain ϕ (Fig. 3*b*). Changes in length of diagonals must be consistent with the biaxial strains, which leads to the relationship shown in Eq. (2), where E is Young's modulus

$$E_s = \frac{E}{2(1+\mu)} \qquad (2)$$

and μ is Poisson's ratio. The value of μ can be found when E_s and E are experimentally determined. *See* STRESS AND STRAIN.

John B. Scalzi

Shear center

A point on a line parallel to the axis of a beam through which any transverse force must be applied to avoid twisting of the section. A beam section will rotate when the resultant of the internal shearing forces is not collinear with the externally applied force. The shear center may be determined by locating the line of action of the resultant of the internal shear forces. A rolled wide flange beam section has two axes of symmetry, and therefore the shear center coincides with the geometric center or centroid of the section. When such a beam member is loaded transversely in the plane of the axes, it will bend without twisting.

However, a channel section with one axis of symmetry is subject to twisting unless the externally applied load passes through the shear center of the section (**Fig. 1**). The shear center for a channel bending about the strong axis (x–x) is located a distance e from the center of the web and lies behind the web. An expression may be derived to determine this eccentricity for different sizes of channels of uniform thickness, such that $e = B^2h^2t/4I_x$, where B is the

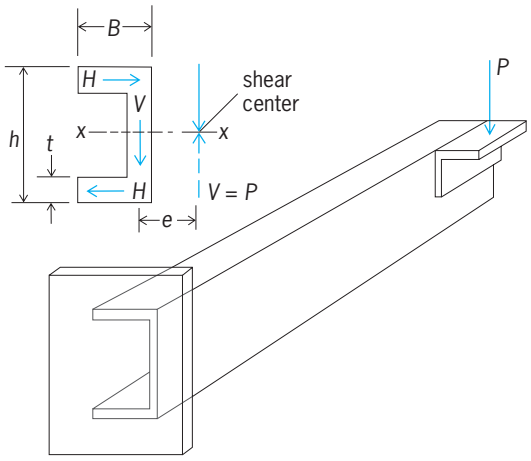


Fig. 1. Shear center under *P* at distance *e* from centerline web, on symmetrical axis.

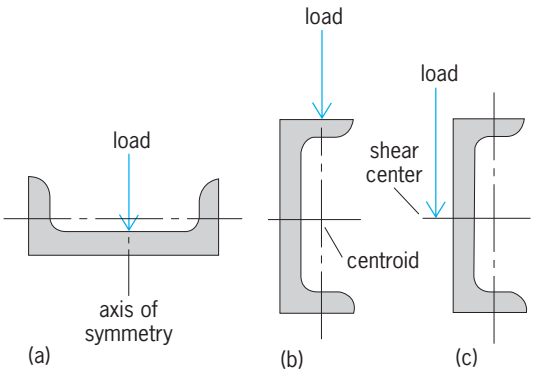


Fig. 2. Three conditions of channel section under load. Channel (*a*) resists bending, (*b*) twists, and (*c*) resists twisting.

flange width, b is the depth of section, t is the thickness of web and flange, and I_x is the moment of inertia about the x axis.

Other structural sections with one axis of symmetry will twist unless properly restrained or loaded through the shear center. **Figure 2***a* illustrates the loading of a channel through the shear center causing bending about the nonsymmetrical axis and no twist; Fig. 2*b* illustrates the condition which will cause twisting; and Fig. 2*c* indicates a condition which will resist twisting. *See* LOADS, TRANSVERSE.

John B. Scalzi

Bibliography. J. S. Iffland (ed.), *Steel Structures*, 1989.

Sheep

Sheep are members of the family Bovidae in the order Artiodactyla, the even-toed, hoofed mammals. Sheep were possibly the first animals to be domesticated about 11,000 years ago in southwestern Asia, a time when sheep were closely associated with the religious, civic, and domestic life of the people. They are referred to more often in the Bible than any other animal.

Wild sheep range over the mountainous areas of the Northern Hemisphere. They have long, curved, massive horns and mixed coats with a hairy outercoat and a woolly undercoat. Although wild sheep vary in size, they are usually larger than domestic sheep and have shorter tails.

Distribution. Over 1 billion domestic sheep extend over most of the land areas of the world. They are most heavily concentrated in the warmer parts of the temperate zones and are found in greatest numbers in the Southern Hemisphere in Australia, New Zealand, South Africa, and South America. Other countries where there are high sheep numbers include Russia, China, India, Iran, Turkey, and the United Kingdom. Sheep are more common in drier regions and at high elevations. Hair sheep are found in the wet tropics.

Domestic sheep were first brought to the Americas by Spanish explorers. Sheep were introduced into the English colonies almost as soon as they were settled. By 1840, New England was the center for sheep production in the United States; the center moved westward to Ohio by 1850. This westward movement of sheep continued, encouraged by the availability of year-round grazing, and by 1860 there were many sheep in California and Texas. The sheep population reached a peak in 1884 with over 50 million head. Since then numbers have declined to under 11 million in the 1980s. Sheep are common in the north-central states and from Missouri to Virginia, but the bulk are in the western states and Texas.

Terminology. Sheep are called lambs until about 12 months of age, from which time to 24 months they are referred to as yearlings, and thereafter as two-year-olds, and so on. The female sheep is called a ewe and the male, a ram or buck; the castrated male is called a wether. Sheep meat is called lamb or mutton depending on the age at slaughter: Young carcasses, where the forefeet are removed at the break joint or temporary cartilage just above the ankle, are sold as lamb; mutton carcasses are those of older animals where it is necessary to take the feet off at the ankle. Mutton is sometimes less desired, brings a lower price than lamb, and may be used in manufactured meats.

General description. A precise description is difficult since sheep are so variable. Horns, if present, tend to curl in a spiral. If horns occur in both sexes, they are usually larger for the ram. In some breeds only the rams have horns, but many breeds are hornless. The hairy outercoat common to wild sheep was eliminated by breeding in most domestic sheep. Wool may cover the entire sheep, or the face, head, legs, and part of the underside may be bare; or wool may be absent as in some breeds (bred for meat) with a hairy coat or a coat which constantly sheds. Although domestic sheep are mostly white, shades of brown, gray, and black occur, sometimes with spotting or patterns of color. The faces and legs are white in some white breeds while black or mottled in others. The shape of the face and head varies from a small dished face to one with a large prominent nose. The ears may be small and pointed; somewhat straight and moderate-sized or long and pendulous; and erect or drooping. Body form varies from the angular, uneven shape of wool-type breeds to the smooth, rounded, blocky appearance of meat-type breeds. The tail tends to be long in most domestic breeds, but is often removed or docked soon after birth. It may be short in some breeds, but may be bare or completely covered with wool in others. Some desert breeds have large fat deposits on the rump or the tail.

The shape of the body of a sheep is determined by the bony skeleton and the amount and distribution of muscle and fat. The thoracic and lumbar vertebrae have both lateral and vertical extensions which contribute to the shape of the back. The front legs have no direct connections to the body skeleton, but the shoulder blades are attached by muscles. Body weights for mature animals may vary from well under 100 to 400 lb (45 to 181 kg) or more. Body measurements vary widely with breed, sex, age, and degree of fatness.

As in other ruminants the stomach of the sheep has four compartments and is adapted to regurgitation and cud chewing. Prior to regurgitation and further chewing, food is stored in the rumen, the largest compartment where roughages are broken down by microorganisms. When food is swallowed a second time, it passes into the third stomach or omasum, which is made up of lengthwise leaves and is connected to the abomasum or true stomach. *See* DIGESTIVE SYSTEM.

The sheep has 20 lamb or temporary teeth, and 32 permanent teeth of which 8 are incisors; a cartilaginous pad is found instead of incisors in the upper jaw. The sheep bites by clamping herbage between the incisor teeth and the dental pad and then jerking or twisting the head.

Although sheep mature in 1–2 years, they may not reach maximum weights and production until 3–5 years of age. The average life-span is about 6–7 years with the productive life sometimes being shorter: survival beyond 10–11 years is rare. Replacement of the lamb incisors with a pair of larger permanent incisors is generally used to estimate age of sheep, although this practice is quite inexact. The first central pair of permanent incisors appears around 1 year of age, and an additional pair is added approximately yearly until a full mouth of four pairs is obtained. After 4 years, age may be estimated from the spread, wear, or loss of the teeth. However, the only sure way to determine age is through individual or group identification from birth throughout life.

Wool and skin. Wool fiber is made up largely of insoluble, sulfur-containing proteins or keratins. The crimps or waves of the wool fiber vary from only 1–2 up to 25 or more per inch (1 up to 10 or more per centimeter), but the waves of adjacent fibers are generally parallel so that blocks of fibers in locks of wool or even over the entire fleece tend to be uniformly crimped. Fiber diameter varies from 10 to 70 micrometers. The number of crimps per inch is greater in fine wool which has fibers of smaller diameters. Fiber length is generally recognized as length of staple, that is, length of wool as it grows on the sheep, and this may vary from about 1 to 6 in. (2.5 to 15 cm) or more for 1 year of growth. Length and diameter are associated to some extent and tend to vary with breed, nutrition, and possibly other environmental factors. Wool grows continually at a somewhat regular rate over the lifetime of the sheep even if it is not removed.

The skin of a sheep is about $^1/_{12}$ in. (2 mm) thick, may be pigmented to various degrees, and contains sweat and sebaceous glands which, together with the hair and wool follicles, are generally arranged in a typical pattern. Those follicles which develop prenatally are called primary follicles, and each is associated with a sweat gland, oil gland, and a smooth muscle. Secondary follicles develop later and have only an oil gland. Primary follicles usually occur in groups of three with a variable number of secondary fibers. There are great variations within breeds and on individual animals regarding the total number of fibers per square inch; totals range from less than 5000 to more than 60,000 per square inch (775 to more than 9300 per square centimeter).

Reproduction. Sheep may reach puberty as early as 100–150 days, although 5–10 months is more common. The first estrus in ewe lambs may occur somewhat later than the first sperm production in ram lambs. Sexual maturity may be more closely related to weight than to age and usually occurs when body weights reach 40–60% of mature weight; it may be delayed even beyond 1 year of age if nutrition is poor. Sexual maturity varies with breed, being earlier for fast-growing breeds such as the Suffolk than for slow-growing breeds such as the Merino, and is generally earlier for crossbred lambs than for purebreds. Ram lambs can be used successfully in breeding, but males are more commonly mated first at 18–20 months of age. Although ewes are often bred first to lamb at 2 years of age, it is not uncommon to breed ewe lambs. Ewes breeding first as lambs generally will have greater lifetime lamb production than ewes bred first as yearlings. Conception rates for ewe lambs may vary 10–60% or more.

The breeding season is related to day length and generally commences as days become shorter; most domestic breeds of sheep have a breeding season restricted to the fall and winter months. There is considerable variation in estrous periods among breeds and among individuals, varying from a few periods to almost year-round breeding. The presence of rams near the beginning of the breeding season may stimulate the onset of the first estrus. Rams do not show a restricted breeding season, but semen production and quality are generally higher in the fall and may decline in the spring and summer; summer sterility may occur under continuous high temperatures.

Estrus. In the ewe, estrus can only be recognized by willingness to stand and to allow the ram to mount. Sometimes the vulva is enlarged, and there is usually a liquefaction and flow of mucus from the cervix. Cornified epithelial cells may be observed in the vaginal smear following estrus. Estrus may vary in duration from a few hours to 3 days or more with an average of 1–2 days; it may be shorter when rams are with ewes continuously. Ovulation normally occurs near the end of estrus.

The normal estrous cycle in the ewe varies from 14 to 19 days with an average of 17 days. Ovulation without estrus generally occurs one cycle before the first estrus. A postpartum estrus may occur within the first day or two after parturition but is not accompanied by ovulation. Ovulation and estrus may occur during lactation but both are more likely to occur after the first 4–10 weeks. Giving progesterone for 12–14 days during the breeding season may synchronize estrus, with estrus and ovulation usually occurring 1–5 days after the treatment is stopped. There may be some reduction in fertility if ewes are bred at the first synchronized estrus. *See* ESTRUS.

Ovulation. The number of eggs shed per heat period (ovulation rate) limits the number of lambs produced and therefore is an important aspect of productivity. It varies with year, season, age, breed, and nutrition. The most common ovulation rate as judged from the number of lambs born is one, although two occur often, three occasionally, four sometimes, and five or six rarely; up to nine have been reported. Ovulation rate increases with age, reaching a maximum at 3–6 years of age. It tends to be high in the early part to the middle of the breeding system but decreases toward the end.

Pregnancy. The average duration of pregnancy varies among breeds from 144 to 155 days; normal pregnancies may range from 138 to 159 days in individuals. Early-mating breeds such as the Southdown and Shropshire have short gestation periods of about 145 days, as do high-fertility breeds such as Finnsheep or Romanov with periods of 142–146 days, while late-maturing breeds such as the

Merino or Rambouillet may have average periods as long as 151 days. Crossbred types such as the Columbia or Targhee tend to be intermediate. Pregnancy is shorter for twin lambs than for singles; it may be longer in older ewes.

The sex ratio is generally 49–50% males. Identical twins are rare. A few cases of freemartin lambs have been reported.

A number of reliable pregnancy tests may be used with sheep. Methods which are both rapid and accurate include palpation by laparotomy, observation with peritoneoscope, and ultrasonic detection of pregnancy either by echo or by the Doppler technique. This last method is highly accurate at about 60 days following conception or even earlier. Diagnosis with x-rays and with fetal electrocardiography is possible but of limited usefulness. Hormonal methods are coming into practical use with accurate diagnosis as early as 18 days after conception.

Parturition. The termination of pregnancy is determined in part by hormonal changes in which fetal adrenal activity initiates a series of hormonal trends culminating in parturition. The pelvic ligaments, vagina, and cervix tend to relax near the end of pregnancy. Parturition may be more common in the forenoon and late afternoon. The first signs of parturition are uneasiness, pawing and frequent turning, or lying down and standing up. The forefeet of the lamb generally appear first in front of the head; one or both front legs turned back are regarded as malpresentations. The lamb is licked dry by the mother following birth and sucks milk within an hour or two. The afterbirth is generally expelled 2–4 h after parturition. The uterus normally returns to nonpregnant size in 2 weeks and is completely involuted by 4 weeks. Synchronization of parturition is feasible with hormonal treatments.

Male physiology. The spermatogenic cycle in the ram is about 49 days. The average sperm production has been measured at 5.5 billion per day. High-quality ram semen is characterized by having at least 2 billion sperm per cubic centimeter, a rapid swirling movement, a pH about 7.0 or with more acidity, and with less than 25% abnormal sperm or less than 5% abnormal heads. Rams show wide variation in frequency of copulation from once in about 10 h to over once an hour. The number increases in direct proportion to the number of available ewes in heat.

Altering reproductive cycles. Artificial insemination techniques have been developed for sheep, but fresh semen gives higher fertility than frozen semen. Practical use of artificial insemination generally requires synchronization of estrus, involvement of genetic improvement, and repeated reproductive cycles to reduce the length of dry periods.

The reproductive rate of sheep may be increased by selection, choice of breed, crossing with high-fertility breeds, crossbreeding, control of reproduction, multiple lambing per year, single-gene high ovulation rate, immunizing techniques, and milk replacers. Combinations of these may be used, but selection is basic because it improves all aspects of weight of lambs weaned of ewes mated, a large part of efficiency, and also improves adaptability to the particular environment. Complete control of reproduction in sheep is now feasible through synchronization of estrus, artificial insemination, early pregnancy diagnosis, and synchronization of parturition, but is not in general use. *See* BREEDING (ANIMAL).

Adaptation. Although sheep are well adapted to cold climates, they have excellent tolerance to heat because of the excellent temperature regulation effected by their wool. Shearing may increase heat tolerance in summer and decrease it in cooler periods. Body temperature maintained at about 102°F (39°C) may vary from about 98 to 106°F (37 to 41°C). The most important source of heat loss is evaporation from the respiratory tract; however, sheep do sweat to a limited extent.

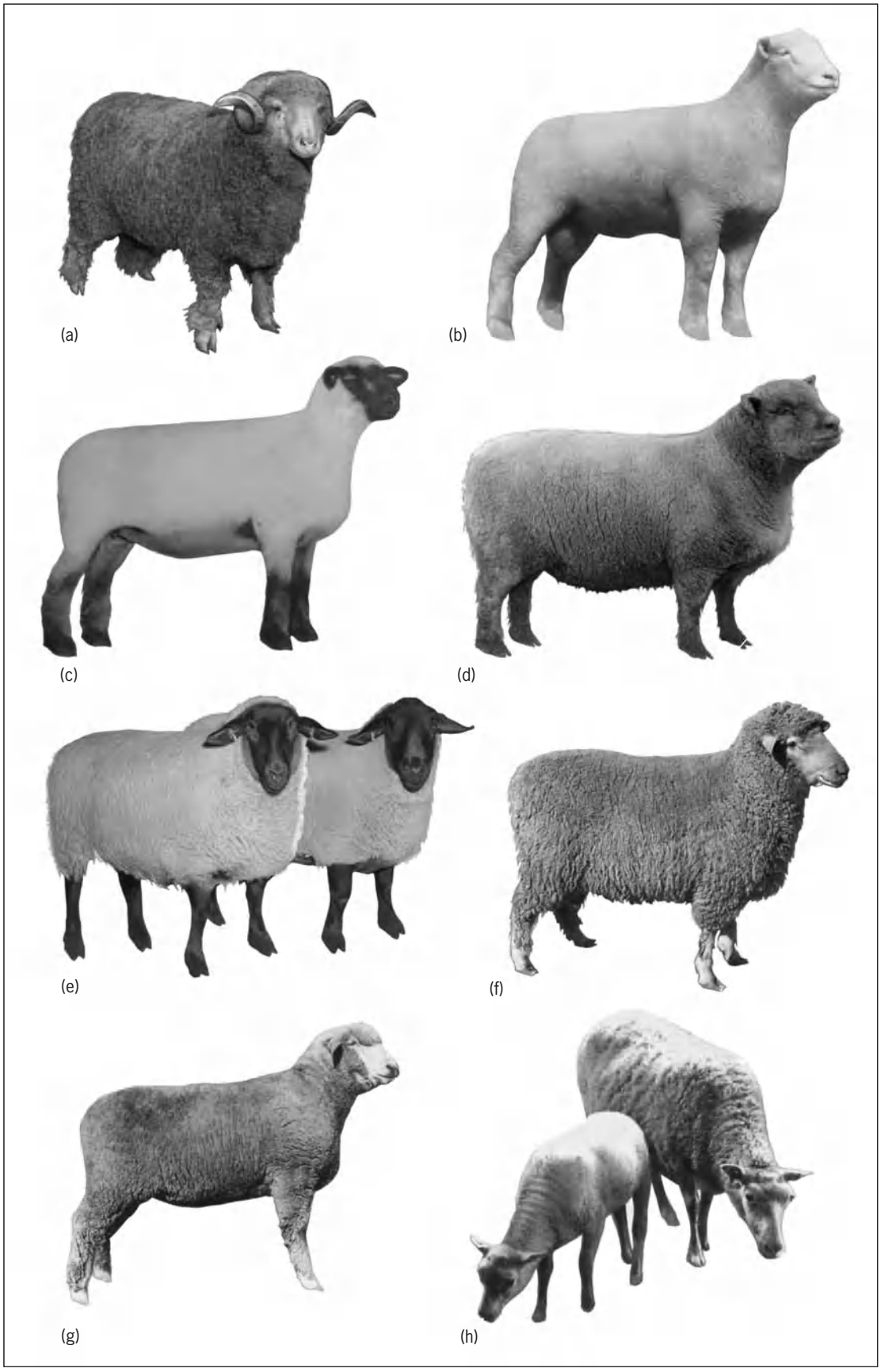
Sheep are probably most productive in temperate climates. Hot climates are unfavorable for lamb meat or market lamb production because of reduced fertility and reduced feed intake. Wool may usually be produced efficiently in hot, dry climates. Internal parasites become abundant and cause serious losses in warm, humid climates. Both lamb and wool are produced well in cold climates, but environmental protection may be necessary at lambing, after shearing, and under wet or stormy conditions.

Breeds. Hundreds of breeds of sheep of all types, sizes, and colors are found over the world; some of the more prominent breeds are shown in the **illustration**. Wool-type breeds, mostly of Merino origin, are important in the Southern Hemisphere, but both fine- and long-wool types are distributed all over the world. Sheep with fat tails or fat rumps are common in the desert areas of Africa and Asia. These usually produce carpet wool. Milk breeds are found mostly in central and southern Europe. Meat breeds from the British Isles are common over the world.

Over 40 breeds of sheep, about half of which were imported from the British Isles, are represented in the United States. The Suffolk and Hampshire breeds represent over half of the purebred registered sheep of the country and are used as sires of slaughter lambs. Other British breeds used for this purpose include Southdown, Shropshire, Dorset, Oxford, and Cheviot. The Dorset is favored for its ability to breed in the spring and fall. About 90% of registered Dorsets are polled.

In the United States, fine wool breeds of the Merino from Spain and the Rambouillet from France and Germany have formed the basis for much of the western range sheep which are often crossed with meat-type rams to produce slaughter lambs. These two European breeds and the Debouillet resulting from their cross are used in the Southwest where wool production is important.

Breeds from long-wool–fine-wool breed crosses have spread over the entire United States and are truly dual-purpose in type, although they are often crossed with meat-type rams for commercial lamb production. The Corriedale, imported from New Zealand, was first used in the West but is now more common in the Midwest; the Columbia and Targhee



Examples of prominent breeds of sheep. (*a*) Rambouillet ram (*Sheep and Goat Raiser*). (*b*) Dorset ram (*Continental Dorset Club*). (*c*) Hampshire ram. (*d*) Southdown ram (*Southdown Association*). (*e*) Suffolk ewe and ram (*photograph by A. L. Henley*). (*f*) Columbia ram (*photograph by A. Sponagel*). (*g*) Targhee ram (*Cook and Gormley*). (*h*) Finnsheep ewe and lamb (*U.S. Department of Agriculture*).

breeds, developed by the Department of Agriculture, are large, rapidly growing, white-faced, polled sheep. The Columbia, which resulted from crossing Lincoln rams on Rambouillet ewes, produces heavy fleeces of 50–58s in spinning count while the Targhee, a three-quarter Rambouillet and one-quarter long-wool, produces fleeces of 60–62s in spinning count. Another crossbred type is the Montadale, from Columbia–Cheviot.

The Romney is the most numerous of the long-wool breeds, although some Lincolns are maintained primarily for crossing with fine-wool sheep. The Cotswold and Leicester breeds have largely disappeared from the United States. The Tunis and Karakul breeds are quite minor in importance.

Breed improvement. Performance and progeny testing of rams both in central stations and on farms has been carried on in many states. The proportion of purebred sheep selected on production records is small but increasing. The Rambouillet and Hampshire breeds have led in development of plans for certifying rams on the performance of their offspring.

The average flock size for purebred flocks tends to be as small as 25–30 head with only one or two rams used per season. This small flock size limits selection so that breed improvement results from decisions made in only a very few flocks. Wool-type breeds tend to be kept in larger flocks than the meat-type breeds.

Crossbreeding for slaughter lambs has been used extensively for a number of years. Crosses of meat-type rams on wool-type or dual-purpose ewes not only produce desirable slaughter lambs but also result in hybrid vigor which, from two-, three-, and four-breed crosses, has produced important increases in number and pounds of lambs weaned per ewe bred over comparable averages of the purebred parents.

Some exotic breeds might be used to advantage in the United States if disease restrictions on importation could be overcome. Finnsheep have shown their ability to increase lambing rates. Other desirable traits for improvement include adaptability to particular environments, less restricted breeding seasons, ability to produce more than one lamb crop per year, greater efficiency of feed use, and more desirable carcass traits.

Nutrition and feeding. Forages make up about 95% of sheep diets and supply most of the energy needs. Diets should contain 10–15% protein, generally from leguminous plants. Sheep prefer fine forages, but they eat a large variety of grasses, legumes, weeds, herbs, and shrubs. Legume hays are preferable, but corn or sorghum fodder, cereal or grass hays, and straws are often fed. Silages from corn, sorghum, cereal, and other plants may be used to replace part of the hay ration. Nonleguminous roughages may need protein and mineral supplements, but lack of energy may be the most common nutritional deficiency of sheep. Concentrates, including the common farm grains of oats, barley, corn, wheat, and sorghum, are fed sparingly, usually in winter, also prior to and following lambing, and to growing and fattening lambs. Common protein supplements include cake or oil meal from soybeans, flaxseed, or cottonseed. Feeding of by-products, crop residues, and wastes such as poultry or other manures is increasing. *See* ANIMAL FEEDS.

Minerals such as calcium, phosphorus, iodine, copper, and cobalt may need to be supplied. Sheep may do without salt, but if supplied they will normally consume about 1 lb (0.45 kg) per month. Vitamins A, D, and E are required and may have to supplement certain rations. Vitamins C and the B complex are synthesized by sheep.

Feed required for maintenance and growth or production varies with age, weight, phase of reproduction, and climatic conditions. Nonlactating ewes and those in early gestation require 1.6–2.2% of body weight of air-dry feed. Requirements of mature ewes increase in late pregnancy and are greatest in early lactation with a range of 3.2–4.8% of body weight. Lambs and yearlings require 2.2–4.5% of their body weight with greater requirements per unit of body weight at lighter weights.

Pelleting of feeds reduces handling, storage space, and waste. Pelleting of high-roughage diets generally results in increased feed consumption and eliminates selection of the more palatable portions of the roughage. With pellets there is greater feed intake resulting from a more rapid passage of the finely ground material through the digestive tract.

Antibiotics may improve performance of growing and fattening lambs, particularly under conditions of stress. Supplemental or creep feeding of lambs from 2 to 6 weeks of age up to the time of weaning is often advantageous, particularly on ordinary pastures.

Production systems. Sheep under farm conditions are generally kept in fenced pasture from the spring to fall months. Housing, or some shelter, and harvested feed are usually provided in winter, especially in the colder and more humid climates. Flock size is typically small, 25 head or less, but may range up to 150 and sometimes up to several hundred.

Range sheep of the western states are often herded in bands of 700–1000 head of ewes with lambs and up to 2000 head or more of dry ewes. Some rangeland is fenced, and range sheep are sometimes herded only part of the year or not at all. Flock size is generally large with individual holdings often of several thousand up to 15,000 or 20,000. Lambing is often done in sheds although lambing on the range is not uncommon.

Intensive production involving large numbers of sheep in concentrated areas, generally with confinement for part or all of the year, may be practical. This system often involves lambing in the fall and winter months and may include lambing more often than once per year.

Management. Management practices are similar for different production systems although wide variations do exist. Breeding normally occurs in the fall for a period of 45–90 days. In purebred matings individual rams are mated to groups of ewes, while in commercial flocks rams are generally group mated, with from 2 or 3 rams per 100 ewes.

Extra care is usually provided at spring lambing time to increase the survival rate of the lambs. Ewes are commonly shorn or crutched just prior to lambing. The newborn lamb may be assisted in suckling and is often given extra protection from cold or wet weather. Docking and castrating may be done with rubber bands at birth or soon after, or by cutting within 6–10 days or up to 1 month after birth. The umbilical cord of the newborn lamb is dipped in iodine at birth to prevent infection. Vaccination for tetanus, overeating disease, and sore mouth may be done on the young lamb or sometimes later. Purebred lambs may be individually identified with ear tags at birth. Lambs are often identified with their mothers with scourable paint brands.

Shearing to remove the annual growth of wool is usually done in the spring but may range from January to June. Fleeces are individually tied with paper string and are packed in bags holding about 20 to 40 fleeces. Average fleece weight is about 8 lb (3.6 kg).

Lambs may be weaned as early as 3 weeks but weaning under weights of 30–45 lb (13.6–20.4 kg) is not recommended. While weaning age may extend up to 5 months, most ewes decline in milk production as lactation progresses. Lambs are often practically weaned by their mothers by 4 months more or less. The artificial rearing of lambs with milk replacer from birth is desirable when the lamb would otherwise die.

Culling is usually done in the fall. Ewes which fail to reproduce, are unsound, have bad udders, have poor or missing teeth, or are losing weight with age should be eliminated. Ewes in their sixth year or older which fail to wean lambs should generally be culled. Ewes with low lamb production may be culled at any age. Ewes with light fleeces should generally be culled when young.

Disease and parasites. Losses from diseases are moderate to low, although sheep suffer from many maladies. Common diseases include abortion, pneumonia, and overeating disease. Foot rot can be eliminated or controlled by tedious trimming and treating of feet and isolation of infected animals. Vaccination is effective for some diseases, particularly blue tongue and sore mouth. Epididymitis may interfere with ram fertility.

External parasites such as lice or keds are troublesome and may be controlled by dipping, spraying, or dusting. Sheep scabies has been eradicated in the United States.

Internal parasites cause tremendous losses, particularly in warm humid areas. The stomach worm is probably the most common parasite. Weaning the lambs onto clean pasture from dry lot is generally effective in preventing or delaying the increase of parasites on summer pastures.

Products. Wool is an important source of income from sheep but it is less important than lamb meat. Wool is valued or graded chiefly on fineness or fiber diameter, staple length, clean yield, and freedom from contamination with foreign material. Color, strength, handle, and crimp also are often considered. Wool is a superior fiber, even though synthetic fibers are often cheaper, and it is especially useful for warmth, moisture absorbability, and resistance to fire. It is unique in extensibility, resilience or wrinkle resistance, felting ability, and durability. Fine and medium wools are generally used for apparels, and coarse wools for rugs and carpets.

Lamb meat is the most important product of sheep in the United States. High-quality carcasses are thick and blocky with a high proportion of lean, large loin-eye muscle, and with a thin covering of fat. The fat has a high melting point, and the meat has a characteristic flavor. The loin chops, rib chops, and legs are the favored cuts. Consumption of lamb is about 2 lb (0.9 kg) per person per year. Lamb is not well distributed in the United States and therefore not available to many.

Sheep are used for milk production as well as for meat and wool in many countries, particularly in southern Europe. Ewe milk averages about 6% fat and is superior for making cheese. Sheep may yield up to 130 gal (492 liters) of milk annually. Sheep manure is valued in some areas. Skins are used for a variety of lightweight leathers. Pelts may be tanned with the wool on for items such as coats, rugs, and slippers. Fur pelts are produced from young Karakul lambs. Sheep have been used rarely for work and sport. *See* AGRICULTURAL SCIENCE (ANIMAL).

Clair E. Terrill

Bibliography. I. E. Coop (ed.), *Sheep and Goat Production*, 1982; M. E. Ensminger and R. O. Parker, *Sheep and Goat Science*, 1985; R. Jensen and B. L. Swift, *Diseases of Sheep*, 1982.

Sheet-metal forming

The shaping of thin sheets of metal (usually less than $^1/_4$ in. or 6 mm) by applying pressure through male or female dies or both. Parts formed of sheet metal have such diverse geometries that it is difficult to classify them. In all sheet-forming processes, excluding shearing, the metal is subjected to primarily tensile or compressive stresses or both. Sheet forming is accomplished basically by processes such as stretching, bending, deep drawing, embossing, bulging, flanging, roll forming, and spinning. In most of these operations there are no intentional major changes in the thickness of the sheet metal. *See* METAL FORMING.

There are certain basic considerations which are common to all sheet forming. Grain size of the metal is important in that too large a grain produces a rough appearance when formed, a condition known as orange peel. For general forming an ASTM no. 7 grain size (average grain diameter 0.00125 in. or 32 micrometers) is recommended. Another type of surface irregularity observed in materials such as low carbon steel is the phenomenon of yield-point elongation that results in stretcher strains or Lueder's bands, which are elongated depressions on the surface of the sheet. This is usually avoided by cold-rolling the original sheet with a reduction of only 1–2% (temper rolling). Since yield-point elongation reappears after

some time, because of aging, the material should be formed within this time limit. Another defect is season cracking (stress cracking, stress corrosion cracking), which occurs when the formed part is in a corrosive environment for some time. The susceptibility of metals to season cracking depends on factors such as type of metal, degree of deformation, magnitude of residual stresses in the formed part, and environment. *See* METAL, MECHANICAL PROPERTIES OF.

Anisotropy or directionality of the sheet metal is also important because the behavior of the material depends on the direction of deformation. Anisotropy is of two kinds: one in the direction of the sheet plane, and the other in the thickness direction. These aspects are important, particularly in deep drawing.

Formability of sheet metals is of great interest, even though it is difficult to define this term because of the large number of variables involved. Failure in sheet forming usually occurs by localized necking or buckling or both, such as wrinkling or folding. For a simple tension-test specimen the true (natural) necking strain is numerically equal to the strain-hardening exponent of the material: thus, for instance, commercially pure annealed aluminum or common 304 stainless steel stretches more than cold-worked steel before it begins to neck. However, because of the complex stress systems in most forming operations, the maximum strain before necking is difficult to determine, although some theoretical solutions are available for rather simple geometries. *See* ALUMINUM; STAINLESS STEEL.

Considerable effort has been expended to simulate sheet-forming operations by simple tests. In addition to bend or tear tests, cupping tests have also been commonly used, such as the Swift, Olsen, and Erichsen tests. Although these tests are practical to perform and give some indication of the formability of the sheet metal, they generally cannot reproduce the exact conditions to be encountered in actual forming operations.

Stretch forming. In this process the sheet metal is clamped between jaws and stretched over a form block. The process is used in the aerospace industry to form large panels with varying curvatures. Stretch forming has the advantages of low die cost, small residual stresses, and virtual elimination of wrinkles in the formed part.

Bending. This is one of the most common processes in sheet forming. The part may be bent not only along a straight line, but also along a curved path (stretching, flanging). The minimum bend radius, measured to the inside surface of the bend, is important and determines the limit at which the material cracks either on the outer surface of the bend or at the edges of the part. This radius, which is usually expressed in terms of multiples of the sheet thickness, depends on the ductility of the material, width of the part, and its edge conditions. Experimental studies have indicated that if bending is carried out in a pressurized environment, with pressures as high as 500,000 lb/in.^{2} (3.4 gigapascals), the minimum bend radius decreases substantially. This observation is important in forming brittle materials. Heating the workpiece is another method of improving bendability.

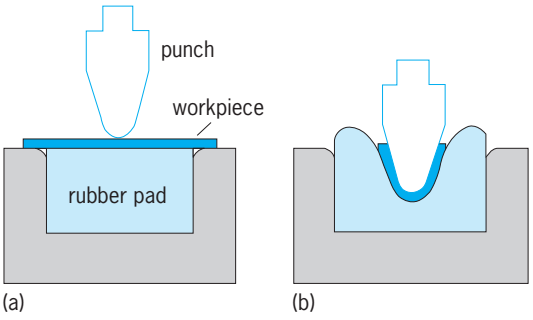


Fig. 1. Bending process with a rubber pad. (*a*) Before forming. (*b*) After forming.

Springback in bending and other sheet-forming operations is due to the elastic recovery of the metal after it is deformed. Determination of springback is usually done in actual tests. Compensation for springback in practice is generally accomplished by overbending the part; adjustable tools are sometimes used for this purpose.

In addition to male and female dies used in most bending operations, the female die can be replaced by a rubber pad (**Fig. 1**). In this way die cost is reduced and the bottom surface of the part is protected from scratches by a metal tool. The roll-forming process replaces the vertical motion of the dies by the rotary motion of rolls with various profiles. Each successive roll bends the strip a little further than the preceding roll. The process is economical for forming long sections in large quantities.

Rubber forming. While many sheet-forming processes are carried out in a press with male and female dies usually made of metal, there are four basic processes which utilize rubber to replace one of the dies. Rubber is a very effective material because of its flexibility and low compressibility. In addition, it is low in cost, is easy to fabricate into desired shapes, has a generally low wear rate, and also protects the workpiece surface from damage.

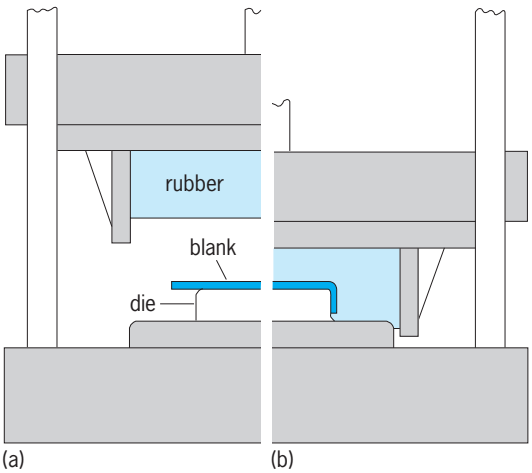


Fig. 2. Guerin process, the simplest rubber-forming process. (*a*) Before forming. (*b*) After forming.

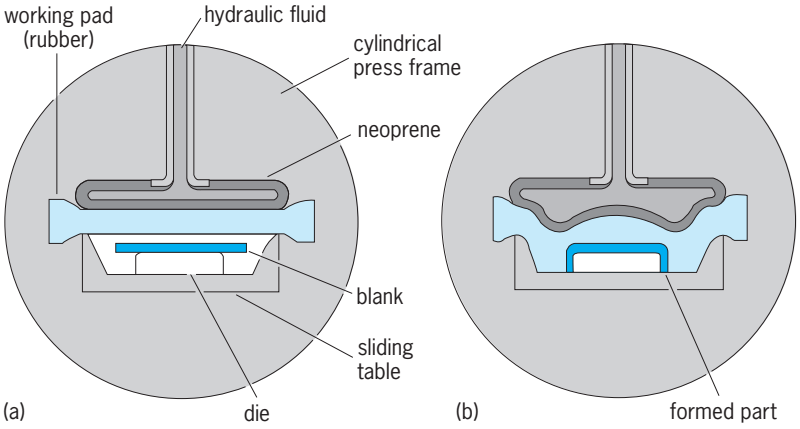


Fig. 3. Verson-Wheelon process. (*a*) Before forming. (*b*) After forming.

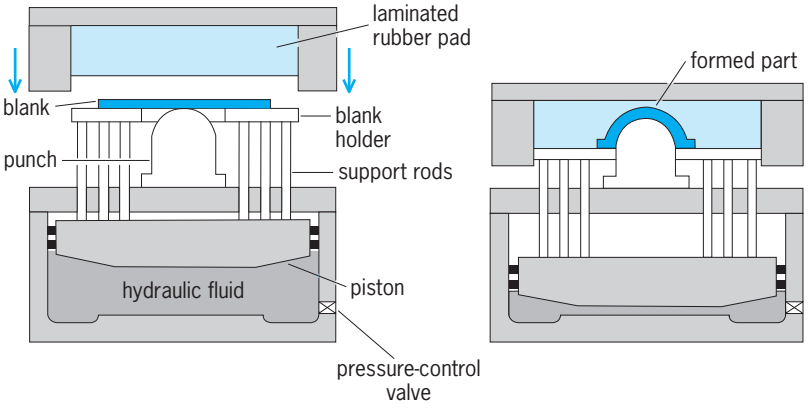


Fig. 4. Marform process.

The simplest of these processes is the Guerin process (**Fig. 2**). Auxiliary devices are also used in forming more complicated shapes. In the Verson-Wheelon process (**Fig. 3**) hydraulic pressure is confined in a rubber bag, the pressure being about five times greater than that in the Guerin process. For deeper draws the Marform process is used (**Fig. 4**). This equipment is a packaged unit that can be installed easily into a hydraulic press. In deep drawing of critical parts the Hydroform process (**Fig. 5**) is quite suitable, where pressure in the dome is as high as 15,000 lb/in.^{2} (100 megapascals). A particular advantage of this process is that the formed portions of the part travel with the punch, thus lowering tensile stresses which can eventually cause failure.

Bulging of tubular components, such as coffee pots, is also carried out with the use of a rubber pad placed inside the workpiece; the part is then expanded into a split female die for easy removal.

Deep drawing. The basic components of a deep-drawing operation are shown in **Fig. 6**. A great variety of parts are formed by this process, the successful operation of which requires a careful control of factors such as blank-holder pressure, lubrication, clearance, material properties, and die geometry. Depending on many factors, the maximum ratio of blank diameter to punch diameter ranges from about 1.6 to 2.3. *See* DRAWING OF METAL.

This process has been extensively studied, and the results show that two important material properties for deep drawability are the strain-hardening exponent and the strain ratio (anisotropy ratio) of the metal. The former becomes dominant when the material undergoes stretching, while the latter is more pertinent for pure radial drawing. The strain ratio is defined as the ratio of the true strain in the width direction to the true strain in the thickness direction of a strip of the sheet metal. The greater this ratio, the greater is the ability of the metal to undergo change in its width direction while resisting thinning.

Anisotropy in the sheet plane results in earing, the appearance of wavy edges on drawn cups. Clearance between the punch and the die is another factor in

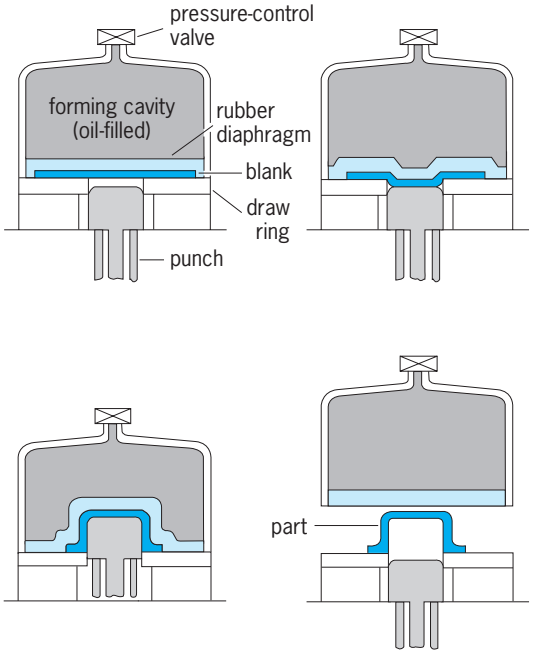


Fig. 5. Hydroform process.

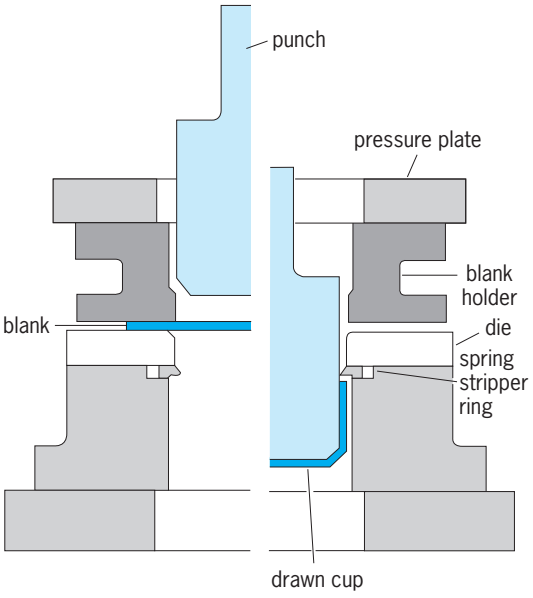


Fig. 6. Deep-drawing process.

this process; this is normally set at a value of not more than 1.4 times the thickness of the sheet. Too large a clearance produces a cup whose thickness increases toward the top, whereas correct clearance produces a cup of uniform thickness by ironing. Also, if the blank-holder pressure is too low, the flange wrinkles; if it is too high, the bottom of the cup will be punched out because of the increased frictional resistance of the flange. For relatively thick sheets it is possible to draw parts without a blank holder by special die designs.

Miscellaneous processes. Many parts require one or more additional processes; some of these are described briefly here. Embossing consists of forming a pattern on the sheet by shallow drawing. Coining consists of putting impressions on the surface by a process that is essentially forging, the best example being the two faces of a coin. Coining pressures are quite high, and control of lubrication is essential in order to bring out all the fine detail in a design. Shearing is separation of the material by the cutting action of a pair of sharp tools, similar to a pair of scissors. The clearance in shearing is important in order to obtain a clean cut. A variety of operations based on shearing are punching, blanking, perforating, slitting, notching, and trimming. *See* COINING.

Sheet-forming equipment. The most common equipment is a mechanical or hydraulic press. Presses may be single, double, or triple acting, depending on the number of slides to do a certain portion of the operation. Common mechanical presses are the eccentric, crank, knuckle joint, and toggle. The design of dies requires considerable experience because of the great number of problems that may arise in production. Tool and die materials for sheet forming are generally made of cast irons, cast alloys, die steels, and cemented carbides for high-production work. Nonmetallic materials such as rubber, plastics, and hardwood are also used as die materials. The selection of the proper lubricant depends on many factors, such as die and workpiece materials, and severity of the operation. A great variety of lubricants are commercially available, such as drawing compounds, fatty acids, mineral oils, and soap solutions. *See* HYDRAULIC PRESS.

Pressures in sheet-metal forming generally range between 1000 and 8000 lb/in.^{2} (7 and 55 MPa) (normal to the plane of the sheet); most parts require about 1500 lb/in.^{2} (10 MPa).

Spinning. This process forms parts with rotational symmetry over a mandrel with the use of a tool or roller. There are two basic types of spinning: conventional or manual spinning, and shear spinning (**Fig. 7**). The conventional spinning process forms the material over a rotating mandrel with little or no change in the thickness of the original blank. Parts can be as large as 20 ft (6 m) in diameter. The operation may be carried out at room temperature or higher for materials with low ductility or great thickness. Success in manual spinning depends largely on the skill of the operator. The process can be economically competitive with drawing; if a part can be made by both processes, spinning may be more economical than drawing for small quantities. *See* SPINNING (METALS).

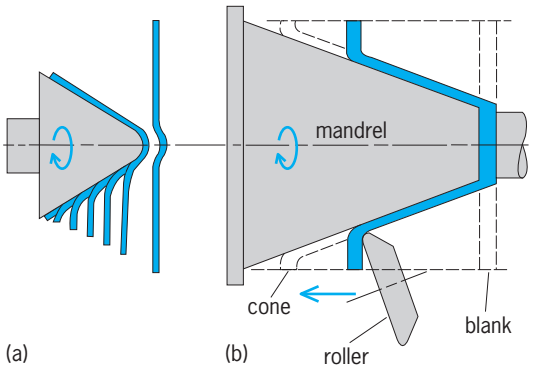


Fig. 7. Spinning processes of (*a*) conventional and (*b*) shear type.

In shear spinning (hydrospinning, floturning) the deformation is carried out with a roller in such a manner that the diameter of the original blank does not change but the thickness of the part decreases by an amount dependent on the mandrel angle. The spinnability of a metal is related to its tensile reduction of area. For metals with a reduction of area of 50% or greater, it is possible to spin a flat blank to a cone of an included angle of 30° in one operation. Shear spinning produces parts with various shapes (conical, curvilinear, and also tubular by tube spinning on a cylindrical mandrel) with good surface finish, close tolerances, and improved mechanical properties. *See* METAL COATINGS. Serope Kalpakjian

Bibliography. R. Baril, *Modern Machining Technology*, 1987; D. F. Eary and E. A. Reed, *Techniques of Pressworking Sheet Metal: An Engineering Approach to Die Design*, 2d ed., 1997; R. A. Lindberg, *Processes and Materials of Manufacture*, 4th ed., 1990.

Shellac

The lac resin (secreted by the lac insect) when used in flake (or shell) form. Shellac varnish is a solution of shellac in denatured alcohol.

Shellac varnish is used in wood finishing where a fast-drying, light-colored, hard finish is desired. Drying is by simple evaporation of the alcohol. Shellac varnish is not water-resistant and is not suitable for exterior coatings. When used as a finish, it has the distinct advantage that it remains soluble. When touch-up is required, it therefore merges completely with the original finish, and no scratches or worn spots show.

Natural shellac has a brown color and is often referred to as orange shellac. Bleaching with chlorine produces bleached or white shellac. Bleached shellac is slightly less durable and less hard than orange shellac but is lighter in color.

Shellac consists of a number of acids and esters and about 3% shellac wax, which is removed to make dewaxed shellac. Shellac varnishes are usually sold

on the basis of the number of pounds of shellac dissolved in a gallon of alcohol, as 3-, 4-, or 5-lb (1.4-, 1.8-, or 2.3-kg) cuts. Shellac varnish should be freshly prepared because it deteriorates on aging. Drying rates become slower as the number of pounds of shellac per gallon increases. *See* PAINT; SURFACE COATING; VARNISH. C. R. Martinson; C. W. Sisler

Bibliography. R. Lambourne, *Paints and Surface Coatings: Theory and Practice*, 1999; D. Stoye (ed.), *Paints, Coatings, and Solvents*, 2d ed., 1998.

Ship design

A process which translates a set of owner's requirements into the drawings, specifications, and other technical data necessary to actually build a ship. Naval architects lead the process, but engineers and designers with many other skills contribute. These other skills include marine engineering, structural design, and production engineering. The ship design process is iterative, and is subdivided into several phases during which the design is developed in increasing degrees of detail. Typically, the owner's requirements specify the mission that the new ship must perform and define such parameters as required speed, fuel endurance, and cargo weight and capacity. Generally, the cost to build and operate a ship is constrained by the prospective owner since the ship will be part of a transportation or industrial system, which must be profitable. The ship design process involves numerous trade-off studies in order to achieve the desired capability and, at the same time, stay within the established cost. *See* NAVAL ARCHITECTURE.

Mission requirements. Ships are designed and built to satisfy the mission requirements and constraints specified by the prospective owner. These requirements and constraints are unique to each ship being considered and form the basis for design development. For some ships, such as point-to-point cargo ships, the mission requirements can be simply stated; for example, "Transport 5000 20 ft ISO standard cargo containers at an average sea speed of 18 knots with 10,000 nautical miles between refuelings. On- and off-load the 5000 containers using shore-based cranes in less than XX hours." For other ship types, such as industrial ships performing missions at sea, the mission requirements are more complex. The requirements for a fisheries research vessel, for example, might specify the ability to catch fish using several different techniques, radiated noise limitations, required sonar performance, and several different aspects of maneuvering and seakeeping performance, such as low-speed stationkeeping and the ability to maintain a specified track over the sea floor in the face of cross currents, winds, and seas. *See* MERCHANT SHIP; OCEANOGRAPHIC VESSELS.

Generally, constraints as well as requirements are imposed on the designer by the prospective owner. Cost, both to design and build the ship and to operate it, is usually constrained. The two primary elements of operating cost are crew and fuel, so there is nearly always pressure on the designer to reduce crew size and fuel consumption. Physical constraints may also be imposed on the design related to construction, operational, or maintenance requirements. Weight or dimensional constraints may be imposed if the ship is to be built or maintained in a specific dry dock. Pier or harbor limitations may also impose dimensional constraints. Ship length may be limited by the requirement to tie up to a certain pier. Ship air draft (vertical distance from the water surface to the highest point on the ship) may be limited by the need to pass under bridges of a certain height. Ship navigational draft (vertical distance from the water surface to the lowest point on the ship) may be limited by the depth of a dredged channel in a particular harbor.

In addition to unique mission requirements and constraints, every ship must satisfy certain physical principles. The fundamental principles are that (1) the ship hull and superstructure must have adequate storage space, and (2) the ship must float at an acceptable waterline (draft neither too great nor too small) when it is fully loaded. A ship floats because the buoyant force acting vertically upward is equal to the total weight of the ship and its contents. Per Archimedes' principle, the buoyant force is equal to the weight of the water displaced by the hull and its appendages (propeller, rudder, and so on). This weight is termed the displacement. Another principle is that the ship must be statically stable; that is, when it is displaced from its equilibrium condition, it must tend to return to that condition. For example, when the ship is heeled to one side by a disturbing force such as a wind gust, it must tend to return to the vertical rather than continuing to roll and capsizing. The ship's hull must have sufficient strength to withstand the forces that will act upon it over a range of loading and sea conditions. The ship must possess sufficient propulsive power to achieve the desired speed even with a fouled bottom and in adverse sea conditions. In addition, it must generate sufficient electric power to satisfy the requirements of mission systems; ship machinery; heating, ventilation, and air conditioning (HVAC) systems; hotel; and other ship services. *See* ARCHIMEDES' PRINCIPLE; BUOYANCY; HYDROSTATICS.

Design approach. Ship design is an iterative process. The various synthesis and analysis tasks are repeated in increasing detail as the design is developed, as shown by the design spiral in **Fig. 1**. In the figure, each complete circuit around the center is intended to represent a complete set of the steps necessary to configure and assess a design concept. The steps are repeated in the second circuit, the third, and so on. Initially, assumptions are made as to the eventual size and configuration of the ship in order to get the process started. As the design proceeds, these initial assumptions are replaced by better data as more becomes known. The nature of the analysis work also changes. Initially, rough order-of-magnitude estimates are made using simplified procedures. Later, more detailed calculations are made using more complex procedures. The initial circuits around the spiral

are made quickly by one or two designers. Later in the process, many more persons are involved and the circuits are made more slowly. Figure 1 depicts an idealized design process. In fact, after the initial ship concept is developed, design work in the various design disciplines proceeds more or less in parallel, with periodic halts to integrate the results and assess overall progress.

In planning, managing, and describing the ship design process, a distinction is made between level I (total ship) design and level II (ship system) design. Level I design concerns the synthesis and analysis of total ship attributes such as hull form and general arrangement, and the assessment of total ship attributes such as weights and centers. Level II design concerns the synthesis and analysis of the major individual ship elements such as structure; propulsion machinery; electric power generation and distribution; ship control, navigation, and communication systems; auxiliary and mechanical systems, including piping and HVAC systems; and outfit and furnishings. Many outputs of level II design are inputs to level I design assessments. Examples of such outputs are data on weight, reliability, required crew size, cost, and risk.

Design integration is a major challenge. If the level II elements of a ship design are developed and optimized independently, the result will be a nonoptimum and disappointing level I total ship design. The elements of the design must be integrated so that they work well together. This will generally require that compromises be made in the level II system designs.

Design phases. The four traditional phases of ship design are concept, preliminary, contract, and detail design (Fig. 1). The phases correspond to milestones in the process used to develop and build a new ship; they permit periodic top-level reviews of the development status by the top management of both the design/build team and the owner (and enable the project to be halted if desired); and they reflect the fact that the design team changes size and composition during design development.

Concept design. The first design phase helps the owner and the designer to define the top-level requirements for the new design and to develop a ship concept design that reflects those requirements. Typically, the top-level design requirements are defined through an iterative dialog between the owner and the designer. The designer will develop numerous sketch designs, called ship feasibility studies, that explore parametric variations of the key requirements, which are major design drivers in terms of ship size, cost, and performance. Costs are estimated for each option. The prospective owner will review the sketch designs and their estimated costs and make tentative decisions to narrow the range of investigation. The designer will then develop a second set of design alternatives, representing a more focused set of options, and present them to the owner for review. By means of such an iterative dialog, the major requirements for the ship are defined, and a feasibility study, representing the agreed-to requirements, is developed. Subsequently, this brief study is further developed into a concept design.

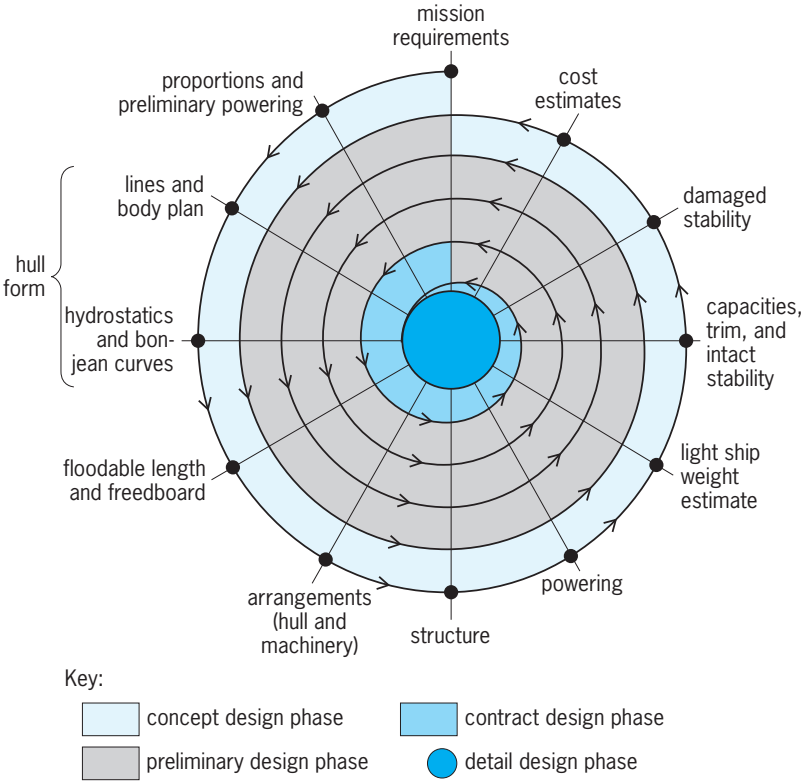


Fig. 1. Basic design spiral, showing the iterative ship design process.

Preliminary design. During the second design phase, the major ship characteristics which drive ship cost and performance are emphasized. For each characteristic, trade-off studies are performed in which a range of alternatives is studied. Issues typically addressed during this phase include hull shape (proportions and form coefficients), propulsion plant [prime movers: number and type(s), number of shafts], general arrangement options, topside configuration, structural configuration, hull material, and electric plant configuration. The initial strategy for building the ship is developed. At the conclusion of this phase, a budget-quality cost estimate is prepared. Major ship characteristics, such as the principal hull dimensions, installed propulsion power, and cargo capacity, are not expected to change upon completion of this phase. *See* PROPELLER (MARINE CRAFT).

Contract design. During this phase, the design is further developed and refined. Drawings and specifications are produced which form an integral part of the contract for ship detail design and construction. Hydrodynamic model tests are run, including resistance, propulsion, seakeeping, and maneuvering tests. Hull form and appendage shape definitions are completed. The primary hull structure is defined, including stiffener and plate scantlings, based on refined load estimates and finite element analyses. Distributive systems are defined in diagrammatic form. The equipment in ship operating and living spaces is selected and arranged. The ship weight and center-of-gravity estimate, which accounts for the weight

and location of each major item in the ship, is further refined. Intact and damage stability analyses for a range of ship loading conditions are refined. The ship specifications, a primary product of this phase, define quality standards for the hull and hull outfit as well as performance standards for each item of ship machinery and equipment. The specifications also define the tests and trials that must be performed by the shipbuilder as the ship nears completion. The specified test criteria must be satisfied for the ship to be considered acceptable. *See* FINITE ELEMENT METHOD.

Detail design. The fourth and final design phase is devoted to completing the ship design in every detail and to translating the completed design into easily understood work packages that can be used directly by the artisans who will build the ship without further translation or interpretation. The work packages are the fabrication and installation instructions to the ship fitters, welders, sheet metal workers, pipefitters, and electricians. In the detail design phase, especially in the development of work packages, great emphasis is given to production engineering, to make the ship construction process as speedy and efficient as possible. *See* ENGINEERING DESIGN; PRODUCTION ENGINEERING; SHIPBUILDING.

Level I (total ship) design elements. Level I design includes such topics as hull form and general arrangement, weights and centers, stability, hydrodynamic performance, crew size, cost, reliability, supportability, operability, and risk.

Hull form. Naval architects describe hull forms by their principal dimensions (length, L; beam, B; draft, T; and depth, D), proportions (ratios of length to beam, L/B; length to depth, L/D; and beam to draft, B/T), and by form coefficients, which characterize the hull shape. L is the hull length between perpendiculars at or near the ends of the ship and may also be the length on the waterline at which the ship is designed to float when fully loaded. This waterline is called the design waterline, or DWL. B is the hull beam, generally the beam at the DWL midship (at the mid-length of the ship). T, the ship draft, is measured amidship vertically down from the DWL to the lowest point on the hull. D, the hull depth, is also measured vertically at midship from the lowest point on the hull to the main or weather deck. Freeboard is the height of the hull above the water surface and is equal to the difference between D and T, or $D - T$.

The hull form for a new ship design must satisfy the ship's mission requirements as well as the physical principles mentioned above. The hull must provide adequate internal volume, buoyancy when floating at or near the DWL, and stability. In addition, the freeboard must be sufficient to provide adequate stability at large heel angles, and adequate reserve buoyancy in the event that the hull suffers damage and a portion of the hull is flooded. The freeboard must also be adequate to keep the weather deck reasonably dry in rough seas. Hydrodynamic performance is a major consideration in defining the hull form. Resistance and propulsion, seakeeping, and maneuvering behavior must all be considered. *See* SHIP POWERING, MANEUVERING, AND SEAKEEPING.

First among the considerations is resistance and propulsion. The hull proportions and form coefficients must enable the ship to be propelled efficiently at the desired top speed as well as at the specified economical cruising speed. These two speeds may be very different, as on a warship, or may be the same, as on a point-to-point cargo ship. The two most significant form coefficients from the standpoint of resistance and propulsion are the displacement-length ratio and the prismatic coefficient. The displacement-length ratio indicates the hull fatness or slenderness. It is the ratio of the hull's displacement volume divided by the hull length cubed. Traditionally, in English units, the displacement-length ratio is defined as the hull displacement in long tons of salt water divided by hull length in feet/100. In this form, the ratio varies from a low of about 10–15 for a racing shell used by college crews, up to about 500 for a harbor tug. The prismatic coefficient C_p indicates the bluntness or fineness of the underwater hull form and is defined as the displacement volume divided by the volume of a circumscribing prism with length L and cross section equal to the largest cross section of the ship's immersed form, at or near midlength. A very fine hull has a C_p value of about 0.55, while a very full hull such as that of an oil tanker might have a C_p of 0.90 or higher.

The speed-length ratio (SLR) is an important concept when discussing ship resistance and propulsion. SLR is defined as the ship speed in knots divided by the square root of the hull length L in feet. The two primary components of hull resistance to forward motion are frictional resistance and wavemaking resistance. Frictional resistance is due to the friction between the hull's wetted surface and the water passing over it. Wavemaking resistance is due to the waves generated by the moving hull; creating these waves requires energy input. Below an SLR of about 0.75, virtually all of the hull resistance is frictional resistance. As the SLR increases above 0.75, the importance of the wavemaking component or resistance steadily increases. At an SLR of about 1.0, the frictional and wavemaking components of resistance are about equal in importance. Above an SLR of 1.0, the wavemaking component dominates. At an SLR of about 1.30, a conventional displacement ship is moving at its "hull speed." At this speed, the ship's speed-power curve is very steep and it is very difficult to increase speed further, that is, a great deal of power is required to increase speed even a small amount.

Frictional resistance is a function of the hull's wetted surface and hull smoothness. The wavemaking portion of resistance is affected by the hull shape, and its importance thus increases as the SLR increases above about 0.75. For SLR values in the range of 0.8–1.1, the optimum C_p is low, in the 0.55–0.57 range. For SLR ratios approaching 1.30 and higher, the optimum C_p increases to about 0.64. At such high speeds, the effect of displacement-length

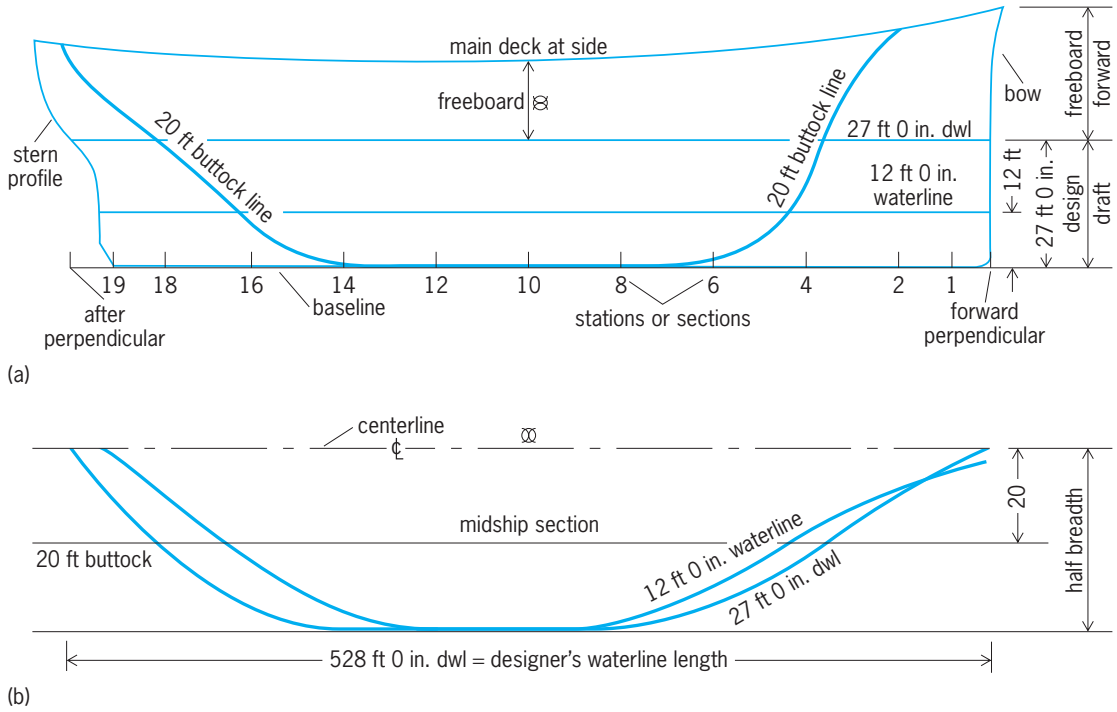


Fig. 2. Geometry of a ship. (*a*) Profile view. (*b*) Half-breadth plan. These views, along with the transverse view given in Fig. 3, make up the lines drawing, or lines of the vessel. 1 ft = 0.3 m; 1 in. = 2.5 cm.

ratio is dominant and low values are desired. *See* FRICTION.

For the naval architect, length is the most important hull dimension. It has a greater effect on construction cost than any other dimension. It also has a powerful effect on hull resistance characteristics through its effect on both the SLR and the displacement-length ratio. Trade-off studies are performed during the design process to find the optimum balance between reduced length to save construction cost and increased length to improve hull resistance characteristics. Hull beam is the second most costly dimension in terms of construction cost, and has the most important effect on transverse static stability. Thus, selection of the optimum beam is also a subject of study during design development. Hull depth is important because of its effect on freeboard and on the strength of the ship's hull girder.

The ship's hull form is delineated in a so-called lines drawing. This drawing depicts the hull form in three views: the elevation or profile plan, looking at the hull from the side; the half-breadth plan, looking down from above; and the body plan, looking at the hull from the bow or stern (**Figs. 2** and **3**). Each view depicts the lines of intersection of the hull with a series of parallel planes. In the profile view, these planes are parallel to the hull centerplane, the plane of symmetry of the hull, and the lines of intersection are called buttocks. In the half-breadth plan, the planes are parallel to the design waterplane, and the lines of intersection are called waterlines. In the body plan, the planes are parallel to the transverse midship plane of the hull and are perpendicular to the other two sets of planes. The lines of intersection are called sections.

Ordinarily, only one side of the hull is depicted in the lines drawing since most hulls are symmetrical about their longitudinal centerplane. Using the three projections of the lines drawing in conjunction with one another, it is possible to determine the position in space of any point on the hull surface. The lines drawing is used to perform many of the analyses during ship design development. Also, many other ship drawings are based on the hull geometry defined by the lines drawing.

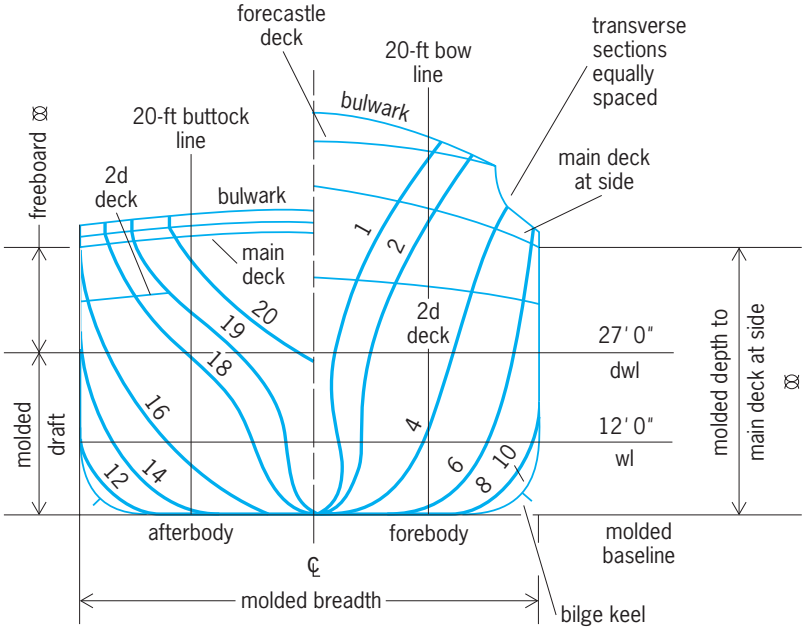


Fig. 3. Transverse view, or body plan, of ship. 1 ft = 0.3 m; 1 in. = 2.5 cm.

Subdivision. The internal hull volume is subdivided into compartments by horizontal surfaces called decks, and vertical surfaces called bulkheads. Bulkheads are either transverse or longitudinal, depending on whether they are perpendicular to or parallel to the ship's centerplane.

Decks permit the internal space to be efficiently used for living and working spaces. Decks may be omitted from main machinery spaces or cargo holds where large heights are required. Bulkheads serve several purposes. They add to the strength of the hull, restrict fire and flooding, support the decks, and separate compartments with dissimilar functions.

Longitudinal bulkheads are sometimes used to add structural strength or to bound vital spaces such as shaft alleys. Generally they are avoided, however, since in the event of hull damage a watertight longitudinal bulkhead restricts flooding to one side of the ship. This can result in a large heel angle after damage and may cause the ship to capsize.

The primary function of transverse bulkheads is to restrict the longitudinal extent of flooding in the case of hull damage. Transverse bulkheads are watertight up to the bulkhead deck. Ships are generally designed so that the flooding, resulting from a specified extent of damage, will not immerse the hull so deeply that the bulkheads restricting the flooding are immersed. If this occurs, seawater can flow over the tops of the bounding bulkheads into adjacent undamaged compartments, and the resulting progressive flooding can cause ship loss, as occurred with the *Titanic*.

The spacing of transverse bulkheads is a critical design decision. If the spacing is too great, the ship will not be able to survive flooding beyond the capability of the ship's pumps to dewater the flooded space. Too close a spacing may interfere with cargo handling and stowage, passenger accommodations, machinery arrangements, or other ship functions. This could cause the ship to be unsuccessful commercially. Naval architects develop floodable length curves to aid in locating transverse bulkheads (**Fig. 4**). The floodable length at any point along the hull is the length of flooding, centered at that point, which will cause the ship to sink down to a waterline 3 in. (7.6 cm) below the bulkhead deck. Compartment permeabilities are reflected in the floodable length curve, as noted in the figure. The permeability of a compartment μ is the volume of flooding water that can enter that compartment, expressed as a percentage of the compartment volume. An empty cargo hold or ballast tank would have a permeability of close to 100%, while a full ballast tank would have a permeability of zero. Most ship spaces other than tanks have permeabilities in the 60–80% range. There are international regulations that specify the subdivision requirements for various ship-type categories, such as passenger ships, oil tankers, and dry cargo vessels. The passenger ship regulations are driven by passenger safety concerns, while the oil tanker regulations are driven by pollution concerns.

General arrangement. Naval architects define the arrangement of external and internal ship features in a drawing termed the "general arrangement." This drawing typically shows the exterior of the ship in several views. For the ship's interior, the inboard profile (a profile view along the ship's centerplane) and plan views of each deck in the hull and superstructure are shown. The general arrangement shows all internal tanks and compartments, including passageways and vertical trunks provided for access. The development of the ship's general arrangement during the design process is one of the principal ship design tasks, and the drawing is used to help guide and control the entire design process. During arrangement development, top priority is given to the ship's mission requirements and the associated payload. After the hull form and principal subdivisions have been defined, blocks of space are assigned to accommodate the various ship functions. Alternatives are often considered and evaluated, such as propulsion machinery midships versus aft. Later, the principal access routes within the ship are outlined, and individual compartments are sized and located. Additional detail is developed as the design proceeds.

Capacities. Available cargo hold and tank capacities (fuel, fresh water, ballast) are calculated from the hull form and general arrangement. The results are

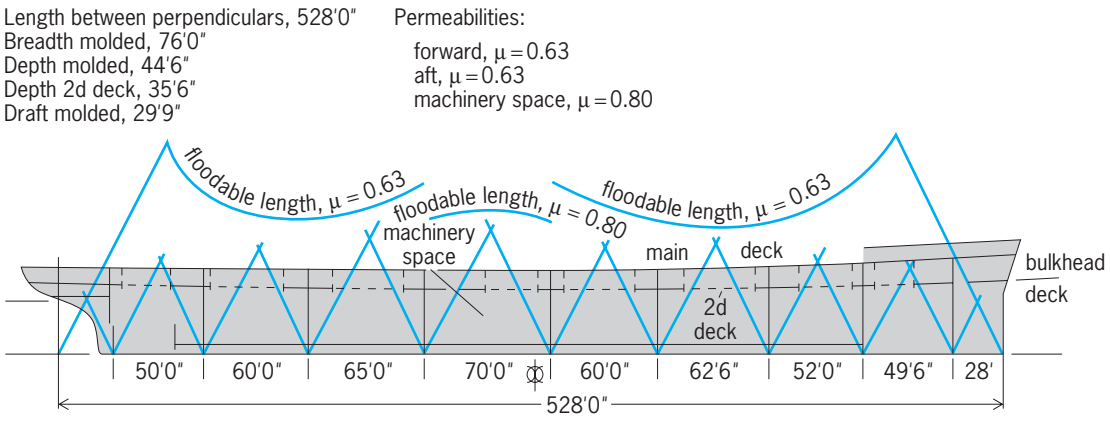


Fig. 4. Subdivision diagram and floodable length curve. 1 ft = 0.3 m; 1 in. = 2.5 cm.

compared with the required capacities, and adjustments are made if necessary.

Weights and centers. Careful analysis of ship weight and center of gravity is essential to a successful ship design. The fully loaded weight of a ship has two elements: the light ship weight, the weight of the ship itself; and the loads, called the deadweight. The light ship weight includes the weight of the hull structure; propulsion machinery; electric plant; ship navigation, control, and communications systems; piping and ventilation systems; cargo-handling equipment; outfit and furnishings; and all other equipment and fittings necessary to make the ship complete. The deadweight consists of all the variable loads that can be moved onto or off the ship, including cargo, fuel, crew and their personal effects, stores, and fresh water. The light ship weight plus the deadweight equals the full load weight or displacement, the weight which must be supported by the buoyancy of the hull. Early in the design process, weights and centers are estimated using similar "parent" ships or historical data, which are analyzed to create estimating relationships or algorithms. Margins are applied to the estimates to allow for unknowns and estimating errors. As the design proceeds, more and more of the total weight can be calculated based on equipment lists and drawings, such as the structural drawings. As the quality of the estimate improves, the margins applied can be reduced. The total weight and center-of-gravity position are calculated for a range of ship operational loading conditions, varying from light to full load. The results are used to calculate trim and list when the ship is at rest pierside and also to analyze stability and seakeeping characteristics. *See* CENTER OF GRAVITY.

Stability. A ship must be stable under all normal conditions of loading and performance at sea. This means that when it is inclined from its at-rest position by a disturbing force, such as a wind gust, it should return to the at-rest position when the force is removed. Stability is addressed in both the longitudinal and transverse directions; for most surface ships, transverse stability is a greater concern. Stability is an issue both when the hull is intact and after the hull has been damaged and partially flooded. Thus both intact and damage stability are addressed by the naval architect during the design of a ship.

Figure 5 shows the cross section of a ship that has been inclined from its normal upright position by a disturbing force. The weight of the ship W acts down at the ship's center of gravity, point G in the figure. When the ship is upright, the buoyant force, equal to W, acts up at the center of buoyancy, point B. After the ship is inclined, the center of buoyancy shifts toward the low side to point B'. The buoyant force is now offset from the weight force by the distance GZ. A couple is formed acting to right the ship. The righting moment is $W \times GZ$. The distance GZ is termed the righting arm. For a given center-of-gravity position G, the righting arm is a function of the ship's hull form. GZ can be computed for a range of heel angles and plotted versus the heel angle to yield the ship's righting arm curve. Generally, GZ will increase with heel angle until a maximum is reached and then decrease as heel is further increased until it reaches zero. At this point the ship will capsize.

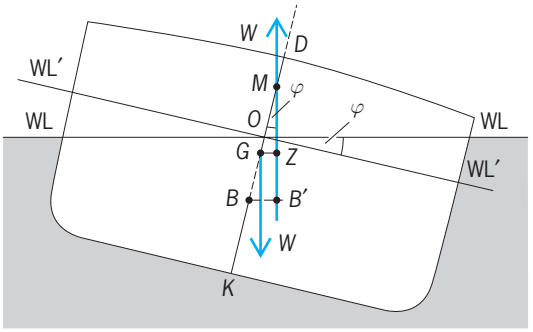


Fig. 5. Diagram of inclined ship, showing points, lines, and angles that are important in determining the ship's stability. W′L′ = inclined water line, WL = position of waterline on upright ship. Other symbols are explained in text.

The range of stability refers to the range of heel angles over which the righting arm is positive. Increasing hull beam will increase the height of the righting arm curve. Increasing freeboard will increase both GZ (at large heel angles) and the range of stability. For a given hull form, GZ is a function of the height of point G, that is, the vertical position of the ship's center of gravity. As G is raised, GZ decreases. If GZ is negative, that is, if Z is to the left of G in Fig. 5, then the ship is unstable—the moment is tending to capsize the ship rather than return it to the upright. Intact stability is evaluated for several assumed hazards. These include a beam wind (perpendicular to the direction that the ship is sailing) combined with ship rolling due to beam seas, and the case of all the ship's crew and passengers crowding to one side of the ship.

The point M in Fig. 5 is called the metacenter. It is the point about which the center of buoyancy is rotating. For small heel angles, M lies on the ship's centerplane at the point where the line of action of the buoyant force intersects it. This fact permits naval architects to analyze stability at small heel angles, called initial stability, using simplified methods. Initial stability is checked very early in the design process. Later, large-angle intact stability and stability after damage are analyzed. For initial stability, the distance GM is called the transverse metacentric height. If M is above G, GM is positive and the ship is stable. If M is below G, then GM is negative and the ship is unstable. The height of M above the ship's keel, distance KM in the figure, is a function of the ship's hull form. KM has two parts: the distance KB, the height of the ship's center of buoyancy above the ship's keel, plus the distance BM, the height of the metacenter above the ship's initial center of buoyancy. Distance BM is called the metacentric radius. BM is equal to I/V, where I is the transverse moment of inertia of the ship's waterplane at the water surface and V is the immersed hull volume. Both KB and BM are functions of the ship's hull form and can be readily computed in early-stage design (or estimated using simplified relationships). I can be expressed

as a constant $\times LB^3$, and V can be expressed as a constant $\times L \times B \times T$ (B is the hull beam in these expressions). Thus I/V varies as beam squared. This is why a small beam increase has such a strong effect on GM and hence initial stability. Note in Fig. 5 that the righting arm GZ is equal to GM times the sine of the heel angle ϕ for small heel angles (when M is on the ship's centerplane). *See* MOMENT OF INERTIA.

A ship's natural roll period varies inversely with the square root of its transverse GM. Increasing GM increases the ship's stiffness and improves its initial stability but reduces its roll period. Shorter roll periods mean greater lateral accelerations and increased discomfort for shipboard personnel and passengers. Thus the naval achitect must find the best balance between stability and seakindliness when establishing GM in a new design. Generally, a GM equal to about 6% of the maximum beam on the waterline in the full load condition is appropriate for commercial vessels. Passenger ships and warships have higher GM values in order to meet more stringent damage stability requirements. Damage stability is analyzed after one or more adjacent compartments have been flooded. The presence of a beam wind combined with rolling is assumed. Naval architects use the terms one-, two-, or three-compartment ships, referring to the number of adjacent compartments that can be flooded while still satisfying the established stability criteria.

Freeboard and load lines. A load line is a mark on the side of a ship's hull which indicates the deepest draft to which it can be legally loaded in fresh or salt water in various operating regions and seasons of the year. International conventions have established regulations for establishing load lines. The regulations are based on considerations of reserve buoyancy, hull structural strength, and the provisions for making the hull watertight in extreme conditions, such as strong, tight hatch covers. Reserve buoyancy is the volume (potential buoyancy) of the watertight hull above the load waterline. Credit is also taken for weathertight superstructure. The regulations establish the minimum acceptable freeboards for the various operating conditions. The naval architect must ensure that the ship being designed will satisfy the load line regulations when carrying the desired payload. For passenger ships carrying more than 12 passengers, in addition to the above regulations, there is a limit termed the subdivision draft. The subdivision draft is a function of the spacing of internal watertight transverse bulkheads and hence the number of compartments that can be flooded in a damaged condition before reserve buoyancy and stability disappear. If the subdivision draft is less than the limiting draft per the load line regulations, the subdivision draft governs.

Tonnage. Tonnage measurement is the result of continuing efforts by government authorities to establish a uniform means of measuring ships' relative size and carrying capacities for the purposes of taxation. Since about 1860, tonnage has generally been a measure of internal volume or cubic capacity expressed in units (tons) of 100 ft^3 (3 m^3). Gross tonnage refers to the entire internal volume, less certain exempted spaces. Net tonnage is the tonnage remaining after the non-earning spaces, such as machinery and crew living spaces, have been deducted from the gross tonnage. Net tonnage is intended to be a measure of a ship's earning ability, and is the basis on which charges, such as canal transits and port entry, are determined. Until recently, different countries and canal authorities have had their own tonnage measurement regulations. In 1969, the International Convention on the Tonnage Measurement of Ships established a simplified method of determining tonnage. The Convention replaced the myriad of special tonnage systems previously used.

Hydrodynamic performance. Satisfactory hydrodynamic performance is critical to a new ship's success, both in terms of mission effectiveness and of economic success. Thus the evaluation of hydrodynamic performance is an important element of ship design. The three primary elements of hydrodynamic performance are calm water resistance and propulsion, maneuvering, and seakeeping. The latter includes ship motions as well as the effects of rough water on resistance and propulsion and the structural loads transmitted to the hull. Hull size and shape have strong effects on resistance and ship motions. These effects are primary considerations in early-stage design when hull size and gross shape are being determined. Propulsion, maneuvering, and other aspects of seakeeping get greater attention later in the design process. To date, naval architects have been unable to satisfactorily predict hydrodynamic performance by analytical means. The complexities of the phenomena involved have simply been too great. Thus, in early-stage design, rather crude analytical predictions of hydrodynamic performance have been made. Extensive model testing has traditionally been used in the middle phases of design to evaluate attractive design alternatives and make the best possible performance predictions. Advances in computational fluid dynamics (CFD), coupled with reductions in the cost of high-speed computing, are rapidly changing this picture. Today, model testing is generally used later in design only to confirm sophisticated analytical predictions of hydrodynamic performance; in the future, physical model tests may be phased out entirely. *See* COMPUTATIONAL FLUID DYNAMICS; HYDRODYNAMICS.

Propulsion. The shaft power (SHP) required to propel a ship at a specific speed in calm water is equal to the power required to tow the ship at that speed (EHP) divided by the quasi-propulsive coefficient, P. C. EHP equals the tow rope pull (total ship resistance) times the velocity converted to power. P. C. equals the open-water propeller efficiency times a factor called the hull efficiency, to account for the effect of the propeller on the ship resistance and the effect of the hull and appendages on the inflow velocity to the propeller. Thus, to estimate the SHP at a specific speed, the following must be

estimated: total resistance, open-water propeller efficiency, and hull efficiency. Total resistance consists of bare hull resistance plus appendage resistance plus still air drag (negligible for most ships). Bare hull resistance consists of frictional resistance plus residual resistance, which is the sum of wavemaking resistance plus separation and eddymaking resistance. Appendage drag and hull efficiency can be estimated from self-propelled model test results for similar ship designs. Frictional resistance can be calculated using rather simple relationships based on turbulent flow, knowing the immersed wetted area of the hull. Still air drag is significant only for high-speed ships, and can be estimated based on the projected frontal area of the ship topsides and superstructure and an assumed drag coefficient. Open-water propeller efficiency can be estimated using propeller model test data or lifting line theory. This leaves residual resistance, which can be estimated from the model test results for a similar hull form, the results of model tests on a standard series of hull forms, the results of regression analyses of model tests of a large number of hull forms, analytically computed wavemaking resistance multiplied by a correction factor based on experience, or on model tests of the new hull form. In early-stage design, speed-power estimates are based on combinations of analytical and empirical methods, using previous model tests of similar hull forms. Later in the design process, model tests are generally performed for the new design. These tests include bare hull and appended hull resistance and self-propulsion tests. *See* TOWING TANK.

Seakeeping. Ship motions are the direct effect of the seaway on a ship. In higher sea states, these motions may result in bottom slamming (when the ship's keel emerges from the water and then violently reenters) and deck wetness. Slamming and deck wetness lead to voluntary speed reductions to reduce their frequency and severity. Ship motions themselves result in accelerations which cause loss of crew effectiveness and passenger seasickness, and may even damage equipment. Excessive motions and accelerations may prevent certain missions from being carried out, such as helicopter or small boat operations.

The complex motions of a ship at sea are analyzed by decomposing them into six components, three linear and three rotational, with reference to orthogonal axes through the ship's center of gravity. The linear components are surge (fore and aft motion), sway (side-to-side motion), and heave (vertical motion). The rotational components are roll (rotation about the longitudinal axis), pitch (rotation about the transverse axis), and yaw (rotation about the vertical axis).

Three of the motion components—pitch, heave, and roll—have the dominant effects on seakeeping performance. Pitch and heave are strongly coupled; their effects are most severe in head seas (waves coming from directly ahead). They cause vertical accelerations, deck wetness, and bottom slamming. Once the overall hull size and shape have been established, little can be done to ameliorate pitch and heave, since the forces necessary to do this are very large for conventional hull forms.

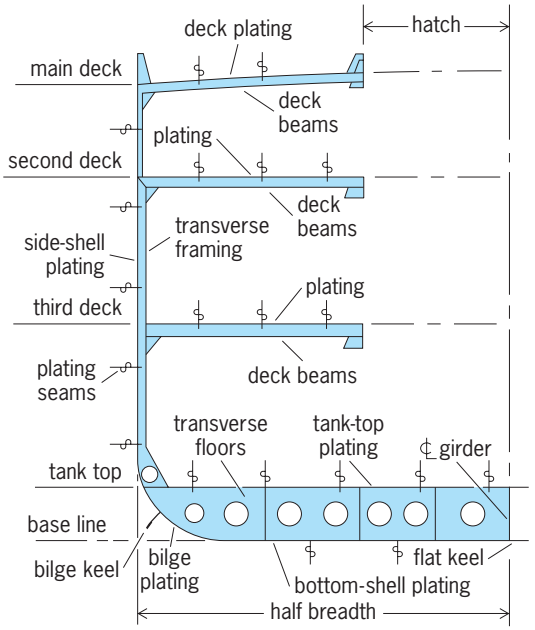


Fig. 6. Schematic midship section showing structural arrangement of plating and transverse framing within a steel merchant ship.

Roll motion, on the other hand, causes lateral accelerations and is caused, and can be reduced, by relatively small forces. For this reason, several practical means of reducing roll motion have been developed and employed. Bilge keels are long projecting structures, fitted at the turn of the bilge (**Fig. 6**), which are inexpensive and quite effective at reducing roll amplitudes over the entire speed range. Active or passive antiroll tanks can be effective at most speeds and headings, but they have significant space and weight penalties, are noisy, and can increase roll at some headings and ship loading conditions. Active antiroll fins are very effective at mid-to-high ship speeds but are costly and ineffective at low speeds.

Seakeeping behavior is assessed analytically in early-stage design and may also be evaluated by means of model tests later in design. Seakeeping assessments are an important part of the hull form trade-off process. Roll stabilization studies are often performed.

Maneuvering. The third primary element of ship hydrodynamic performance is maneuverability. Important aspects of maneuverability are directional stability, turning ability, and responsiveness to course changes, all at significant forward speeds, and maneuverability at low speeds, at pierside, for example. If a ship is directionally stable, it will tend to return to its original course after being forced off course by a minor disturbance such as a wind gust. Turning ability is expressed in terms of the turning diameter at various speeds and rudder angles. Maneuverability is influenced more by the ship's appendages, such as

the rudder and skeg, than it is by the basic hull proportions, although these do have a significant influence. For this reason, in sizing and shaping the hull in early-stage design, speed-power and seakeeping considerations generally play a greater role than maneuverability. Initially, appendages are often sized using empirical relationships or by ratiocination from similar successful designs. Later, analytical models are used to prepare refined performance predictions. Finally, physical model tests are generally performed to confirm the predicted performance. Low-speed maneuverability is often enhanced by installing thrusting devices of various sorts at the bow or stern which can be used to create strong side forces at low forward speeds.

Crew size. The total life cycle cost of a ship is the sum of its design and build (D&B) cost plus its operating cost over the ship's life. Crew cost is invariably the largest fraction of the total life cycle cost, greater even than the D&B cost. Thus ship owners and naval architects give great attention to minimizing crew size. Three scenarios must be considered in establishing crew size: normal ship operations, ship maintenance performed at sea, and damage control in the event of fire or flooding. Federal safety regulations and union rules are important factors for commercial ships. Automation of functions traditionally performed by humans is a technique used to reduce crew size. However, such automation is costly and adds maintenance and crew training requirements which must be considered. Thus crew size versus automation trade-off studies are a typical element of the design process.

Reliability. A ship and its systems must be highly reliable from the standpoints of both safety and good business. During design, reliability analyses are performed for critical ship systems in order to ensure that sufficient reliability is designed in. Techniques for improving reliability include paying more for more reliable system components or increasing redundancy to permit continued operation in the event of a failure.

Supportability. A ship must be properly supported to be capable of sustained, reliable operations. Properly trained crew members must be available when required to fill voids. The ship must receive timely replenishment of consumables, other stores, fresh water, fuel, other liquids, and repair parts. A routine for regular ship maintenance must be established and provisions made to accomplish unscheduled repairs when needed. Supportability considerations must be an integral part of the design process. Decisions must be made on when, where, and how frequently the ship will be resupplied. A ship maintenance plan must be developed which addresses how the ship will be maintained and by whom (for example, maintenance at-sea versus ashore, in-place repairs versus unit replacement, and maintenance by the ship's crew versus shore support personnel). The ship's storerooms and workshops must reflect the maintenance and supply support plans. Adequate internal accesses must be provided to permit stores handling and necessary maintenance actions to be performed. Crew training requirements must be defined and training materials developed for unique ship systems.

Cost. The prospective owner of a new ship will nearly always have cost limits which must not be exceeded. Thus throughout the design process, from the development of initial sketchy concepts through detail design, trade-offs are made in which projected costs are a major consideration. There are two major elements of the life cycle cost of a ship: the acquisition (or design and build) cost, and operating costs, that is, the costs of items such as fuel, crew salaries, supplies, maintenance, and insurance. D&B cost is further split into two categories: recurring and nonrecurring costs. Nonrecurring costs are the costs of design and development associated with building the first ship of a new class, while recurring costs are incurred for each ship of a multiship class. There is often a trade-off between acquisition and operating cost; for example, operating cost may be reduced by spending more initially. A more costly but more efficient propulsion plant might reduce fuel consumption and operating cost. The ship designer must know the owner's D&B and operating cost thresholds, as well as the owner's preference as to the relative weight to be given to D&B versus operating cost in performing design trade-offs.

Risk. Risk is an issue which must be addressed in each new ship design. Technical risk relates to performance, both performance at delivery and over the long term. Will the completed ship meet the owner's performance requirements? Program risk includes cost and schedule as major elements. Will the ship be delivered on time and at a cost at or below the owner's threshold? For the ship designer, there is a generally a trade-off between innovation and risk. Innovative design solutions may have the potential of improving performance or reducing cost, but they generally increase technical or program risks because they have not been fully developed or have not been proven previously in service. Generally ship owners are extremely conservative and reluctant to be the first to try new systems, equipment, or design approaches. A bold owner can gain a significant market advantage with an innovative design but runs the risk of disappointment and financial losses. The ship designer must be sure to understand the owner's willingness to accept risk before starting a design project. Methods have been developed to formally assess and manage risks during the design process so as to reduce the chances of rude surprises after ship completion. *See* RISK ASSESSMENT AND MANAGEMENT.

Level II (ship system) design elements. The total ship consists of major level II elements or systems which can be classified as follows: structure, propulsion plant, distributive (electric, piping, and HVAC) and mechanical systems, and outfit and furnishings. The ship design process includes the design of each of these level II elements as well as their functional and physical integration into a balanced whole.

For information on propulsion plant design *See* MARINE ENGINE; MARINE ENGINEERING; MARINE MACHINERY.

Structure. The ship's hull structure can be thought of as a hollow box beam. The main deck and the shell plating on the exterior of the hull comprise the outer surface of the beam, and the decks and bulkheads within the hull subdivide and strengthen the beam. The hull structure must withstand the overall loads acting on the hull, which tend to bend and twist the beam, as well as the local loads acting at any point. The overall hull loads result from differences between the local weight and buoyancy along the length of the hull in still water, and loads resulting from operation in rough seas (wave loads). Local loads result from sources such as water pressure, breaking seas, the weight of cargo, and loads due to components of the ship itself, such as heavy machinery, masts, and superstructure.

The hull of a typical steel merchant ship (Fig. 6) is a complex structure. The side shell and decks are supported locally by framing members called frames and beams, respectively. The bottom is supported by deep beams called floors. The purpose of the framing members is to maintain the adjacent plate membranes in their designed contours and positions relative to one another, as the plates are subjected to internal, external, and in-plane loads. The local framing members are, in turn, supported by more substantial structure. An important structural design decision is whether the deck and side shell frames should run longitudinally or transversely. Cost, weight, cargo handling, and maintenance considerations all enter into the decision. Stiffener spacing is an important variable. Wider stiffener spacings add weight but reduce cost by reducing the number of structural parts and the amount of welding required. Generally, ship hulls are built of steel, but other materials are sometimes used. In terms of steel grade, higher grades of steel are more costly to purchase and fabricate but may save weight, which is of great importance in some designs. *See* LOADS, TRANSVERSE; STRESS AND STRAIN; STRUCTURAL DEFLECTIONS.

The loads acting on a ship's hull are complex, and some cannot be accurately determined. Thus the structural designer must be guided by experience as well as theory. Classification societies (the American Bureau of Shipping in the United States) have developed rules for the structural design of commercial ships based on a combination of theory and experience. These rules can be used to define the structural dimensions used for a new design, as well as to gain an understanding of preferred construction practices. *See* STRUCTURAL DESIGN.

Distributive systems. The distributive systems in a ship are the electrical; piping; and heating, ventilation, and air conditioning (HVAC) systems. The design process for these systems consists of (1) analyzing the demands or loads which the system must support, (2) postulating one or more system concepts, as defined by a 1-line block diagram that depicts the system elements and the connections between elements, (3) sizing the major equipment elements of the system such as the electric generators, pumps, compressors, fans, and so on, (4) sizing the connections between system elements, such as the electric power cables, pipes, and vent ducts, and (5) developing the physical arrangement of the system within the ship. Historically, this was done using two-dimensional drawings, but increasingly, three-dimensional computer-aided design (CAD) models are employed. As with most other aspects of ship design, the process is an iterative one. For example, the required size of a pump will be dependent on the details of the input and output pipe sizes and runs; these are not defined until quite late in the design process. Thus, estimates are made of the effects of these details to enable an estimate of pump size to be made earlier in the process. *See* COMPUTER-AIDED DESIGN AND MANUFACTURING.

Mechanical systems. A ship's mechanical systems include handling systems for anchors, boats, and cargo. For each such system, requirements must be defined, system concepts postulated, and analyses performed to size system components. Major equipment items are generally selected from vendor options.

Outfit and furnishings. Outfit and furnishings include items such as deck coverings, insulation, joiner bulkheads, ceilings, doors, lighting, shelving, port lights, and compartment furnishings. The ship's living and working spaces must be arranged and furnishings selected.

Warship design. Warship design is necessarily more complex than most commercial ship design. Warships carry and support sophisticated combat systems which include weapons systems and command, control, and communications systems. The design of these systems is a tremendous challenge. *See* NAVAL ARMAMENT.

Survivability is a key consideration in warship design. Low signatures, decoys, and active self-defense systems are employed to avoid being hit by an enemy mine or projectile. Armor and other passive protection features, as well as extensive subdivision and a high standard of damage stability, are built into the ship to contain damage and prevent ship loss if the ship is hit. Damage control features are provided to fight fires and control flooding after a hit, with the goal of restoring some mobility and combat capability.

The top speed of most warships is significantly higher than the speeds of most commercial ships. On the other hand, warships operate at top speed a small fraction of the time. Most of their at-sea time is spent at much lower transit or cruising speeds. Thus, unlike most merchant ships, warship propulsion plants must operate at two widely different power levels; this has a significant effect on the plant design. Weight must be controlled to achieve high speeds at practical power levels. Thus warship designs typically incorporate lighter-weight components, systems, and structure which are more costly than their commercial equivalents since they employ more expensive materials and are more difficult to

fabricate. One way to reduce structural weight, for example, is to reduce stiffener spacings, which increases the number of structural elements and hence fabrication costs.

Warship crews are typically much larger than merchant ship crews, and many naval ship crew members have little experience. These two factors have many ramifications for warship design, including more extensive living facilities, more trash and waste to handle while satisfying antipollution regulations, and greater challenges for crew training to properly operate and maintain the ship.

U.S. Navy warships are designed for long periods at sea. Thus they have large storerooms for supplies and spare parts, as well as extensive capabilities for self-maintenance. In addition, the ships are capable of being replenished with fuel, ordnance, and supplies while under way at sea. This is done by helicopters shuttling to and from the supply ship or by high-line transfers while connected to the supply ship, both ships proceeding on parallel courses. *See* NAVAL SURFACE SHIP.

Peter A. Gale

Bibliography. J. B. Caldwell and G. Ward (eds.), *Practical Design of Ships*, 2 vols., 1992; E. V. Lewis (ed.), *Principles of Naval Architecture*, 3 vols., 1988; H. Schneekluth and V. Bertram, *Ship Design for Efficiency and Economy*, 2d ed., 1998; R. Taggart (ed.), *Ship Design and Construction*, 3d ed., Society of Naval Architects and Marine Engineers, 1980.

Ship nuclear propulsion

Nuclear reactors for shipboard propulsion can be of any type used for the production of useful heat. Nuclear power is particularly suitable for vessels which need to be at sea for long periods without refueling or for powerful submarine propulsion. Only the pressurized water reactor (PWR) and the liquid metal reactor (LMR) have actually been applied to operating vessels. The pressurized water reactor has been most widely applied since it uses a readily available coolant and has a relatively simple cycle and control system and a large industrial and technical base. The supposed advantages of a liquid metal reactor (compactness, fast response, and higher propulsion plant efficiency) have not been proven in application, and liquid metal reactors are not now in marine service. For basic information on fission reactors *See* NUCLEAR POWER; NUCLEAR REACTOR; REACTOR PHYSICS.

In all the shipboard nuclear power plants that have been built, energy conversion is based on the steam turbine cycle, and that portion of the plant is more or less conventional (see **illus.**). There are two types in use: a steam turbine geared to a fixed-pitch propeller (called a geared turbine), and a steam turbine generator whose output drives an electric motor connected to a propeller (called a turboelectric unit). Any energy conversion process that converts heat into mechanical energy could be used to propel a ship. For example, a closed-cycle helium gas turbine has been studied, but none has been built for ship propulsion. *See* MARINE ENGINE; MARINE MACHINERY; PROPELLER (MARINE CRAFT); SHIP POWERING, MANEUVERING, AND SEAKEEPING.

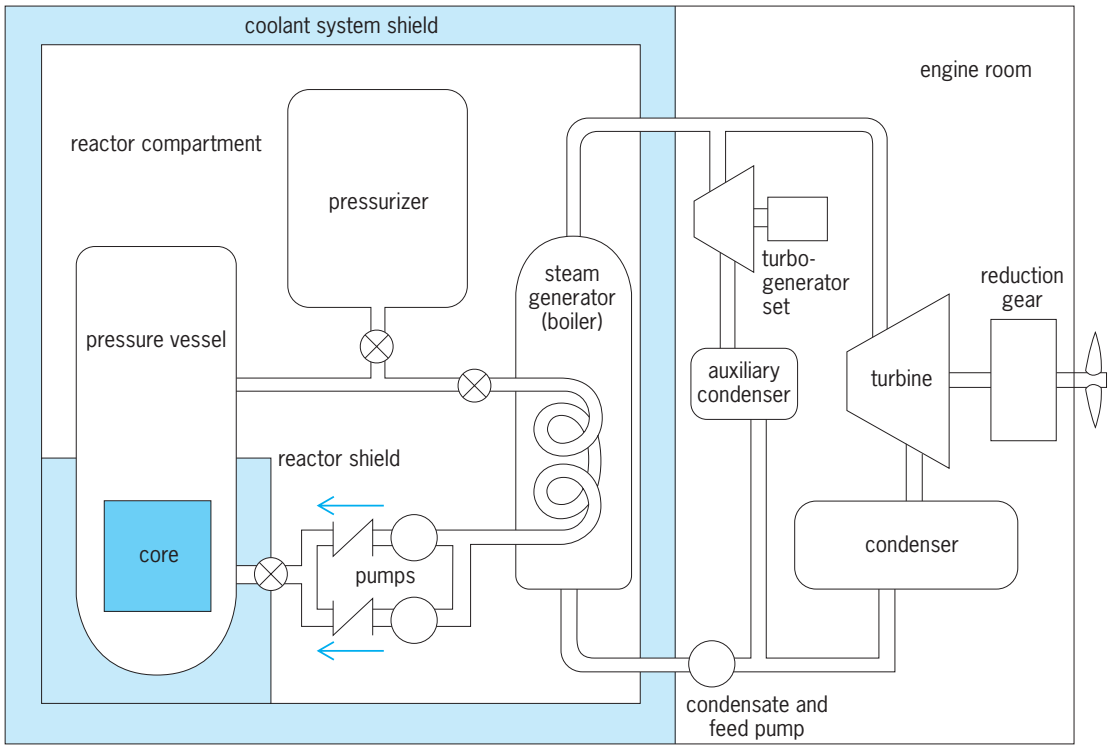
Civil applications. Three early civilian applications, the *Savannah* (United States), the *Otto Hahn* (Germany), and the *Mutsu* (Japan), were technology demonstrations. They did establish that ships using nuclear power could not compete economically with ships using cheap fossil fuel for cargo/passenger applications on the oceans of the world.

In contrast, the use of nuclear propulsion has proven successful in the unique environment of the Russian Arctic. The long time between refueling and the energy required for ice breaking made this application comparatively practical. The icebreaker *Lenin* remained in service for 30 years from 1957, though a major accident is believed to have occurred in one or more of its reactors, which led to the replacement of the three reactors with two others in 1967–1970. The operational use of the *Lenin* led to larger icebreakers, the *Arktika* class of ships, the first of which was launched in 1975. These vessels have two pressurized water reactors and are used in deep Arctic waters to transport cargo (typically ores) and as passenger excursion vessels. (The *Arktika* was the first surface vessel to reach the North Pole.) *See* ICEBREAKER.

For use in shallow water such as estuaries and rivers, several shallow-draft *Taymyr*-class icebreakers were built in Finland and fitted with a single pressurized water reactor in Russia. Eight or nine icebreakers and one nuclear-powered container ship have operated. The nuclear-powered container ship *Sevmorput*, which began operation in 1988, had one pressurized water reactor; it was designed to meet International Maritime Organization (IMO) standards and has transported cargo to the Russian Far East and Cuba.

Military applications. Work on naval nuclear propulsion started in the late 1940s, and the first test reactor, a United States pressurized water reactor, started up in 1953. The first nuclear-powered submarine, the *Nautilus* of the United States, went to sea in 1955. The first Soviet nuclear-powered submarine went to sea 3 years later. *See* SUBMARINE.

Most naval reactors have been pressurized water reactors, which differ from commercial pressurized water reactors producing electricity in that they must have high power density and a small volume, and therefore run on highly enriched uranium (more than 20% uranium-235). They also have a long reactor core life, so that refueling is not needed for 10 years or more (depending on the amount of high-speed operation). This lifetime is extended by the relatively high enrichment of the uranium and by the incorporation of burnable poisons in the reactor assembly, which are progressively used up as the fuel is depleted and fission product nuclear poisons accumulate. By careful design these two effects cancel out, and long reactor life is attained. Unlike their power station counterparts, military reactors operate at varying power levels, which also influences the design of the reactor.



Schematic diagram of a typical pressurized nuclear propulsion plant (relative sizes of the various components are not indicated). (*Society of Naval Architects and Marine Engineers*)

United States. Most of the United States nuclear-powered ships have used a pressurized water reactor with a geared turbine. Several submarines have been turboelectric. All of the submarines were built as either ballistic-missile-carrying (SSBN) or hunter-killer (SSN) submarines. Many of the later SSNs carry cruise missiles. The second nuclear-powered submarine used a liquid metal reactor (sodium) which was replaced by a pressurized water reactor since the sodium design provided no operational advantages and had several disadvantages as a military vessel. The nuclear-powered surface ships, comprising carriers and cruisers, have used multiple pressurized water reactors of various sizes depending on the shaft horsepower required for the application.

The United States has built nearly 200 nuclear submarines. A small deep-diving nuclear-powered research vessel was built by the Navy for oceanographic work. The United States has also built nine nuclear-powered aircraft carriers, all of which (including the first, the *Enterprise*, commissioned in 1961) are still in operation. The nine nuclear cruisers have all been retired or will be soon. *See* NAVAL SURFACE SHIP.

Soviet Union. Most of the Soviet military nuclear-powered ships were pressurized water reactors with geared turbines. The Soviet Union built many classes of submarines (SSNs, SSBNs, and SSGNs) of differing designs depending on desired speed, diving depth, and size. Many of the early submarine classes used two reactors for reliability. The largest submarines in the world, the *Tyfun* class of SSBNs, used two reactors (each in its own pressure hull).

The Soviet Union built four nuclear-powered cruisers. The cruisers had an additional oil-fired steam system to provide superheated steam for a more efficient steam turbine plant. There was a small pressurized water reactor designed to be retrofitted into existing diesel submarines. There were also several deep-diving ocean research minisubmarines, each with a single small pressurized water reactor. Several classes of turbo-electric submarines used a liquid metal reactor with a lead-bismuth alloy as the primary coolant. This allowed a higher steam pressure and a more compact primary and secondary plant. This design had a number of operational and safety problems, and all units have all been removed from service. The Soviet Union built nearly 250 submarines. They also built a nuclear-powered communications ship (with two pressurized water reactors) which operated briefly in the Pacific (probably in support of the Soviet space program).

United Kingdom. The United Kingdom has built about 20 nuclear-powered submarines. All of them have used pressurized water reactors with geared turbines. Both SSN and SSBN types were built. The earliest application was an adaptation of a United States pressurized water reactor, but later submarines have used an indigenous reactor design.

France. All of the French military nuclear-powered submarines have used the pressurized water reactor with a turbo-electric drive. There are two types, SSBN and SSN. France has built more than 10 submarines and a nuclear-powered aircraft carrier using two reactors (based on the current SSBN pressurized

water reactor) driving propellers through geared turbines.

China. The Chinese nuclear-powered submarines, both SSBN and SSN, are believed to be based on early Soviet pressurized water reactor designs and are turbo-electric.

Other countries. The only other country that may be involved in submarine nuclear propulsion is India, which has operated a borrowed Russian nuclear-powered submarine of obsolete design and has announced development of an indigenous design.

Status. At the end of the cold war in 1989, there were more than 400 nuclear-powered submarines operational or being built. By 1994 the number was halved. More than 150 had been taken out of service, and many planned for construction were stopped. The size of the miliary fleets depends on a number of factors. The trend is toward a smaller, more modern fleet. The extra expense of operating and maintaining a nuclear-powered ship has placed pressure on military/national budgets. Underwater detection technology has made many of the early submarines obsolete before the end of their design lives, and the cruisers were not updated with the latest antimissile technology. Arms reduction treaties eliminated many of the SSBNs. *See* ANTISUBMARINE WARFARE.

Currently there are less than 200 nuclear-powered ships with one or two reactors each. Over 5000 reactor years of naval operation have been accumulated. Most are submarines, but the ship types range from icebreakers to aircraft carriers. All of the countries named above continue to build nuclear submarines of various types and, in the case of the United States, an aircraft carrier. There have been no announced plans for new nonnaval nuclear-powered ships.

Alan R. Newhouse

Bibliography. *Jane's Fighting Ships*, 1997–1998 edition; *The United State Naval Nuclear Propulsion Program*, USDoE/USDoD, June 1996.

Ship powering, maneuvering, and seakeeping

Powering, maneuvering, and seakeeping constitute the three central areas of ship hydrodynamics. Although this article focuses on these areas in connection with displacement ships, that is, vessels supported entirely (or almost entirely) by hydrostatic forces, basic concepts of powering, maneuvering, and seakeeping are also critical to an understanding of high-speed craft, such as planing boats, hydrofoil craft, and air-cushion vehicles, even if the physical phenomena governing dynamically supported craft are somewhat different. *See* AIR-CUSHION VEHICLE; HYDROFOIL CRAFT; MERCHANT SHIP; NAVAL ARCHITECTURE; NAVAL SURFACE SHIP; SHIP DESIGN; SHIPBUILDING; SUBMARINE.

Powering

Even in antiquity a body of practical knowledge existed among shipbuilders regarding the relationship between hull proportions and speed. In the ancient world, with propulsion limited to human power acting through paddles and oars, the speed advantages of increased length, a high ratio of length to breadth (beam), and hull slenderness were well known, but without a formal theoretical basis. In naval architecture, the term "hull slenderness" has a special meaning; it is often used to describe the relationship between length and displacement. The volumetric coefficient of the hull, for example, is a measure of slenderness defined as displaced volume divided by the length cubed. Generally, lower values of volumetric coefficient are associated with low-resistance forms: long, slender hulls which are more easily driven through the water. In the era of human power, increased length also increased the space available for paddlers or oarsmen, thus making the effect on speed even greater.

With the introduction of sails, first to supplement oars or paddles, and culminating in pure sailing ships, the effects of a ship's stability on sail-carrying power tended to reduce the benefits of slenderness and high length-to-beam ratio. Consequently, even very fast sailing ships were by no means as slender as many earlier galleys or other oared ships. To some extent, the balance was shifted back again by the advent of steam propulsion and the connection between the powering of a ship and its economic performance. The power required to move a ship determined the size, power, and cost of its engines, the cost of fuel per voyage and, because of the weight and volume required for coal, the range and cargo capacity of the ship. By the middle of the nineteenth century, the need for an adequate theory of ship powering was apparent, and the work of William Froude, in particular, marked the beginning of the scientific study of powering. The field of powering is divided into two related issues: resistance, the study of forces opposed to the ship's forward speed, and propulsion, the study of the generation of forces to overcome resistance.

Resistance of ships. A body moving through a fluid experiences a drag, that is, a force in the direction opposite to its movement. In the specific context of a ship's hull, this force is more often called resistance. Resistance arises from a number of physical phenomena, all of which vary with speed, but in different ways. These phenomena are influenced by the size, shape, and condition of the hull, and other parts of the ship; and several of them are interrelated. Frictional resistance is produced by the viscosity of the fluid and the velocity of the ship. Frictional resistance is directly related to the ship's wetted surface area. Therefore, reducing wetted surface, other things being equal, is favorable to reduced resistance. Because of frictional forces, the pressure in a stream of fluid is also lower on the afterbody than on the forebody of the ship. Form drag is the result of this pressure drop. When the afterbody shape is particularly blunt, such as on an immersed transom stern, large-scale eddies are visible, and the form drag is very high. For this reason, form drag is sometimes also identified as eddy-making drag. Because both

frictional resistance and form drag are related to viscous losses, they are often grouped together as viscous resistance. The distinction is that skin friction would exist even on a flat thin plate moving edgewise through the fluid, while form drag exists because of the shape of the body. *See* VISCOSITY.

When a ship moves at or near the surface of the sea, forces are produced which are not the result of viscosity, but the result of surface waves created by pressure variations around the ship. Wavemaking resistance reflects the energy carried away from the ship by the wavetrain. The shape of the wave surface also affects the flow around the hull, and the actual wetted surface as well. Similarly, eddies and other viscous flow phenomena near the stern can affect the formation of the wavetrain. For these reasons, wavemaking and viscous resistance are not entirely independent, although it is convenient to consider them separately.

In addition to the resistance of the bare hull, a ship's appendages can produce significant amounts of drag. Typical appendages include rudders, exposed propeller shafts, shaft bearings and struts, bossings where shafts emerge from the hull, skegs, bilge keels and stabilizing fins, sonar domes, and auxiliary thrusters. Appendage resistance can be a significant part of the total resistance of the ship, often 3–5% for single-screw ships, and 15% for twin-screw ships with exposed shafts and struts.

Because the density of air is only about 1/800 that of seawater, the air resistance of the superstructure and above-water hull of the ship is usually not a very large component of the total. However, superstructures of most ships are quite bluff (broad and flattened) in shape, and the drag contribution may not be negligible. Typically, still air resistance, that is, air resistance due solely to the ship's speed, is a significant part of total resistance only for very fast ships. In addition to still air resistance, wind can add both a direct drag increase and indirect effects. For example, a wind not from dead ahead causes a ship to make leeway (that is, drift slightly sideways through the water), and may also require use of the rudder to hold course. Both of these effects cause additional resistance. Bare hull resistance plus appendage resistance, and generally with still air resistance included, is often referred to as resistance fully appended.

Viscous resistance. Basic to the understanding of viscous resistance is the notion of the boundary layer. This is a relatively thin layer of water (thin with respect to ship dimensions), adjacent to the hull surface, in which viscous forces dominate the flow. Osborne Reynolds, as a result of studies involving flow in pipes, introduced in 1883 a dimensionless ratio of inertial forces to viscous forces in a fluid system. The Reynolds number is often written as $\mathrm{Re} = VL/\nu$, where V is the velocity of flow, L is the length of the system or body under consideration, and ν is the kinematic viscosity of the fluid [given by $\nu = \mu/\rho$, where μ is the (dynamic or absolute) viscosity and ρ is the fluid mass density]. The Reynolds number largely determines the characteristics of the boundary layer, including its thickness, velocity and pressure gradients, turbulence, and friction coefficient. The frictional resistance of a body is given by $R_f = {}^1/_2\rho S V^2 C_f$, where ρ is the mass density of the fluid, S is the wetted surface, V is the velocity, and C_f is the friction coefficient, which is a function of Reynolds number only. This function has been determined experimentally and has been given many empirical formulations, of which the present standard used in ship design is $C_f = 0.075/[\log_{10}(\mathrm{Re}) - 2]^2$. *See* BOUNDARY-LAYER FLOW; REYNOLDS NUMBER.

Because form drag is directly related to viscous phenomena, it is usually expressed as a multiplier of the friction coefficient, the form drag factor. There is a wide variation in form drag factor for hulls of differing proportions, but a typical range of values is from 1.05 for a slender, sharp-ended hull to 1.25 or even higher for bluff forms such as tankers.

Wavemaking resistance. Unlike viscous resistance, wavemaking resistance is characterized by behavior which changes dramatically with speed. Water waves, including those produced by a ship's forward motion, propagate with a speed related to their wavelength. At a very low speed, no significant waves are produced and the ship's entire resistance is viscous in origin. It can also be shown, although it is not as intuitively obvious, that wavemaking vanishes in the limit of very high speed, generally too high a speed to be of practical interest for displacement ships. Again, however, the ship's resistance is mainly viscous, although spray also becomes a drag contributor. At intermediate speeds, wavemaking resistance is extremely important. Depending on the hull form, wavemaking often determines the maximum speed that a displacement hull can reach.

Waves exist because a water surface, subject to an initial disturbance, cannot be in static equilibrium; gravity forces the surface to oscillate about its static level. The dimensionless number which characterizes wavemaking resistance is the Froude number, the ratio between inertial and gravitational forces in a fluid system. The Froude number may be written as $\mathrm{Fr} = V/(gL)^{1/2}$, where V is the velocity, g is the acceleration of gravity, and L is the length of the ship, generally taken at the waterline. The form of the Froude number is highly important; it implies that the wave systems generated by geometrically similar ships of various lengths are "in scale" when the ship speeds vary with the square root of length. (A quantity proportional to Froude number, but not dimensionless, appears in many English and American sources. This is the speed-length ratio, defined as $V_k/L^{1/2}$, where V_k is the speed in knots, and the length L is in feet.) *See* FROUDE NUMBER.

In his studies of ship resistance, about 1870, William Froude assumed that the total resistance of a ship, or of a model of the ship, can be decomposed into frictional and residuary resistance components. The frictional resistance of a ship or model was approximated by that of a flat plate of equal area and length. Residuary resistance, then, was obtained by deducting frictional resistance from total resistance.

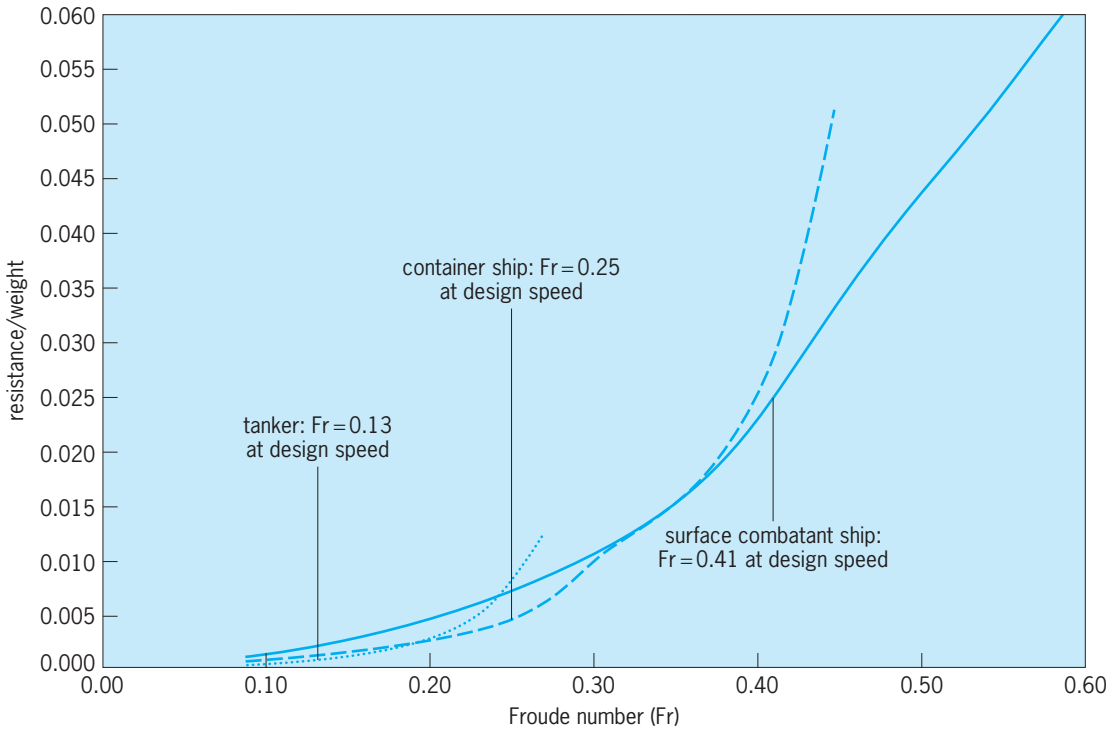


Fig. 1. Comparative resistance of three hull forms intended for different speed regimes. The tanker has displacement of 350,000 metric tons (344,000 long tons), dimensions of 366 × 50 × 22.2 m draft (1200 × 164 × 72.8 ft draft), and design speed of 15 knots or 7.7 m/s (Fr = 0.13). The container ship has displacement of 63,000 metric tons (62,000 long tons), dimensions of 274 × 32.2 × 10.7 m draft (900 × 105.6 × 35.1 ft draft), and design speed of 25 knots or 13 m/s (Fr = 0.25). Surface combatant ship has displacement of 11,300 metric tons (11,100 long tons), dimensions of 195 × 19.5 × 6.5 m draft (640 × 64.0 × 21.3 ft draft), and design speed of 35 knots or 18 m/s (Fr = 0.41).

The essence of Froude's contribution is that to capture the correct wavemaking behavior, a model of a ship must be tested at a corresponding speed to the full-scale ship, that is, at the same Froude number. At corresponding speeds, the model's wavemaking resistance can then be scaled up to the full-size ship simply by multiplying by the cube of the linear scale ratio, that is, $(L_{ship}/L_{model})^3$. Frictional resistance must then be recalculated at the ship's Reynolds number, and added to residuary to obtain full-scale ship bare hull resistance. This simple approach, barring the details of how form drag is to be accounted for, is the foundation of ship resistance model testing.

The influence of hull form on wavemaking is profound and rather complicated. Analytical predictions of wavemaking provide a conceptual understanding of the wavemaking phenomenon, but purely mathematical treatments have not yielded reliable design methods for all types of ship forms. Instead, in early stages of a ship design, estimates of wavemaking resistance are based on past data, using systematic series data (results from a series of model tests on a closely related family of hull forms), or regression methods (results derived from a statistical summary of tests for a range of ships, not necessarily systematically related, but generally similar to the hull being designed). As a hull design is developed, it is increasingly common practice to obtain more detailed wavemaking estimates using computational fluid dynamic methods, with final verification by model testing of the design hull. *See* COMPUTATIONAL FLUID DYNAMICS; TOWING TANK; WATER TUNNEL.

The two most important elements in determining a ship's wavemaking resistance are slenderness and whether the hull's volume is spread out into the ends or concentrated nearer amidships. Other form parameters, such as the sharpness of the bow and stern waterline endings, and the shape of the hull sections, also are known to affect both wavemaking resistance and form drag. Ships designed for different purposes and speed regimes have markedly dissimilar hull forms, and their resistance characteristics are quite different (**Fig. 1**).

Unlike viscous losses, the waves generated by a ship's hull may be partially canceled at a particular speed by incorporating hull features which produce waves of opposite phase. A widely used feature of this type is the bow bulb, often seen on ships of moderate to high Froude number, such as container ships, and some naval ships. Bulbous bows are also found on slower ships, operating at low Froude number, such as oil tankers; however, in these applications the effect of the bulb is related to moderating local wave-breaking phenomena around a bluff bow, rather than overall wave cancellation.

Hull roughness correction and correlation allowance. The friction coefficient discussed above applies to a hydrodynamically smooth surface, typically a freshly painted model surface. Consequently, when resistance predictions are based on model test extrapolation,

an additional correction is added to account for differences between the typical surface condition of a ship's hull and an ideal smooth surface. This is sometimes referred to as a hull-roughness correction, or δC_f. Often, however, a roughness correction is included as part of a more general correction, termed a correlation allowance, which encompasses all differences between model test and ship trial conditions not explicitly contained in the friction coefficient. Consequently, the correlation allowance involves not only the resistance of the ship but propulsion as well, and is not independent of specific model-test and trial practices.

In service, the frictional resistance of a ship is obviously affected by changes in surface roughness. In fact, excessive hull roughness and fouling by marine organisms can increase the powering requirements dramatically, increase fuel consumption, prevent efficient powerplant operation, and ultimately impair economic performance. Maintaining the smoothness of hull surfaces is a major concern in ship operation.

Propulsion. Many devices have been used to propel a ship. In approximate historical order they include paddles, oars, sails, draft animals (working on a canal towpath), paddlewheels, marine screw propellers, vertical-axis propellers, airscrews, and waterjets. A key distinction is whether or not propulsive forces are generated in the same body of fluid that accounts for the main sources of the ship's resistance, resulting in hull-propulsor interaction.

Any propulsor can be understood as a power conversion device. Delivered power for a rotating propulsor is the product of torque times rotational speed. The useful power output from the system is the product of ship resistance times ship speed, termed effective power. The efficiency of this power conversion, often termed propulsive efficiency, is of great importance. *See* BOAT PROPULSION.

Wake, thrust deduction, and hull efficiency. The efficiency of a propeller working in an undisturbed uniform inflow is called open-water efficiency. This is a function of propeller characteristics alone: diameter, pitch, blade geometry, revolutions per minute (rpm), and speed of advance, that is, the average velocity of flow into the propeller. For several reasons, however, when a propeller works behind a ship's hull the overall efficiency can depart significantly from the open-water efficiency. *See* PROPELLER (MARINE CRAFT).

Working near the hull, the propeller's speed of advance is different from the speed of the ship through the water. To a great extent this difference is due to the presence of the hull's viscous boundary layer, but the velocity is also influenced by the form of the stern, and in some cases, the presence of the ship's wave system. The ratio of speed of advance V_a to ship speed V is given by $V_a/V = (1 - w)$, where w is the wake fraction, representing the amount of fluid velocity change caused by the hull at the propeller's location. Wake fraction typically varies from values near zero (for twin-screw ships with large propellers extending well outside the hull boundary layer, such as naval combatants) to values of 0.3 or more (for ships with relatively full sterns and propellers deeply embedded in the hull's viscous wake, such as large single-screw tankers).

At the same time, the action of the propeller influences the flow over the after portions of the hull, compared with conditions without the propeller at work. Typically, the effect of a screw propeller near the stern is to reduce the pressure acting on the afterbody. Consequently, the propeller must generate a thrust somewhat in excess of the hull's resistance. The ratio of thrust to resistance is denoted by $T/R = 1/(1 - t)$, where t is the thrust deduction. Typically, thrust deduction is somewhat smaller than wake fraction for single-screw ships, and of the same order or slightly larger for twin-screw.

The ratio of effective power to delivered power involves both wake and thrust deduction. Effective power is defined as $P_e = RV$, where R is the total resistance and V is the ship's speed. The power transferred from propulsor to fluid, the so-called thrust power, is defined as $P_t = TV_a$. From the definitions of wake, thrust deduction, and effective power, this can be rewritten as $P_t = [R/(1 - t)][V(1 - w)]$.

The quantity $(1 - t)/(1 - w)$ is called hull efficiency. It is usually denoted by η_b, and can also be written in terms of power as $\eta_b = P_e/P_t$. It is noteworthy that for single-screw ships, η_b generally exceeds 1. For full-sterned hulls in particular, such as tankers, with propellers located deep in the wake, values of η_b as high as 1.25 to 1.3, or more, are by no means uncommon. Values of hull efficiency greater than 1 do not violate overall conservation of energy; they merely reflect the fact that the ship's hull has transferred energy into the fluid wake which can be partially recovered by the propeller.

Wake adaptation. In open water, the ratio between thrust power and delivered power is simply the open-water efficiency, $\eta_o = P_t/P_d$. However, the distribution of velocities in a ship's wake varies considerably at different radii from the shaft line, and consequently it is often possible to design a propeller to suit the wake distribution of a given hull significantly better than a propeller designed for open flow. The result is that such a wake-adapted propeller design may have a rather low value of η_o when tested in open-water conditions, but is actually more efficient in the wake for which it has been designed. The ratio between the propeller efficiencies in the "behind" and open-water conditions is a measure of the propeller's wake adaptation to a given hull, and is usually called relative rotative efficiency, η_r. Like hull efficiency, its value may exceed unity, but only slightly.

Propulsive efficiency is often defined as effective power divided by delivered power. This can be expressed as a product of the propeller's open-water efficiency, the hull efficiency, and the relative rotative efficiency, that is, $\eta_{\text{propulsive}} = P_e/P_d = \eta_o\eta_b\eta_r$. Mechanical losses between the propeller and the engine include frictional losses in shaft bearings, and in other portions of the transmission system, such

as couplings and gears. These losses are sometimes included as an additional factor in the efficiency, the transmission efficiency, η_t. In this case, the propulsive efficiency is defined as $\eta_{\text{propulsive}} = P_e/P_{\text{brake}} = \eta_o\eta_b\eta_r\eta_t$, where P_{brake} is the brake power of the propulsion engine or motor.

Propulsive effects of fouling. Increased surface roughness of propeller blades can impair efficient powerplant operation to an even greater extent than hull roughness. Blade roughness increases the torque required to turn the propeller at a given rpm, while at the same time blade lift forces and propeller thrust begin to degrade, even at a constant rpm. The result is a significant decrease in propulsive efficiency, even for a constant required thrust. However, when this occurs concurrently with excessive hull fouling, which increases the required thrust, the results are compounded.

Maneuvering

Maneuvering (more generally, ship controllability) includes consideration of turning, course-keeping, acceleration, deceleration, and backing performance. The field of maneuvering has also come to include more specialized problems of ship handling, for example, the production of sideways motion for docking or undocking, turning in place, and position-keeping using auxiliary thrusters or steerable propulsion units. In the case of submarines, maneuvering also includes depth-change maneuvers, either independently or in combination with turning.

Turning. When a ship's rudder is turned from its neutral position, it forces the stern toward the outside of the turn (**Fig. 2**). As a result, the hull is forced to operate at a drift angle, that is, with an angle between the ship's centerline and the path of its center of gravity. It is the hull's side force which actually moves the ship in its curved path, not the rudder. The rudder must provide sufficient moment to start the turn expeditiously and then to balance other moments which would tend to return the ship to a straight course (for example, moments generated by a fixed skeg or by the propellers).

Typically, a ship's tactical diameter is approximately three to five ship lengths. Smaller tactical diameter is generally associated with larger rudder area, larger maximum rudder angle, and higher rudder rate (the angular speed with which the rudder can be moved by the steering engine). Typically, maneuverable ships may have rudders sized as large as 2% or even 3% of the lateral plane area of the hull (approximately the product of length times draft). Many merchant ships, however, have rudders considerably less than 1% of lateral plane area. Normally, maximum rudder angle is limited to 35° port and starboard, although some ships have rudders designed for high maneuverability that can be deflected further. However, smaller tactical diameter is almost always associated with the hull assuming a larger drift angle. Because this produces extremely high drag as well as side forces, the speed loss in a sustained tight turn is very large. Further, at high speeds, extreme turns also can produce large angles of heel. For a displacement ship, the heel is toward the outride of the turn, and the angle in a high-speed turn may become large enough to produce a hazardous condition.

Directional stability. Depending on hull, skeg, and rudder design, a ship may be either directionally stable or unstable. A stable ship, if perturbed from a straight trajectory with the rudder remaining fixed in its neutral position, will tend to return to a straight trajectory. An unstable ship, by contrast, will tend to go into a turn, either to port or to starboard, and remain in a turn, even with the rudder neutral. Consequently, once a turn is established, merely returning the rudder to its neutral position may or may not result in recovery from the turn, even over a long period of time. Further, a directionally unstable ship can keep to a straight course only by moving the rudder slightly from one side to another, more or less continually, which produces an additional drag.

However, while an excessively unstable ship is difficult to control, some degree of directional instability is not necessarily unsafe or even undesirable. Most large tankers, for example, are directionally unstable. However, the inertia of large ships is so great, and their responses are consequently so slow, that tendencies to diverge from the intended course can be easily countered by the helmsman or an automatic controller or autopilot. By contrast, some excessively stable ships are also slower-turning.

Acceleration, deceleration, and backing. Compared with aircraft or land vehicles, even the most powerful

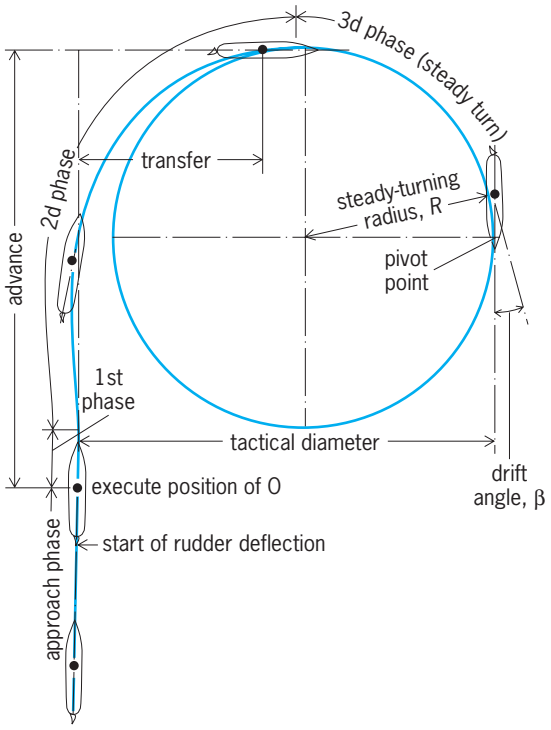


Fig. 2. Turning path of a ship. (*After E. V. Lewis, ed., Principles of Naval Architecture, Society of Naval Architects and Marine Engineers, 2d ed., 1988*)

ships have a very low power-to-weight ratio. Consequently, the accelerative performance of ships is generally a secondary design issue at best. However, acceleration, deceleration, and backing have important effects on controllability. The effectiveness of a rudder, whether in turning or course-keeping, is increased when the rudder is located directly behind a propeller, because the outwash from the propeller, when going ahead, increases the velocity of flow over the rudder, and consequently the available rudder side force. For this reason, when a propeller is producing additional ahead thrust for acceleration, the rudder is also able to act with more authority. Conversely, when a propeller is going astern, the flow over the rudder is reduced and the controllability of the ship may be compromised. Because of the geometry of a ship's stern, a propeller generally produces side force as well as an axial thrust. Consequently, when backing, it is not unusual for a single-screw ship to have markedly asymmetric turning behavior. Further, when attempting to decelerate hard (the so-called crash-back maneuver), single-screw ships tend to yaw in one direction, depending on the direction of propeller rotation, while the rudder may not be sufficiently effective to prevent this.

Lateral thrusters. Many ship-handling tasks require fine control of lateral position and speed. These include docking and undocking (large ships not fitted with lateral thrusters usually employ tug assistance); dynamic positioning (hovering over a position on the sea floor for oil exploration or drilling, for example), and precision track-keeping (for example, while cable laying or surveying).

Lateral thrusters of several different types are often fitted on merchant ships. The most common type is the tunnel thruster, a screw propeller mounted inside a transverse tunnel open at either side of the ship. Thrusters may be mounted at the bow only, or at both bow and stern, where the hull is relatively narrow. Tunnel thrusters are effective only at low speeds, typically below 3 knots (1.5 m/s). Other types, using steerable waterjets or steerable propellers, may remain effective at somewhat higher speeds.

Seakeeping

The modern term "seakeeping" is used to describe all aspects of a ship's performance in waves, affected primarily by its motions in six degrees of freedom (**Fig. 3**). Seakeeping issues are diverse, including the motions, accelerations, and structural loads caused by waves. Some are related to the comfort of passengers and crew, some to the operation of ship systems, and others to ship and personnel safety. Typical issues include the incidence of motion sickness, cargo shifting, loss of deck cargo, hull bending moments due to waves, slamming (water impact loads on sections of the hull), added powering in waves, and the frequency and severity of water on deck.

Even in the era of the wooden sailing ship, some elements of ship geometry were adapted to reduce deck wetness, and to avoid extremely high structural loads due to waves. High freeboard, sheer (the upward sweep of the decks toward bow and stern), elevated forecastles and sterncastles, well-rounded bows and sterns, and sections with higher deadrise (sloping rather than flat-bottom frames) were common features that were known to contribute to seaworthiness and seakindliness, a general and now largely obsolete term referring to deep but "easy" motions (in effect, low accelerations).

A more scientific approach to seakeeping improvement began with the consideration of the ship's roll, especially quick roll, as an impediment to accurate gunnery. It was known centuries ago that a ship with excessive initial stability also tended to roll quickly. More recently, the development of seakeeping theory and design features intended to improve seakeeping has been led by a desire for passenger comfort, and by naval requirements, in particular the operational need for helicopter operations from small or medium-sized surface warships.

Linear theory of ship motions. Beginning in the early 1970s, a simple but very useful theory of ship motions became widely used in ship design. Mathematically, the ship's oscillatory behavior in waves may be considered as that of a linear system consisting of springs, masses, and dampers, subjected to a harmonic excitation due to the forces and moments imposed by an incident wave. Six coupled linear equations of motion are involved, one for each degree of freedom, and the relevant characteristics of the ship may be expressed as coefficients for restoring force, inertia, and damping.

Computation of these coefficient matrices is required. The restoring force coefficients are computed from simple hydrostatics. The inertia matrix elements, however, must include not only the actual mass and moments of inertia of the ship itself, but also the component of the hydrodynamic force which acts in phase with the acceleration of the ship's harmonic motion. This hydrodynamic component is known as added inertia or added mass. Once the required coefficients are computed, the coupled equations may be solved. The resulting solution, at a given wave frequency, speed, and heading, gives the magnitudes and phases of the six ship motions in response to an assumed sinusoidal wave of unit amplitude. Repeating the computation for a range of

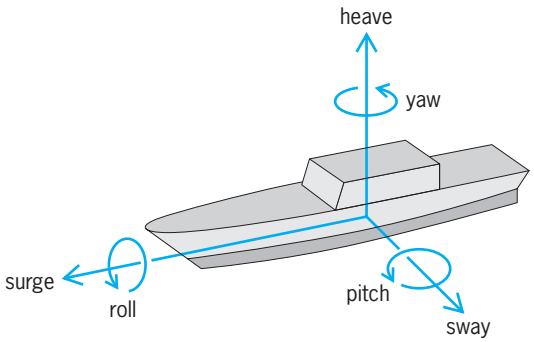


Fig. 3. Ship motion degrees of freedom.

wave frequencies yields six functions of frequency, each known as a response amplitude operator.

Actually, ocean waves are not regular sinusoids, but are random and irregular, characterized by sea spectra analogous to the spectra of other oscillatory natural phenomena such as sound or light. However, linear theory permits real sea conditions to be modeled as a superposition of sinusoidal components, which are linearly additive. Consequently, if a sea spectrum is specified, and the ship's response amplitude operators have been calculated, then spectra of all six motions of the ship can be computed as well. From the motion spectra, statistical measures of the ship's motions and accelerations, including predicted probabilities and extreme values of the motions (the largest values that can be expected to occur within a given time), can be derived.

In spite of the strong simplifying assumptions required to formulate a linear theory of ship motions, results for the vertical plane motions, heave and pitch, are surprisingly accurate, provided that certain limits are not exceeded. However, for roll and lateral motions coupled with roll (sway and yaw), the results of a purely linear treatment are less accurate. Largely this is because roll-restoring moments are often quite nonlinear even at moderate roll angles, while roll damping also includes both linear and viscous (strongly nonlinear) components. As a result, linear theory often tends to underestimate roll in mild conditions but overestimate it in severe conditions.

Added resistance in waves. When a ship operates among waves, even in relatively moderate sea conditions, its resistance and required power generally increase. Added resistance in waves arises from several sources. Waves generated by the ship's motions radiate away from the ship, and the radiated energy in these waves corresponds partly to work which must be done against an added drag force. Also, the incident wave is partially reflected by the presence of the ship, even if its motions are small, and this entails a transfer of momentum or force.

Typically, ship motions contribute most significantly to the added resistance of fast, fine-hulled ships, while reflection forces contribute more in the case of large, massive ships, such as tankers. Apart from direct wave-induced forces, a ship's wetted surface and underwater form are constantly changing due to waves and ship motions, and the result is often increased viscous drag. Finally, the cyclic use of the rudder to remain on course adds yet another drag.

The effects of waves and ship motions on propeller performance are less straightforward. The average thrust required tends to increase due to added resistance, and in general efficiency is reduced. In extreme seas, of course, where the propeller may move vertically a great deal, additional cavitation, air drawing, and even emergence of the propeller can occur, with dramatic losses of thrust. Consequently, even if a ship's speed is not reduced voluntarily in rough seas, a loss of speed often occurs due to power limitations.

Motion reduction. Because roll is relatively lightly damped and easily excited, it is often troublesome as a cause of motion sickness even in relatively benign sea conditions. The incidence of seasickness is particularly high for passengers, who are often less accustomed to the ship's motion than crew members. Most large ships are fitted with bilge keels, appendages which increase roll damping and added inertia. On passenger and cruise ships, among other types, roll stabilization systems are often fitted in addition to bilge keels. Dramatic roll reduction, occasionally approaching 90%, can be obtained with active antiroll fins that are controlled by signals from a gyroscope sensing unit. However, fins are only fully effective at relatively high speeds. Other antiroll systems include active and passive tanks, in which a fluid is either actively pumped or permitted to flow passively, and tuned in such a way as to produce rolling moments out of phase with the wave-induced roll, even if the ship speed is low.

Progress has been made in the design of hull forms with increased damping of pitch and heave motions, including modification of section shapes, and even the experimental use of antipitch fins. In general, these efforts have only been moderately successful, and forms with improved seakeeping performance are often compromised from the standpoint of powering. Another approach to improved seakeeping is reduced wave excitation. Excitation forces are largely generated by changes in the ship's buoyancy during a wave encounter. Consequently, a ship with a small waterplane area is inherently less affected by the motion of the sea surface. Developments such as the small-waterplane-area twin-hull (SWATH) concept have been successful in improving motions in adverse sea conditions (**Fig. 4**).

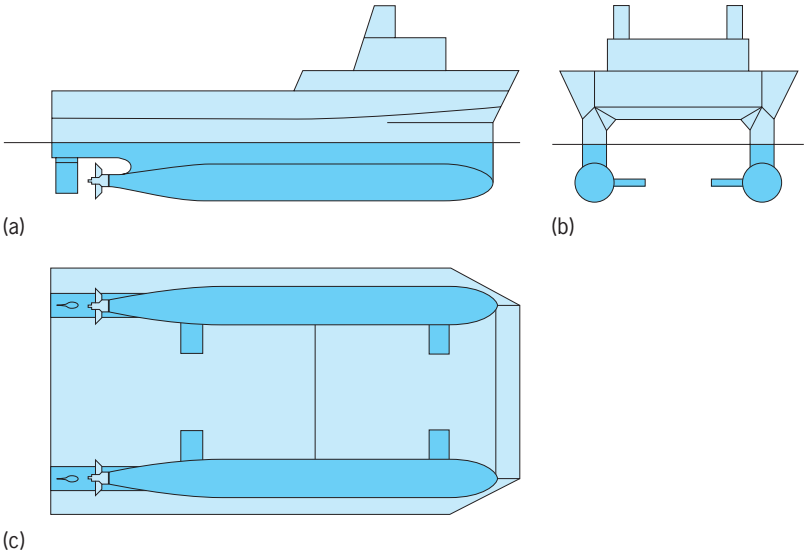


Fig. 4. Small-waterplane-area twin-hull (SWATH) vessel for reduced motions in waves. (*a*) Side view. (*b*) Front view. (*c*) Bottom view.

Role of Computational Fluid Dynamics

In the past, initial powering, maneuvering, and seakeeping predictions for new designs depended

almost entirely on design rules of thumb or, at best, applicable series or regression data from previous model tests, subsequently refined by additional model tests. With the increase in computing power available to the naval architect, computational fluid dynamic methods are now applied to some of these problems in various stages of the ship design process.

Numerical methods for computing the pressures and velocities in the flow around a ship's hull facilitate estimates of resistance and maneuvering forces. These methods include potential flow codes, which give approximate solutions in which viscous forces are neglected, and more complete viscous flow solutions, making certain modeling assumptions regarding the boundary layer. Computational fluid dynamic methods permit estimation of the resistance and of the geometry of the ship's wave system and wake (including predictions of propeller-hull interactions and propeller-induced vibrations), and are also useful in improving propulsor designs. These methods have been applied with some success in simulating surface ship and submarine maneuvers. Computational fluid dynamics (CFD) offers the potential for solving nonlinear seakeeping problems, especially important in extreme conditions where linear theory proves inadequate. While extremely useful in some areas, computational fluid dynamic methods have not been uniformly well developed for practical ship design, and many promising approaches remain in the area of underlying research. Robert M. Scher

Bibliography. R. L. Harrington (ed.), *Marine Engineering*, 1992; E. V. Lewis (ed.), *Principles of Naval Architecture*, 2d ed., 3 vols., 1998; H. E. Saunders, *Hydrodynamics in Ship Design*, 3 vols., 1956; R. Taggart (ed.), *Ship Design and Construction*, 1980.

Ship routing

The selection of the most favorable route for a ship's transit in the open ocean. Safety, economics, and environmental protection play a key role in determining this selection. In the late 1890s, the practice of following predetermined routes for shipping was adopted by shipping companies operating in the North Atlantic. As ships grew both in size and number, the ability to plan open-ocean transits based upon the environmental conditions expected during a voyage and the effect of these conditions upon the ship's performance became important factors in planning a ship's route. Ship routing is also referred to as weather routing, optimum track ship routing, and meteorological navigation.

Figure 1 illustrates the concept of ship routing. A ship is scheduled to depart the English Channel bound for Norfolk, Virginia, and the owners desire a least-time track. Routes can be computed to minimize transit time, avoid waves over a specific height, or, more typically, provide a compromise between travel time and the roughness of the seaway to be encountered. The shortest track is the modified great circle, but in this case the severe storm off Ireland would generate strong opposing winds and heavy head seas. In addition, the ship would obtain little benefit from the favorable, following winds that could be realized from the second storm as it proceeds eastward. The standard track avoids the head seas to some degree, but is a longer route and approaches the second storm so closely that the benefits of following winds are negated by the heavy seas and associated violent ship motions. An optimum route for reducing travel time avoids the head seas, yet takes advantage of the favorable conditions by paralleling the second storm just the right distance to the north. This example is oversimplified, since storms move erratically at times, change intensity, or become stationary. To compensate for this variability, routing offices compute the initial route based on the long-range weather forecast, but maintain a daily plot of both the ship and the storm positions. If conditions warrant, an amended route is sent to the ship master based on the latest weather changes. *See* OCEAN WAVES; STORM DETECTION; WEATHER FORECASTING AND PREDICTION.

Effect of environmental conditions. In addition to all the latest environmental predictions, the ship router must understand how these conditions will affect the specific ship under routing. For this, ship performance graphs are used, similar to **Fig. 2**, which relate the average speed that a ship of a particular class will make under increasing heights of head, beam, and following waves. For some ships, a relationship has been developed between ship speed and both wave height and length. This knowledge is useful, because for some wave lengths the ship's pitching motion becomes accentuated through resonance. The loss of ship's speed is due more to the master's reducing engine speed to ease the ship motions than to direct resistance of the seaway to the ship's passage. *See* SHIP DESIGN; SHIP POWERING, MANEUVERING, AND SEAKEEPING.

Although the effect of waves is the predominant factor considered, the ship router also utilizes knowledge of ocean currents, wind, and hazards to navigation such as ice fields or fog banks. On U.S. east coast runs, the currents are more important than the waves. Most major oil companies have their tankers follow the Gulf Stream on northbound runs and avoid it on the return trip. This practice is logical, since a ship can gain a 3–5-knot (1.5–2.5 m/s) push by riding the current. The National Oceanic and Atmospheric Administration (NOAA) issues the weekly *Gulf Stream Bulletin*, which provides ships with the latest position of the Gulf Stream as measured by meteorological satellites carrying infrared sensors that detect warm water. *See* GULF STREAM; METEOROLOGICAL SATELLITES.

Extent of application. Ship routing is available from a number of environmental consulting companies, and is utilized by many of the large shipping companies. The U.S. Navy operates its own routing service, which provides some 1200 transits per year to everything from seagoing tugs to aircraft carriers. Routing services are available from at least five countries other than the United States: the United

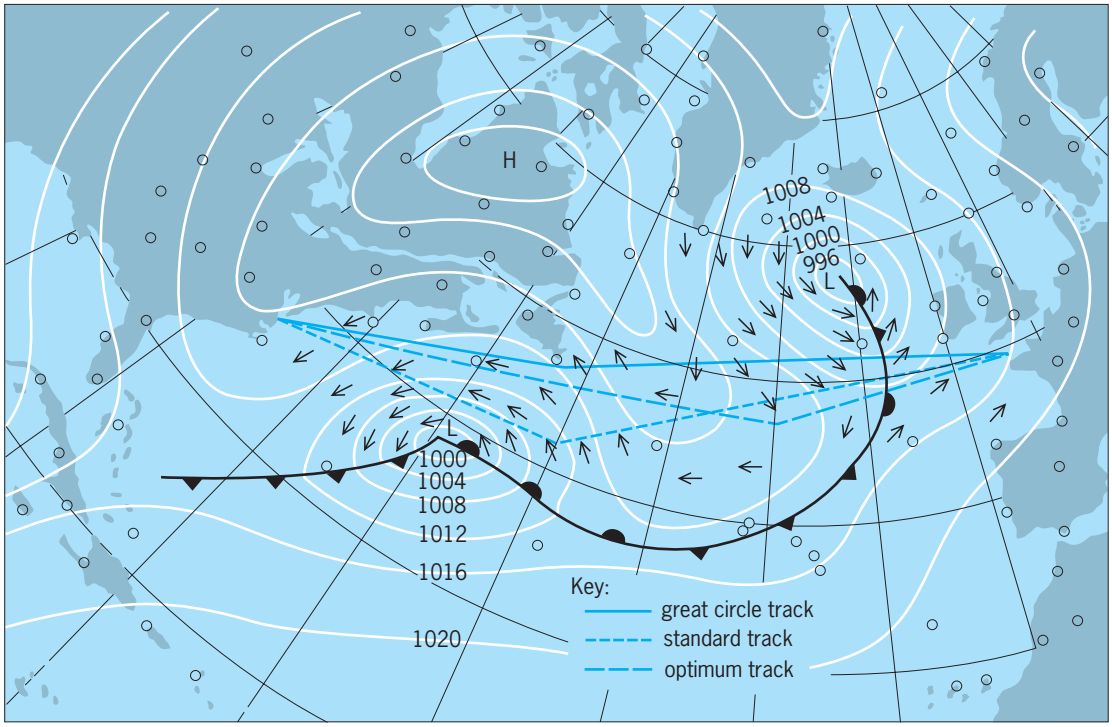


Fig. 1. Meteorological chart illustrating concept of ship routing. (*After R. W. James, The present status of ship routing, in Proceedings of Interocean '70, Düsseldorf, November 1970*)

Kingdom, Germany, Netherlands, Norway, and Russia. Additional information on weather routing services is available from the World Meteorological Organization.

Ship routing also benefits from improvements in all weather navigation capabilities such as those provided by the Global Positioning System (GPS). Additionally, new standards being implemented by the International Maritime Organization (IMO) provide for the automated exchange of ship information through a Universal Shipborne Automatic Identification System (AIS). This system employs automated transponders designed to operate in a ship-to-ship mode for collision avoidance and a ship-to-shore mode when a ship's route begins to approach coastal areas. Providing coastal authorities with information about a ship's speed, course, position, and cargo assists in the efficient management of ship traffic in congested shipping lanes and ports. *See* SATELLITE NAVIGATION SYSTEMS; VESSEL TRAFFIC SERVICE.

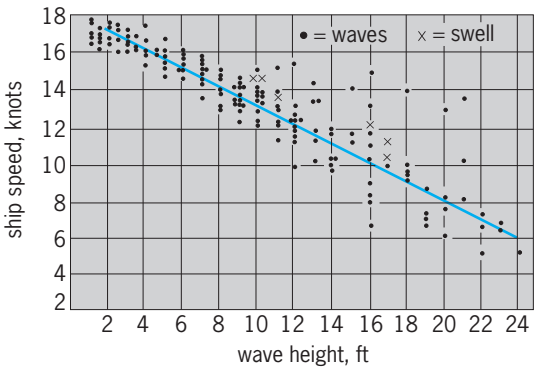


Fig. 2. Ship performance graph showing relationship between speed of ship and head waves. (1 ft = 0.3 m; 1 knot = 0.5 m/s.)

Benefits. Ship routing is very cost-effective. The charges for a route are low compared to the savings realized by reduced storm damage, less fuel consumption, and shorter travel times. For a large company, the reduction in transit time means that more cargo is carried in fewer ships than would be possible if weather routing were not used. Additional intangible benefits accrue to routed ships in terms of crew and passenger well-being, since the ship has a smoother crossing than nonrouted ships. When the savings of ship routing are coupled with improved navigation capabilities and automated ship reporting systems for coastal and port management, further cost saving are realized.

Ship routing has progressed from the simple concept of the sailing masters to "follow the winds" to modern practice which utilizes such technology as satellite data to track storms and measure currents, high-speed computer calculations, and modern communications to follow the ship's progress and provide route diversions where necessary for safety, economics, and protection of the environment.

Richard W. James; Douglas Taggart

Bibliography. N. Bowditch, National Imagery and Mapping Agency, *The American Practical Navigator, 2002 Bicentennial Edition*, Celestaire, Inc./Paradise Cay Publications, 2002; T. J. Cutler, *Dutton's Nautical Navigation*, 15th ed., Naval Institute Press, 2003.

Ship salvage

Voluntary response to a maritime peril by other than the ship's own crew. Salvage is usually time-critical. The property in danger can be any type or size of vessel or maritime cargo. Salvage is encouraged by compensating the salvor based on risk to the salvor and the salvor's equipment, conditions under which the salvage service was performed, the value of the vessel and cargo saved and, more recently, minimizing environmental damage.

Aside from international agreements requiring the saving of life at sea, salvage is voluntary, and is driven by economics. Given sufficient time and money, salvage masters and salvage engineers can accomplish virtually any salvage operation. However, the payoff may not be worth the risk and expense, and the ship and cargo owners or the underwriters may not want to spend more than the value of the ship and cargo in order to save them.

Contrary to popular belief, the salvor does not own or gain title to recovered property, but establishes a maritime lien. The salvor's award may be established by agreement between the salvor and the owner, master, or insurance underwriter, or it may be adjudicated either by the courts under admiralty law or by an arbitrator agreed to by the owner, master, or underwriter and the salvor. Changes in maritime law and fee structures, and public response to major pollution casualties, have focused the salvor's attention on preventing or minimizing oil discharge when attending a casualty.

The salvor must often work with limited resources, usually in a remote location in bad weather. Salvors are a combination of maritime ambulance, engineer, and adventurer. Critical decisions must be made and implemented quickly.

Ships, equipment, and personnel. Salvage services can be performed using either dedicated salvage vessels or vessels hired for the specific operation. Salvage ships are typically large seagoing tugs, usually capable of 5000–10,000 hp (3750–7500 kW), designed and outfitted to work at remote sites in all weather (**Fig. 1**). The largest tugs can be as powerful as 40,000 hp (30,000 kW) and exert a pull against a fixed object, called bollard pull, of up to 250 tons (227 metric tons) with their main tow wire. Wire used in salvage is made from multiple twisted steel wires and can exceed 2.5 in. (6.4 cm) in diameter. Salvage ships carry a variety of portable salvage equipment such as firefighting gear, pumps and patching material, electrical generators, compressors, diving equipment, wire and chain, beach gear, pollution control equipment, and the material to make whatever might be needed on-site.

Maritime countries sometimes maintain ships on "salvage station" near high-density shipping lanes or areas where problems historically occur. With improvements in navigation, communications, and engineering reliability, there has been a reduction in the number of casualties, although the potential for pollution and damage and award are generally larger per casualty. The net result has been a reduction in the number of dedicated salvage ships.

Some salvage companies maintain portable equipment at a central location and mobilize it quickly by truck, rail, or air to vessels of opportunity, usually a tug or barge, near the salvage site. Other salvors

Fig. 1. USS *Safeguard* (ARS50). (*U.S. Navy*)

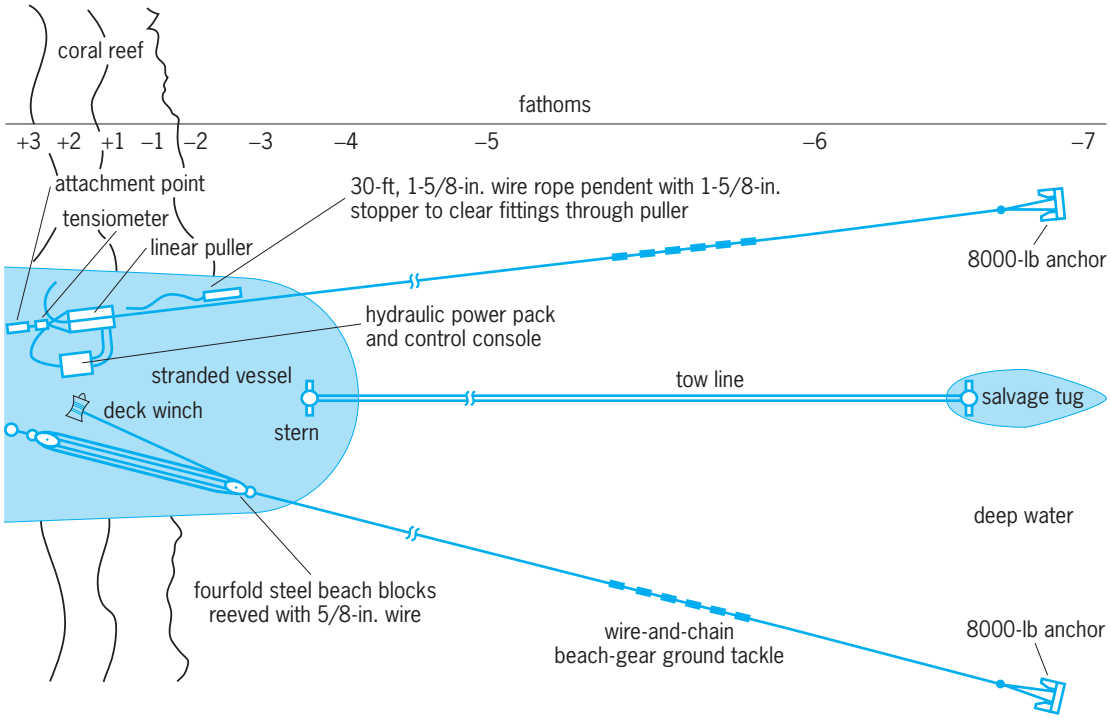


Fig. 2. Beach-gear deployment. Beach gear enables a stranded ship to assist in pulling itself into deep water. 1 fathom = 1.8 m; 8000 lb = 3600 kg; 30 ft = 9 m; 1–5/8 in. = 41 mm; 5/8 in. = 16 mm.

have established loosely allied groups of specialists and equipment providers that they call upon as required.

Trained salvors frequently have multiple skills, and can be riggers, welders, divers, crane operators, carpenters, and pump engineers. They are supervised by an experienced salvage master. A salvage engineer is often part of the team and can determine the hull strength and stability of a casualty. Together, the salvage engineer and the salvage master determine the best way to refloat a vessel, and develop a detailed salvage plan to accomplish the work.

Salvage situations. True salvage is time-critical, involving assets that are immediately available. Assistance is provided in situations involving collisions, firefighting, flooding and damage control, damage from hostile action, and breakdowns of propulsion or steering systems. The salvage team may fight fire from off-ship or board the casualty to fight fires. Salvors also may board the ship to control flooding and stabilize the casualty if required. Rescue towing often results from one or more of the above and the need to reach a safe haven or avoid going aground.

Stranding. This results when a ship drives, or is driven, aground through engine or steering casualty, error, problem with navigation or navigation aid, or bad weather, and cannot extricate itself. Time-criticality may be due to coming bad weather, tidal range, ongoing damage to the ship from the grounding, or other requirement such as clearing a blocked shipping channel. The salvor must assess the condition of the casualty and what is required to refloat it, the potential for further damage to the ship, and the potential for loss of cargo and damage to the environment. A wide range of other factors such as weather, tides and currents, the weight of the ship now supported by the ground (ground reaction), and the type of bottom (coral, mud, rock, sand, and so forth) also must be considered in developing a salvage plan. The situation will usually deteriorate quickly without immediate intervention by the salvor.

A lightly grounded vessel can frequently be freed by tugs pulling at high tide. This is often combined with moving fuel or cargo, especially liquids, within the ship to lighten the grounded portion of the hull, or jettisoning or removing fuel or cargo.

The tops of damaged, flooded compartments such as tanks or engine rooms can be temporarily sealed. Compressed air is then pumped into the compartments. The air displaces the liquid tank contents, usually a combination of flooding water and tank contents such as oil or water, and provides a bubble of air in the tank to restore some buoyancy to the ship.

Beach gear adds great pulling force with relatively little equipment (**Fig. 2**). Beach gear consists of anchors, wires and chain, and a pulling power source such as a winch or hydraulic puller. Anchors are placed to seaward and connected by chain and wire to the stranded vessel. The wire is led either to a hydraulic puller, which has clamps that "walk" a wire in using hydraulic rams, or to a multiple purchase tackle system leading to one of the ship's winches. Each leg of beach gear can exert a force of about 50–60 tons (45–54 metric tons) to help pull the ship off.

Environmental salvage. This is the protection of the environment from damage by pollutants that may result

from a maritime casualty. International agreements such as the International Convention for the Prevention of Pollution from Ships (MARPOL 73/78) and national laws such as the Oil Pollution Act of 1990 highlight the need for protecting the environment. This is most effectively accomplished by keeping pollutants in a ship during a casualty, minimizing dispersion, and rapidly recovering any spilled pollutants, usually cargo crude oil or petroleum products, or ship's fuel stored in bunkers. Salvage actions such as pumping air into a damaged tank to restore buoyancy may blow oil out through the damaged hull below the waterline. Salvors must be prepared to deal with pollutants when this occurs, and plan salvage operations to minimize oil release. Regulations and standard salvage agreements provide financial incentives for pollution reduction. Many salvage companies are capable of controlling pollution.

Wreck removal. This involves recovery of a stranded or sunken vessel, and is usually not time-critical. The Suez Canal clearance in 1974 is an example of a large wreck removal operation.

Many salvage techniques can be brought to bear. If the main deck of the stranded or sunken vessel is above water, a combination of patching and pumping may be sufficient to refloat the casualty. If the main deck is not submerged too deeply, a watertight wall, known as a cofferdam, may be built around the main deck to above the water surface. Patching and pumping may then refloat the casualty.

Floating cranes, some capable of lifting up to 14,000 tons (12,700 metric tons), can be used to lift vessels using wires attached to the hull by divers. The casualty is raised to the surface and pumped out, lifted onto a barge, or moved to a shallower location for patching and pumping. In 1982, a floating crane and barge-cradle combination was used to raise Henry VIII's flagship, the *Mary Rose*, sunk in 1545, and place it in dock in Portsmouth, England, for conservation.

Parbuckling is one of the techniques used to right a vessel that is on its side or inverted (**Fig. 3**). Wires are run from the edge of the main deck on the uppermost side of the hull, around the hull, across the keel, under the hull (passing the main deck on the opposite side), and then to a source of lift, usually a crane, on the surface. Applying a lifting force on the wires causes the vessel to be rolled to an upright position. Other techniques can then be applied to float the vessel.

Pontoons, attached to the sunken vessel by wires and then inflated, can provide lifting force. They are difficult to manage, and are generally used for lighter lifts or to supplement other forms of lift or buoyancy.

Heavy-lift craft can raise sunken ships in stages (**Fig. 4**). Wires are run under the casualty between the flooded-down heavy-lift craft. The heavy-lift craft are pumped out on a rising tide, and the casualty is moved to shallower water, where it rests on the bottom as the tide falls, and the heavy-lift craft are again flooded. This process is repeated until the casualty is in water shallow enough for patching and pumping efforts to begin.

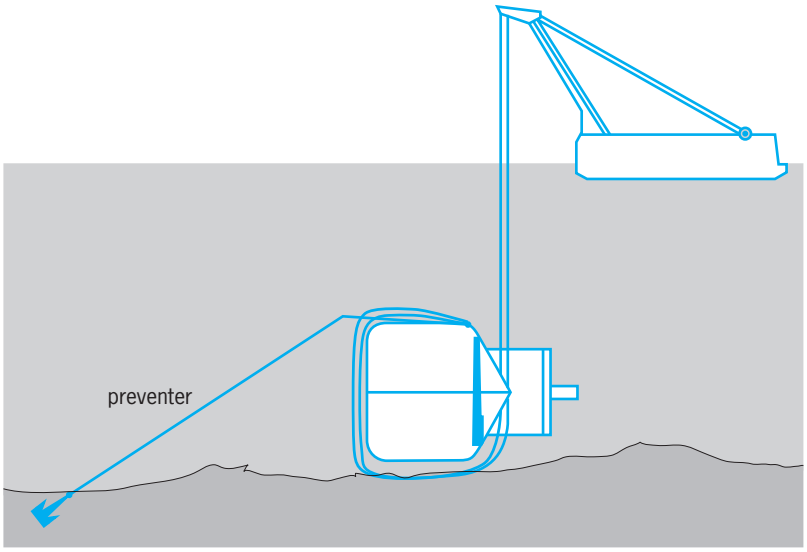


Fig. 3. Parbuckling technique, used to right a ship that is on its side or capsized. Wires are passed from the uppermost deck edge, around under the hull, and to a source of lift. Tensioning the wires rolls the ship upright, and it can then be raised by customary means. A preventer may be used to keep the ship from dragging.

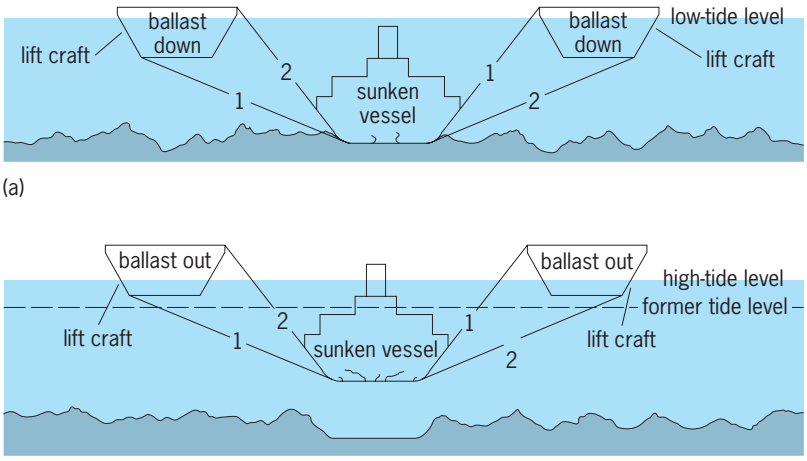


Fig. 4. Lift craft operation. (*a*) At low tide, the lift craft are ballasted down, and cables are passed beneath the sunken vessel and pulled tight. (*b*) At high tide, the lift craft are deballasted, and the sunken vessel is raised above the ocean bottom and towed away.

Wrecks can also be removed by breaking them into pieces small enough to be lifted using cranes on site. This can be accomplished with cutting torches, explosives, or something as simple as a large steel plate, sharpened on one edge, repeatedly raised by a crane, and dropped onto the same spot, cutting the wreck into liftable pieces.

Cargo. A ship's cargo may be of more value than the ship itself, and the salvor may recover the cargo without recovering the ship. The best-known type of cargo salvage is the recovery of jewels or precious metals. But even something as mundane as copper ingots may be considered worth recovering by both the salvor and the owner or underwriter.

Salvage contracts. The concept of "no cure/no pay" is unique to marine salvage contracts. With the exception of recovering expenses for environmental

protection efforts, the salvor who does not save at least some property for the owner is not entitled to compensation. Most time-critical salvage agreements incorporate these terms.

Perhaps the best-known agreement is the Lloyds Open Form (LOF) and its variants, originated in 1892 at the early center of maritime insurance, Lloyds of London, and updated at infrequent intervals. The LOF is an agreement to have an arbitrator at Lloyds evaluate the salvage service in light of the degree of danger, value of the property saved, salvor's skill and effort in preventing environmental damages, and salvor's expenses and long-term investment in salvage equipment. The amount of a salvage award cannot exceed the value of property saved.

A salvor who finds a ship abandoned at sea and tows it to safe haven without a contract has performed voluntary salvage. The salvor does not own the derelict, but has a claim against the ship for an award. If the claim cannot be resolved with the ship owner, the salvor may go to court to obtain an award.

Technology changes. Advances in technology have had a significant impact on ship salvage. Improvements in navigation and piloting equipment, training, engineering reliability, and ship safety have reduced the number of ship casualties requiring salvage assistance. On the other hand, the increased size of ships, particularly oil tankers, has increased the potential impact of each casualty. Ship owners are now required to have extensive casualty plans.

Larger salvage tugs and larger floating cranes have greatly increased the pulling and lifting forces that can be brought to bear on a casualty, and enabled salvage operations to take place in worse weather than they could before.

Salvage engineering advances have enabled salvors to more precisely calculate required lifts, pulls, and pumping. Computer programs enable modeling of ships and rapid calculation of casualty strength and stability, including the impact of damage on hull strength and of flooding on ship stability.

Deep-ocean search, observation, and recovery techniques, including towed sonars and cameras, remotely operated vehicles (ROVs), small crewed submersibles, and saturation diving have brought many otherwise unrecoverable and potentially abandoned wrecks within reach. In 1992, the U.S. Navy located and recovered a crashed helicopter in over 17,000 ft (5180 m) of water. Inaccessible for years, the submerged Titanic is now visited routinely. This access to deeper wrecks has resulted in much legal activity in admiralty law concerning ownership, abandonment, and salvage claims, particularly for high-value cargoes or vessels having significant archeological or historic value. Search, observation, and recovery techniques using towed sonars, ROVs, divers, and salvage ships and equipment are also now used routinely in aircraft salvage such as the crash of TWA Flight 800 off Long Island, New York, in 1996. *See* SONAR; UNDERWATER VEHICLE; UNDERWATER SOUND; UNDERWATER TELEVISION.

Richard P. Fiske

Bibliography. E. Ellsberg, *Under the Red Sea Sun*, 1946, paperback 1974; G. Forsberg, *Salvage from the Sea*, 1977; W. I. Milwee, Jr., *Modern Marine Salvage*, Cornell Maritime Press, 1996; F. Mowat, *The Grey Seas Under*, 1958 reprint, Lyons Press, 2001; National Research Council, *A Reassessment of the Marine Salvage Posture of the United States*, 1994; G. H. Reid, *Marine Salvage: A Guide for Boaters and Divers*, Sheridan House, 1996.

Shipbuilding

The construction of large vessels which travel over seas, lakes, or rivers. A ship must be capable of self-sufficiency in a hostile marine environment for extended periods of time while simultaneously providing the necessities and comforts expected by passengers, the safety and timely arrival of cargo, the offensive and defensive capabilities required by warships, and a variety of other particular functions. The bulk-cargo carrier, such as an oil tanker in which the cargo is voluminous and homogeneous with no premium on speed of transport, is the most efficient mode of freight transportation.

Many different approaches have been used in the construction of ships. Sometimes a ship must be custom-built to suit the particular requirements of a low-volume trade route with unique cargo characteristics. Such a ship may be one of a kind, and as a consequence there may be little opportunity to employ labor-saving techniques during construction due to the low incidence of repetitive processes. A cable-laying ship, used to lay communication lines across bodies of water, could be of this type. On the other hand, there are many instances where a significant number of similar ships are constructed, providing an opportunity to employ procedures which take advantage of repetitive processes. Examples of this construction procedure, known as series production, include the production of U-boats in Germany during World War II, oil tankers in Japan during the 1960s, and destroyers by the Ingalls Shipyard in Mississippi during the 1970s.

Shipyards can be laid out so as to be ideally suited for series production; however, such a layout may be entirely unsuitable for the construction of dissimilar ships if market conditions change. Obviously, the layout used to build U-boats would not be suitable for the construction of ocean liners. Consequently, shipyard facilities are designed to take advantage of labor-saving techniques which can be used in the construction of ships which have an immediate market, while, at the same time, avoiding inflexible facilities that cannot be easily modified to construct ships of different types.

Shipyard layout and facilities. A shipyard site must be on deep water accessible to the ocean (or, in the United States, to the Great Lakes) and should have railroad connections for the delivery of materials. **Figure 1** shows the layout of a large American shipyard. In such shipyards, where a variety of ships are built concurrently, sufficient space must be provided for the storage of vast amounts of material and equipment due to the impracticability of scheduling such items to be delivered to the shipyard

at precisely the time required. In addition to the space required for the storage of purchased steel plates, rolled steel sections, pipe, and other materials, space is required for the proper storage of steam boilers, turbines, reduction gears, pumps of all types, and the broad variety of specialized mechanical, electrical, and electronic equipment that is usually purchased by the shipbuilder. Some shipbuilders manufacture their own specialized equipment; others have an extensive manufacturing capability, but elect to purchase such equipment from other suppliers. The shipbuilder's decision is determined by shipyard capability, worker-power availability, and economics. Since equipment manufacturers serve many shipyards and limit their production to specific types of equipment, they are able to take advantage of labor-saving techniques in their area of specialization, and this often gives them an economic advantage. However, such is not always the case. The production of slow-speed diesel engines for ship propulsion does not lend itself to a small-scale operation. These diesel engines may be more than 42 ft (13 m) high and weigh 1375 short tons (1250 metric tons), with many of the engine parts weighing several metric tons. Consequently, the large, heavy, and expensive facilities required to manufacture and assemble these diesels are similar to those required for other shipbuilding processes. As a result, some shipbuilders manufacture their own diesel engines.

The handling and processing of steel through the production processes require the greatest amount of facilities and space in a shipyard. Not only must the steel be received, inspected, sorted, and stored, but it must also be blasted, primed, cut to shape, perhaps formed to the proper configuration, welded to make assemblies, and then stored again until needed to make larger erection units.

The building of a ship can be divided into seven phases: design, construction planning, work prior to keel laying, ship erection, launching, final outfitting, and sea trials. Ship design and sea trials are discussed in other articles; therefore, this article covers the remaining five phases and is generally applicable to the building of both merchant and naval ships. *See* SHIP DESIGN; SHIP POWERING, MANEUVERING, AND SEAKEEPING.

Nomenclature. The process of constructing the steel hull of a ship requires an understanding of terms which are not used in other industries. The type of structure used varies considerably for different types of ships, but the descriptive terms do not. For illustration purposes, consider the structure of a simple cargo vessel shown in **Fig. 2**. The keel consists of a flat keel and a vertical keel. The flat keel is the center strake (row of plates) of the bottom shell plating; and the vertical keel is a vertical plate on the bottom centerline, running the entire length of the ship. The bottom shell and side shell form the outer, watertight envelope of the ship.

Floors are vertical plates extending across the bottom of the ship, perpendicular to the centerline. Longitudinal girders are similar vertical plates, generally running parallel to the centerline. An inner bottom is several feet above the bottom shell, and provides protection in case of minor bottom damage and a smooth bottom surface for the holds and machinery spaces. Decks are horizontal surfaces corresponding to the

Fig. 1. Newport News shipyard. (*Newport News Shipbuilding*)

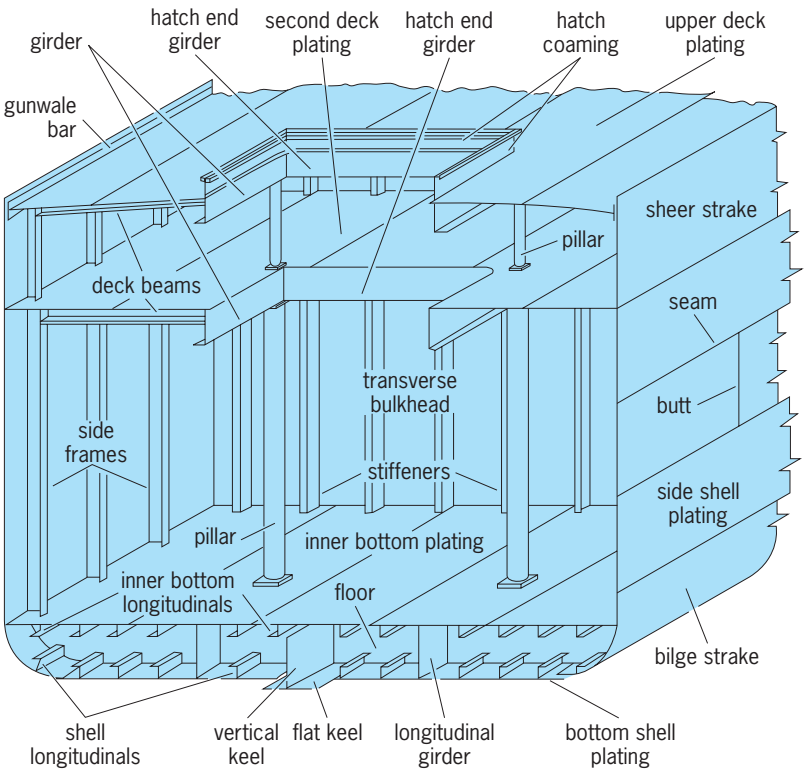


Fig. 2. Section through the structure of a simple cargo vessel.

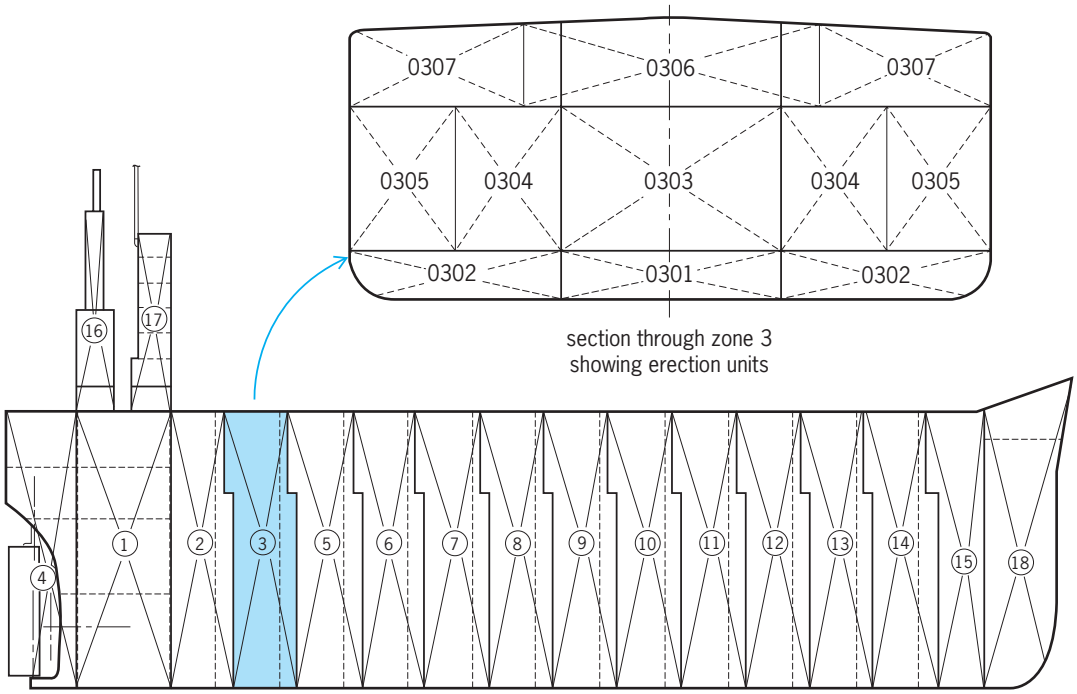


Fig. 3. Erection diagram for a tanker.

floors in a building; and bulkheads are vertical partitions, either longitudinal (fore and aft) or transverse (perpendicular to the centerline of the ship), that divide the ship into compartments. Frames, either longitudinal or transverse, are the structural members which hold the shell against the pressure of the sea. Deck beams support the deck and its loads, and in turn are supported by girders which rest on stanchions or pillars. Seams are the weld connections between the longer edges (usually fore and aft) of plates, and butts are weld connections between the shorter edges of plates. Seams connect strakes, and butts connect the plates within a strake.

Construction planning. The construction planning process is one of the most important aspects of shipbuilding because it establishes the construction techniques to be used and the schedules which all of the shipbuilding activities must follow. Construction planners generally start with an erection diagram (**Fig. 3**) on which the ship is shown broken down into erection zones and units. The first zone to be erected includes the machinery space which is beneath the deck houses; following that, zone 2 (which is just forward of the machinery space and includes the cargo pump room) is erected; and then zone 3 (the first of a series of cargo tanks) is placed in position. The typical midship section shown by Fig. 3 indicates a breakdown of the units which compose erection zone 3. To facilitate the fabrication of steel, insofar as possible, the erection units are designed to be identical. The size (or weight) of the erection units selected is usually limited by the amount of crane capacity available. Crane capacity at the various shipyards varies widely from less than 110 short tons (100 metric tons) to more than 1100 short tons (1000 metric tons), and it is the construction planner's job to plan the work to make optimum use of the facilities available.

Larger crane capacities do not necessarily equate to greater shipbuilding productivity, inasmuch as some shipyards with low crane capacities can be very competitive; however, the general trend of shipyards is toward higher crane capacities which provide the ability to assemble (erect) a ship by using large erection units which have been preoutfitted to the greatest practical extent. The practice of erecting a ship by using large erection lifts provides the advantage of minimizing the time that a dock is occupied by a ship being built, and permits large erection units to be constructed away from the dock under more favorable conditions. For example, with proper planning and suitable crane service, the steel structure can be turned during fabrication so that most of the welding is done in the more efficient downhand position. Also, the erection units can be transported to paint sheds where they can be painted or coated in the controlled environment required in some climates. In addition, the use of staging (temporary structure, such as ladders, for access) during ship construction is minimized. However, the use of large erection lifts has the disadvantage of requiring facilities for transporting and lifting erection units, as well as requiring facilities and space for the fabrication, processing, and storage of erection units. Of course, many ships have been erected essentially one plate at a time, and there is no reason why this practice should not continue.

Once the construction planners have established the manner in which the ship is to be erected and the sequence of construction, the schedules for construction can be developed. Working backward from the time an erection unit is required in the dock, with

allowances made for the many processes involved, a schedule of working plans is prepared. This is the schedule that the ship designers must follow in order to properly support the construction trades. Similarly, a schedule for the procurement of all purchased equipment must be developed. This requires that the construction planners decide at what point in the construction process the purchased equipment is going to be installed; that is, the equipment may be preoutfitted on an erection unit (thereby requiring an early delivery), it may be landed aboard ship just before the deck is installed over it (thereby permitting a later delivery), or it may be practicable to install the equipment after launch (thereby permitting the latest possible delivery).

Work prior to keel laying. Before the keel of a ship is laid (or when the first erection unit is placed in position) a great deal of work must have been accomplished for work to proceed efficiently. The working drawings prepared by ship designers completely define a ship, but often not in a manner that can be used by the construction trades people. Structural drawings prescribe the geometry of the steel plates used in construction, but they cannot be used, in the form prepared, to cut steel plates. Instead, the detailed structural drawings must be translated into cutting sketches, or numerical-control cutting tapes, which are used to fabricate steel. The trend is toward the use of computer-driven cutting torches which require the use of numerical-control cutting tapes. Several organizations have developed sophisticated computer programs which readily translate detailed structural drawings into machine-sensible tapes which can be used to drive cutting torches (**Fig. 4**). In addition to cutting the steel parts, many of the plates must be rolled into the desired configuration (**Fig. 5**). The parts must also be fabricated to form the erection units, and the erection units must be coated and preoutfitted. The preoutfitting (installation of piping, machinery, or other nonstructural items) of erection units may be required by the construction schedule and, if so, must be accomplished if work is to proceed efficiently.

Fig. 4. Steel plate being cut to the desired configuration by a computer-driven cutting torch. (*Newport News Shipbuilding*)

Fig. 5. Steel plate being formed in a roll press. (*Newport News Shipbuilding*)

Ship erection. If all of the preceding work has been accomplished properly and on schedule, the erection of a ship can proceed rapidly; however, short-term problem areas invariably arise. When erecting a ship one plate at a time, there are no serious fitting problems; but when 990-short-ton (900-metric-ton) erection units do not fit (or align) properly, there are serious problems which tend to offset some of the advantages previously mentioned for this practice.

Figure 6 shows a crude-oil tanker in the process of being erected. In Fig. 6*a* the centerline erection unit, which has been preoutfitted to contain the cargo-oil piping before being placed in the dock, can be seen to be substantially complete. Figure 6*b* shows a number of erection units being preoutfitted alongside the dock and aft of the ship. By referring to Fig. 3, the manner in which the ship is erected may become apparent. Figure 6*c* illustrates an advantage in the use of large erection lifts. The forward deckhouse (zone 17 in Fig. 3), which includes the living spaces, offices, and pilot house, was manufactured and outfitted complete with carpets, chairs, and curtains by a subcontractor and shipped by barge to the shipyard, where it is shown being placed aboard ship in a single 990-short-ton (900-metric-ton) lift.

To reduce the time required at the outfitting pier (the time period between launch and sea trials), the outfitting process is initiated as soon as feasible. For example, the final alignment of the propulsion machinery is accomplished as soon as all of the major structural welds are completed; this permits the propulsion machinery to be substantially ready to

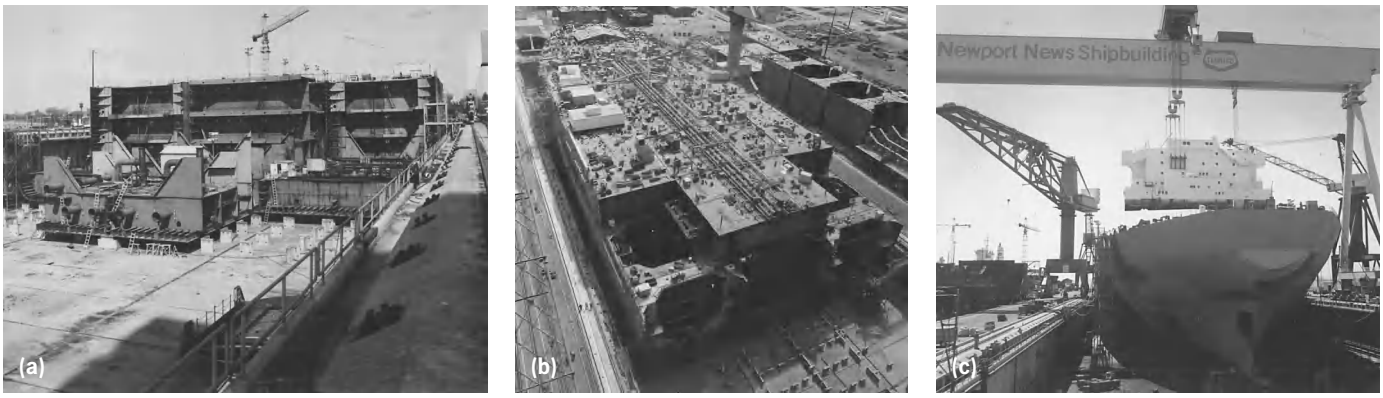


Fig. 6. Crude-oil tanker being constructed with preoutfitted erection units, showing (*a*) centerline erection unit, preoutfitted to contain the cargo-oil piping, (*b*) erection units being preoutfitted alongside the dock and aft of the ship, and (*c*) forward deckhouse being lowered into position. (*Newport News Shipbuilding*)

put to sea when the ship is launched. *See* MARINE MACHINERY.

Launching. The degree of completion that a ship has reached at the time of launching depends on many factors, such as the need for the shipway or dock being used for a subsequent ship, the relative ease of handling materials on the shipway or dock as compared with the outfitting pier, and the degree of completion of the boilers and other equipment which require cooling water for test purposes. However, in general, a ship is launched as soon as the hull structure is sufficiently complete to withstand the strain. Ships may be launched endwise, sidewise, or by in-place flotation. Endwise launchings were the most common method (and are still used) before the advent of side launchings (which are ideally suited for particular circumstances), and the more general use of flotation (for example, graving docks). The use of a graving dock requires a greater investment in facilities than either of the other two methods, but in some cases there may be an overall advantage due to the improved access to the ship and the simplified launch procedure. *See* DRYDOCKING.

Outfitting. The final outfitting of a ship is the construction phase during which checks are made to ensure that all of the previous work has been accomplished in a satisfactory manner; and last-minute details, such as deck coverings and the top coat of paint, are completed. It is considered good practice to subject as much of the ship as possible to an intensive series of tests while at the dock, where corrections and final adjustments are more easily made than when at sea. As a part of this test program, the main propulsion machinery is subjected to a dock trial, during which the ship is secured to the dock and the main propulsion machinery is operated up to the highest power level permissible. The full-power rating of the machinery cannot be developed during a dock trial because rated torque will not turn the propeller at the rated revolutions per minute with the ship restrained from forward motion. Also, for highly powered ships the strength of the pier may limit the power which can be developed. Still, many checks and tests can be made to ensure that the propulsion plant components and systems are correctly installed and function properly. For example, checks can be made to ensure that all relief valves are set at the correct pressure level and are functioning properly. Of particular concern are the extensive automation and remote control systems necessitated by the trend toward reduced crew levels.

When a comprehensive program of dockside tests has been completed, the only capabilities which have not been demonstrated are the operation of the steering gear during rated-power conditions and the operation of the main propulsion machinery at rated power; these capabilities must be demonstrated during trials at sea. However, the remaining equipment and systems can be operated at either actual or simulated rated conditions. For example, the anchor windlass cannot be operated with the anchor and the full scope of 165 fathoms (300 m) of chain extended at dockside; but the anchor and an added weight equivalent to the full scope of chain can be hoisted and lowered at dockside a sufficient number of times to demonstrate the operational readiness of the windlass.

After the ship has satisfactorily passed a comprehensive program of testing and has been determined to be seaworthy, it is ready to be taken on sea trials, followed by a delivery to the owners. *See* MERCHANT SHIP; NAVAL SURFACE SHIP; SUBMARINE.

R. L. Harrington

Bibliography. R. L. Harrington (ed.), *Marine Engineering*, 2d ed., 1992; E. V. Lewis (ed.), *Principles of Naval Architecture*, 3 vols., 2d ed., 1990; H. J. Pursey, *Merchant Ship Construction*, 1987; R. L. Storch et al., *Ship Production*, 2d ed., 1994; R. Taggart (ed.), *Ship Design and Construction*, 3d ed., 1980.

Shipping fever

A severe inflammation of the lungs (pneumonia) commonly seen in North American cattle after experiencing the stress of transport. This disease occurs mainly in 6–9-month-old beef calves transported to feedlots. Many infectious agents, both viruses and bacteria, cause symptoms of respiratory disease in

stressed cattle. However, the characteristic shipping fever pneumonia is caused primarily by the bacteria *Mannheimia* (formerly *Pasteurella*) *haemolytica* serotype A1 and, to a lesser extent, *Pasteurella multocida*; thus a synonym for shipping fever is bovine pneumonic pasteurellosis.

Both bacteria are normally present in low numbers in the nasal passages and tonsil area of unstressed cattle. However, with the stress of shipment and infection with respiratory viruses, these pathogenic bacteria replicate rapidly in the upper respiratory tract and are inhaled into the lungs, where pneumonia develops in the deepest region of the lower respiratory tract (pulmonary alveoli). Viruses can also concurrently damage pulmonary alveoli and enhance the bacterial pneumonia. *See* ANIMAL VIRUS; INFLAMMATION; PASTEURELLA.

Clinical signs. Symptoms of shipping fever occur 1–2 weeks after transport. Disease can occur in 50% or more of a group of cattle. Initially, cattle have reduced appetite, high fever, rapid and shallow respiration, depression, and a moist cough. During the later stages of disease, cattle lose weight and have labored breathing. A nasal and ocular discharge ranging from thin and watery to thick and yellow is usually present. Abnormal lung sounds are heard with a stethoscope. Without vigorous treatment, shipping fever can cause death in 5–30% of affected cattle.

The lesion causing these symptoms is a severe pneumonia accompanied by inflammation of the lining of the chest cavity, and is called a fibrinous pleuropneumonia. The chest cavity contains yellow, cloudy fluid, rich in plasma proteins (especially fibrin), and the lung surface is covered with yellow clots of fibrin. Large areas of the lungs become firm with red and gray mottling because pulmonary alveoli become filled with red blood cells, fibrin, and white blood cells. At autopsy, *Mannheimia haemolytica* can be cultured from affected lungs or pleural fluid. Other bacteria, notably *Mycoplasma bovis, Arcanobacterium* (formerly *Actinomyces*) *pyogenes*, and *Histophilus somni* (formerly *Haemophilus somnus*), are also isolated from cattle with shipping fever, especially in more chronic cases. *See* ACTINOBACILLUS; HAEMOPHILUS.

Virulence factors. *Mannheimia haemolytica* has at least three components, called virulence factors, that are responsible for causing pneumonia. The first is a bacterial cell wall–associated endotoxin that damages pulmonary alveolar capillaries, causing red blood cells, white blood cells, and plasma proteins (particularly fibrin) to escape from capillaries and fill the alveoli. The second is leukotoxin, released from the bacterium into alveoli. This toxin damages the cytoplasmic membranes of white blood cells that are normally present in the lung (pulmonary alveolar macrophages) and white blood cells that infiltrate the lung from the blood. Substances such as enzymes and altered oxygen molecules (free radicals) leak from damaged cells and injure alveolar lining cells (epithelium). Affected white blood cells eventually rupture. The third factor is a thick polysaccharide capsule that surrounds the bacterium and helps prevent the alveolar macrophages and infiltrating white blood cells from killing it. *Pasteurella multocida* also contains endotoxin and is surrounded by a thick capsule. *See* TOXIN.

Treatment and prevention. Treatment is aimed at eliminating bacteria, limiting the inflammatory reaction in the lung, and providing supportive care for cattle. To eliminate bacteria, diseased cattle are injected with antibacterial drugs, usually antibiotics. To reduce inflammation, nonsteroidal antiinflammatory drugs are given. These drugs reduce inflammation, pain, and fever but do not impair the immune response to the viruses and bacteria. Supportive care includes separating affected cattle from the stress of crowding, supplying adequate water, and feeding a ration high in roughage.

Vaccines can stimulate immunity to the viruses and bacteria associated with shipping fever. Although vaccination of cattle does not prevent shipping fever, it may reduce the number of cattle affected and the severity of the disease in individual animals. Viral vaccines should stimulate antibodies that neutralize viruses, by blocking their entry into cells and preventing their spread in the animal, thereby reducing susceptibility to bacterial infection. *Mannheimia haemolytica* vaccines should also stimulate production of two types of antibodies. The first prevents leukotoxin from damaging pulmonary alveolar macrophages and white blood cells, thereby decreasing damage to the lungs. The second attaches to the bacterial surface, allowing the bacteria to be more easily killed by pulmonary alveolar macrophages and white blood cells. *Pasteurella multocida* vaccines stimulate antibodies against the entire bacterium. Vaccines are often administered to cattle soon after shipment. This practice may be of some benefit; however, it does not allow enough time for cattle to develop an adequate antiviral and antibacterial antibody response before the onset of disease. As part of a disease control program called preconditioning, vaccines are given to cattle before shipment. This practice has the potential for better control of shipping fever. Those preconditioned cattle have antiviral and antibacterial antibodies before shipment, and they can better control viral and bacterial infections. *See* PNEUMONIA; RESPIRATORY SYSTEM DISORDERS; VACCINATION.

Anthony W. Confer

Bibliography. M. R. Ackermann and K. A. Brogden, Response of the ruminant respiratory tract to *Mannheimia (Pasteurella) haemolytica, Microbes Infect.*, 2:1079–1088, 2000; D. A. Mosier, Bacterial pneumonia, *Vet. Clin. North Amer. Food Anim. Prac.*, 13:483–493, 1997; B. P. Smith (ed.), *Large Animal Internal Medicine*, 3d. ed., 2001.

Shipworm

A wood-boring mollusk of the family Teredinidae. These borers, also called pileworms, are found around the world in tropical to boreal and marine to nearly fresh-water habitats, from the intertidal to depths of about 2000 ft (600 m). They arose in the

Cretaceous, and since then have been the most important organisms involved in recycling woody plant material entering the sea. It was not until humans became navigators that shipworms became pests. Before the advent of steel ships, they were greatly feared, and even today hundreds of millions of dollars are spent annually on prevention and repair due to their activity.

Teredinids are specialized to bore in woody plant material and utilize it as food. They are elongate and wormlike, with a small shell anteriorly and with siphons and pallets posteriorly (**Figs. 1** and **2**). Adult shipworms vary in length from about 0.8 in. to over 6.5 ft (2 cm to 2 m). There are 15 genera and about 70 species in this variable, specifically adapted, family. They bore by scraping filelike ridges on the anterior part of the shell against the anterior end of the burrow. The posterior or outer end of the burrow has a calcareous lining to which the muscles of the siphons and pallets are permanently attached. If conditions are adverse, the borer can withdraw the siphons and close the burrow entrance with the pallets. In this way they can survive for weeks under anaerobic conditions. *See* FEEDING MECHANISMS (INVERTEBRATE).

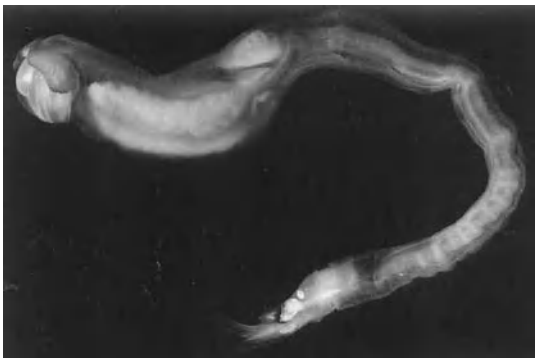


Fig. 2. Living *Lyrodus floridanus*.

Reproduction in the teredinids is similar to that in other bivalves. Fertilization of the egg may take place in the sea (oviparous species), or the eggs may be fertilized as they are deposited in the gills of the female, where they develop to the straight-hinge or first feeding stage (short-term brooders) or until they reach the pediveliger or settling stage (long-term brooders). The length of the free-swimming period is inversely proportional to the length of the brooding period. Tropical species may reproduce throughout the year, but most temperate and boreal species breed during the summer. At settlement the larvae are almost microscopic (200–350 micrometers); they rapidly penetrate the softened wood surface, complete metamorphosis, commence boring, and for the remainder of their life are confined to their burrow. *See* BIVALVIA.

Ruth D. Turner

Bibliography. E. B. G. Jones and S. Eltringham (eds.), *Marine Borers, Fungi and Fouling Organisms of Wood*, Organization for Economic Cooperation and Development, Paris, 1971; N. B. Nair and M. Saraswathy, The biology of wood-boring teredinid molluscs, *Advan. Mar. Biol.*, 9:335–509, 1971; S. P. Parker (ed.), *Synopsis and Classification of Living Organisms*, 2 vols., 1982; R. D. Turner, *A Survey and Illustrated Catalogue of the Teredinidae*, Spec. Publ. Mus. Compar. Zool., Harvard Univ., 1966.

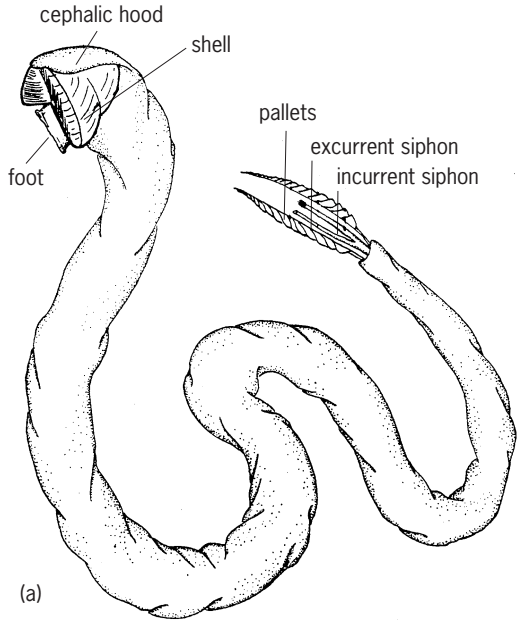


Fig. 1. *Bankia gouldi*. (*a*) Diagram showing position of shells, foot, pallets, and siphons (*after R. D. Turner, A Survey and Illustrated Catalogue of the Teredinidae, Spec. Publ. Mus. Comp. Zool., Harvard Univ., 1966*). (*b*) X-ray of collecting panel exposed for 6 months showing burrows with pallets and shells in place.

Shock absorber

Effectively a spring, a dashpot, or a combination of the two, arranged to minimize the acceleration of the mass of a mechanism or portion thereof with respect to its frame or support.

The spring type of shock absorber (**Fig. 1**) is generally used to protect delicate mechanisms, such as instruments, from direct impact or instantaneously applied loads. Such springs are often made of rubber or similar elastic material. Their design in relation to the natural frequency of the supported system and the forcing frequency of the applied load is most important. *See* SHOCK ISOLATION; SPRING (MACHINES).

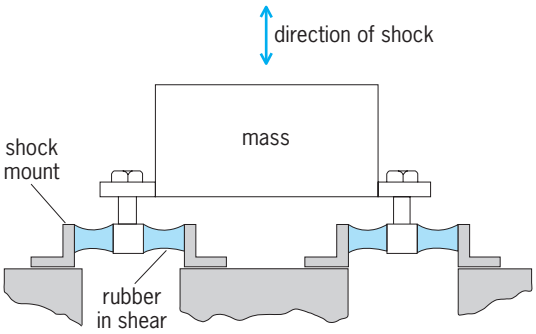


Fig. 1. Spring-type shock absorber.

The dashpot type of shock absorber is best illustrated by the direct-acting shock absorber in an automotive spring suspension system (**Fig. 2**). Here the device is used to dampen and control a spring movement. The energy of the mass in motion is converted to heat by forcing a fluid through a restriction, and the heat is dissipated by radiation and conduction from the shock absorber. *See* AUTOMOTIVE SUSPENSION; VIBRATION DAMPING.

There are also devices available which combine springs and viscous damping (dashpots) in the same unit. They use elastic solids (such as rubber or metal), compressed gas (usually nitrogen), or both for the spring. A flat-viscosity hydraulic fluid is used for the viscous damping.

L. Sigfred Linderoth, Jr.

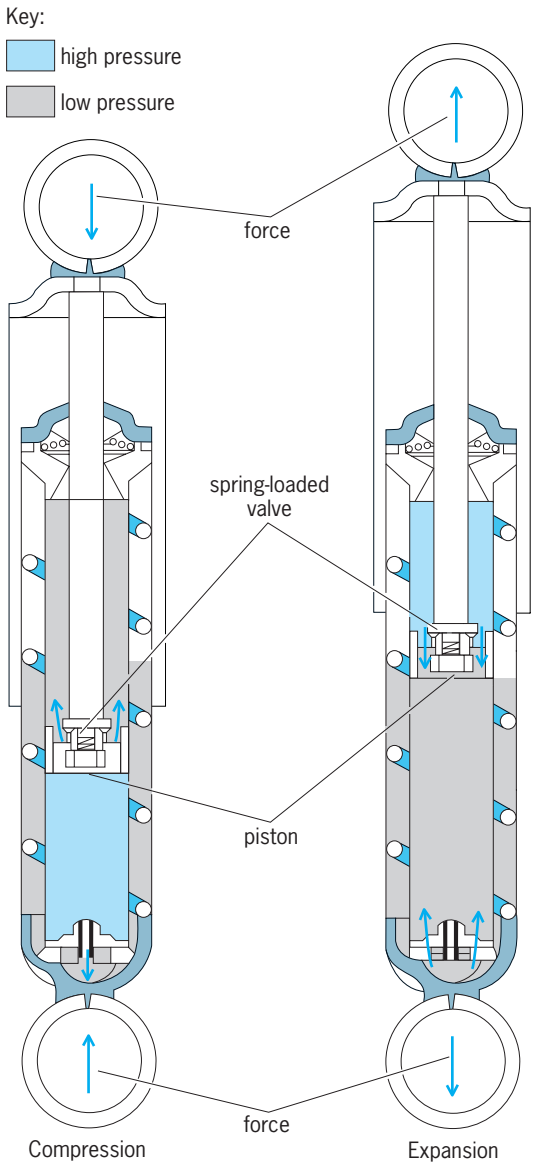


Fig. 2. Dashpot-type shock absorber. (*Plymouth Division, Chrysler Corp.*)

Shock isolation

The application of isolators to alleviate the effects of shock on a mechanical device or system. Although the term shock has no universally accepted definition in engineering, it generally denotes suddenness, either in the application of a force or in the inception of a motion. *See* SHOCK WAVE.

Shock isolation is accomplished by storing energy in a resilient medium (isolator, cushion, and so on) and releasing it at a slower rate. The effectiveness of an isolator depends upon the duration of the shock impact. An isolator may be effective in one case where there is a high G loading with a short duration, 0.001 millisecond or less, but may magnify the shock where there is a lower G loading but a longer duration (0.001–0.015 ms). The quantity G is equal to the so-called limit acceleration a divided by the acceleration of gravity g and is discussed later. Most shock isolators, also known as shock mounts or shock absorbers, that are available commercially are effective for the 0.001-ms or less interval.

Rubber is the most common material used in commercial shock isolators. Rubber isolators are generally used where the shock forces are created through small displacements. For larger displacement shock forces, such as those experienced by shipping containers in rough handling conditions, thick cushions of felt, rubberized hair, sponge rubber, cork, or foam plastics are used. Shock isolation systems which use the various cushion materials are generally custom designed to the particular application and cannot be considered from the standpoint of standardized isolators, but rather from the standpoint of the basic principles involved. *See* RUBBER.

Absorption of shock. The shock load must be divided between the case, the shock cushion, and the equipment. The case, since it must withstand effects of rough handling such as sliding and dropping, is by necessity rigid. The more rigid the case the closer to a 1:1 ratio will be the transfer of the shock from outside to inside. The absorption of the shock is primarily between the cushion and the equipment.

The dissipation of the energy of a 1-ft (0.3-m) drop with a cushion having a linear spring rate would require a thickness of 2 ft (0.6 m) of cushion. Since cushions of such thickness are not feasible, the equipment itself must withstand part of the shock. The cushion that is needed to dissipate the energy

from various heights of drop with equipment sharing part of the load may be determined by consideration of the principles which are involved. *See* SHOCK ABSORBER.

Limit acceleration. When a body moving with velocity v has to be stopped to complete rest, a deceleration (negative acceleration) must be applied. In order to make the stopping process smooth, a maximum value for the acceleration a is prescribed as a limit. Usually the full amount of the limit acceleration cannot be attained for the entire duration of the stopping process. However, an ideal process can be imagined with constant limit acceleration a. Deviation from this ideal case will be considered later.

Such uniformly decelerated motion is exactly the reverse process of uniformly accelerated motion. The same formulas apply.

The total distance traveled S is given by Eq. (1),

$$S = \frac{v^2}{2a} \tag{1}$$

where v is the initial and final velocity. When the velocity to be stopped is produced by a free fall, the necessary height of drop H must be as given by Eq. (2), where g is the gravitational acceleration. From these formulas Eq. (3) is obtained. Since the

$$H = \frac{v^2}{2g} \tag{2}$$

$$v^2 = 2aS = 2gH \tag{3}$$

velocity at the end of the free fall is the same as it was at the beginning of thestopping process, Eq. (4)

$$2aS = 2gH \tag{4}$$

is obtained. It is customary to give the limit acceleration in the form of a G value, defined by $G = a/g$. With this notation, the stopping distance can be written in the form given in Eqs. (5).

$$S = H\frac{g}{a} \qquad S = \frac{H}{G} \tag{5}$$

Force-distance diagrams. It is useful to analyze the ideal case again from another aspect. The energy E stored in the falling body after fall from height H is $E = WH$, where W is the weight of the moving body. The same amount of energy must be consumed during the stopping period.Therefore, Eq. (6) is ob-

$$E = GWS \tag{6}$$

tained. The active force is GW in this stopping phase. Equating the last two equations for the energy yields Eq. (1). As represented in **Fig. 1**, the energy can be visualized as the area under the curve in a force-distance diagram. The curve is a horizontal line in the case of constant deceleration.

Whatever device is used for checking the velocity, it is hardly possible to obtain a constant deceleration

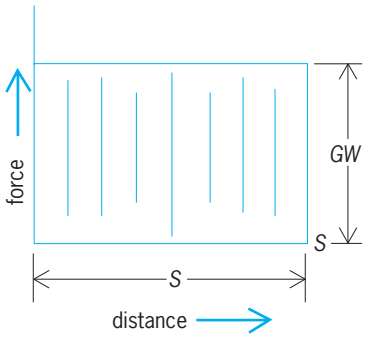


Fig. 1. Force-distance diagram for ideal case.

of exactly the limit value. Therefore, allowances have to be made for practical considerations.

If the opposing force is provided by an ordinary spring, no constant force is produced. The force F is built up gradually with the compression of the spring, as shown in **Fig. 2**. The area under the force-displacement curve is in this case a triangle, representing the energy given in Eq. (7). The kinetic

$$E = {}^1/_2 GWS_1 \tag{7}$$

energy from the fall is the same as before, $E = WH$. Equation (8), the stopping distance for the spring of

$$S_1 = \frac{2H}{G} \tag{8}$$

Fig. 2, is found by equating Eqs. (6) and (7). This means a distance twice as long as that for the ideal case of Fig. 1 is needed. This less favorable condition is caused by the fact that the permitted maximum force is used only at the end of the process. The diagram of Fig. 2 is not '"filled up" completely; it is only half-filled.

Preloaded springs have a better-filled diagram, as shown in **Fig. 3**. An inherent disadvantage of the

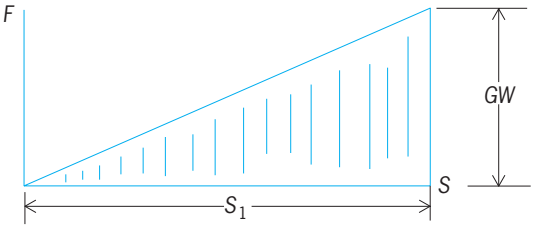


Fig. 2. Force-distance diagram for ordinary spring.

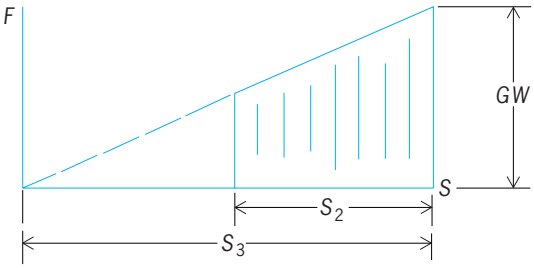


Fig. 3. Force-distance diagram for preloaded spring.

better-filled-up spring diagrams is the steep slope at the beginning, which means that the spring system must be designed for a predetermined shock. This design does not help in isolating small shocks. It has not been proved which shocks do the most damage—the large shock that happens occasionally or the small, repeated shock that occurs almost continuously. From the isolation standpoint, the spring characteristic of an isolator should start up with a moderate slope and continue with gradually sharper increase of force. This is why a system representing a completely filled diagram cannot be applied, and therefore a certain deviation from the ideal condition represented by Eq. (1) is unavoidable.

Damping forces are helpful for attaining longer deceleration force at the beginning of the stopping process. Pure viscous damping alone would result in an elliptical diagram like the one shown in **Fig. 4**.

Since the high velocity at the beginning produces large opposing forces, spring force and damping effects can be combined to make a well-filled diagram, as shown in **Fig. 5**.

Practical shock absorbers usually fall between the two cases of completely filled diagrams and half-filled diagrams. The stopping distance provided therefore lies between Eqs. (5) and (8). The value given by Eq. (8) is considered conservative and therefore is recommended, since a certain tolerance is necessary.

Using Eq. (8) and assuming that the equipment can withstand 30*G*, the operating deflection of the cushion under a 24-in. (60-cm) drop must be as given by Eq. (9). Since *S* is the operating distance, the total

$$S = \frac{2 \times 24}{30} = 1.6 \text{ in. } (4 \text{ cm}) \qquad (9)$$

thickness of cushion will depend on the compression ratio of the material.

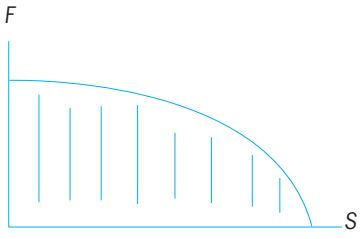


Fig. 4. Force-distance diagram for pure viscous damping.

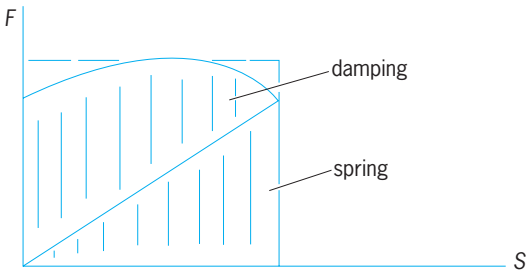


Fig. 5. Force-distance diagram for combined effects of spring force and damping.

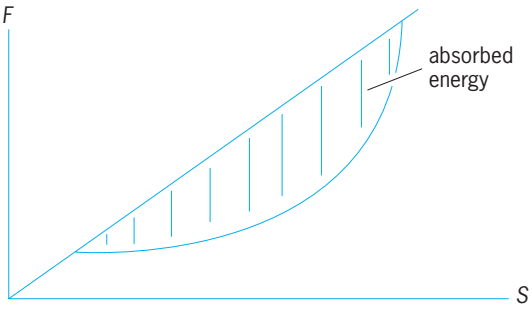


Fig. 6. Diagram of absorbed energy for ideal spring system with damping.

The preceding equation provides the operating thickness needed to bring a given mass to rest with a linear spring cushion, leaving the equipment to share 30*G* of the impact load. It must be remembered that cushions, springs, and so forth store energy and, depending upon their inherent friction, return this energy. Thus, in an ideal spring system, no energy is absorbed. However, by adding damping or by using a material with inherent damping, there is provided a resilient medium (**Fig. 6**).

To accomplish such a system requires engineering information on the cushioning materials under dynamic conditions. Once their behavior is known, the materials that have the desired force-displacement characteristics can be selected, and the damping forces may be added to provide a shock-absorbing system approaching the effectiveness of the one shown in Fig. 6. *See* DAMPING; SPRING (MACHINES); VIBRATION DAMPING. K. W. Johnson

Bibliography. C. M. Harris and C. E. Crede (eds.), *Shock and Vibration Control Handbook*, 4th ed., 1995.

Shock syndrome

A clinical condition caused by widespread impairment of blood flow resulting in an inadequate supply of oxygen to the body's tissues. Shock is a sudden and progressive condition that, if not reversed, can lead to death. Shock affects the entire body, as opposed to other conditions where blood flow is reduced to only a portion of a single organ, such as stroke or heart attack. It is typically, but not always, associated with a low blood pressure that can be felt as a weak pulse.

Shock can be categorized according to the initiating event: hypovolemic (loss of circulating blood volume), cardiogenic (loss of cardiac pumping activity), distributive (shunting of blood flow away from the tissues), or obstructive (a blockage of blood flow through the heart or major vessels).

Regardless of the initiating event, shock will incite responses from the body. These responses will often maintain blood flow to vital organs for a period of time but at the expense of metabolic stress and exhaustion in the tissues. The nervous system reacts to shock by sending signals to the blood vessels in

nonvital organs (such as the skin and bowel) that constrict these blood vessels, reduce blood flow to those organs, and allow blood to go preferentially to the vital organs (heart and brain). The nervous system also increases the heart rate in an attempt to pump more blood.

As shock progresses, the metabolically active tissues are starved for oxygen (a condition termed ischemia) and change their metabolic processes from those that use oxygen (aerobic) to those that do not require it (anaerobic). Anaerobic metabolism is a very inefficient process. Once cells are unable to maintain basal metabolic activities, the cell dies and releases harmful intracellular substances.

Advanced shock is generally associated with impaired cardiac contractility (even if this was not the initial cause) and slugging of blood flow in the microcapillaries. By this time, shock has progressed to the degree that, despite reversal of the inciting process, the impaired tissues do not recover and death results.

Hypovolemic shock. Hypovolemic shock occurs when the circulating blood volume (approximately 70 to 80 mL/kg of body weight or 5 to 6 L in an adult human) is reduced below a critical level. The most common cause of blood loss (termed hemorrhage) occurs when a hole develops in the wall of a major blood vessel and blood escapes either outside the body (external hemorrhage) or to spaces inside the body (internal hemorrhage). The amount of sudden blood loss necessary to produce shock is around 40% of the total blood volume, although this threshold varies according to the individual's physiologic condition and the rapidity of loss.

Blood is approximately 40% red blood cells and 60% plasma, a mixture of water, salt, and proteins. Shock is possible if large amounts of water are lost, as might occur from severe vomiting or diarrhea.

Hypovolemic shock is classically associated with cool, pale skin (due to reduced blood flow to the skin), sweating (due to stimulation of the nervous system), and a fast heart rate (an attempt to pump more blood).

The treatment of hypovolemic shock consists of rapid restoration of circulating blood volume, initially with an isotonic solution of water and salt (termed crystalloid), followed by replacement of lost red blood cells through blood transfusions.

Cardiogenic shock. Cardiogenic shock results when the pumping activity of the heart is suddenly reduced. The main pumping chamber of the heart is the left ventricle. Cardiogenic shock becomes common when approximately 40% of the left ventricular heart muscle loses its ability to contract. The most common reason for cardiogenic shock is a sudden blockage (occlusion) of blood flow in a major artery (coronary artery) suppling blood to the working heart muscle. Even though the heart pumps blood through its four chambers, the heart muscle is, like every other tissue, dependent on blood vessels that course within the muscle itself for its oxygen supply.

A sudden occlusion of a coronary artery (heart attack) is usually caused by a blood clot forming on the site of a long-standing narrowing due to the deposition of cholesterol underneath the inner lining of the artery. Cardiogenic shock is usually associated with pain from the heart attack as well as the previously mentioned signs of shock.

The treatment of cardiogenic shock involves restoration of blood flow through the occluded coronary artery (reperfusion). Reperfusion is possible using drugs which dissolve the occluding blood clot (termed thrombolytics), by mechanically breaking up the blood clot that occludes or narrows the artery using catheters advanced through the circulatory system, or by surgically establishing a route for blood flow around the occluding blood clot (termed coronary artery bypass grafting).

Distributive shock. Distributive shock results from widespread impairment of blood flow to the tissues despite adequate circulating blood volume and cardiac pumping activity. This may appear paradoxical, but in reality it is blood flow though the capillaries that is most important. Only in the capillaries is oxygen delivered to the metabolically active cells. If blood flow is shunted away from the capillaries, then oxygen does not reach the cells and shock may develop. Portions of the nervous system regulate the distribution of blood flow and can finely adjust blood flow to balance the metabolic needs of local tissues.

There are several conditions that lead to distributive shock. A major impairment of neurologic regulation, termed neurogenic shock, is one. Nerve signals pass from the brain, down the spinal cord, and out into the peripheral nerves. Damage to the spinal cord will impair the transmission of these regulatory signals, with the greatest effect if the damage is toward the upper portion of the spinal cord. Thus, patients who sustain spinal cord injury, usually in the neck region, are at risk for developing neurogenic shock.

Infections can produce distributive shock from the toxic substances elaborated by the infecting organisms and by the body's response to the infection. Bacteria produce some of the more severe infections and stimulate the body to produce several chemical substances that impair circulation. The body's response to infection is called sepsis, and if shock develops the process is called septic shock.

Severe allergic reactions can produce widespread improper distribution of blood flow, a process termed anaphylaxis or anaphylactic shock. Such severe allergic reactions may result from insect stings (such as by bees and wasps), foods (such as nuts and shellfish), or medications (for example, penicillin).

Common clinical characteristics of distributive shock are the tendency for skin perfusion to be maintained initially (warm, pink, even flushed) and an absence of sweating. Patients with neurogenic shock have impaired innervation to the heart, so the heart rate is typically low or normal. Patients in septic shock often have a fever. Patients in anaphylactic

shock may have other signs of allergic reaction (such as hives or wheezing).

The initial treatment of distributive shock is the expansion of circulating volume with isotonic crystalloid. The amounts required to restore perfusion may be large: 20 to 40 mL/kg or 2 to 4 L in a typical adult. Many patients will require the administration of drugs that stimulate the cardiovascular system to increase blood flow and improve tissue perfusion.

Obstructive shock. Obstructive shock results from a mechanical blockage of blood flow through the heart or major blood vessels in the chest. For example, a blood clot can form in a leg vein, break off, and travel with the blood from the leg to the heart. It will typically pass through the heart and its valves but become lodged in the smaller pulmonary artery that leads from the heart to the lungs. If the obstruction is large enough, blood flow from the heart is impaired and shock develops.

Other mechanical conditions that may occlude blood flow are the presence of a constriction around the heart that limits pumping action (pericardial tamponade) and blockage of blood flow out of the heart into the major artery of the body (aortic dissection).

Patients with obstructive shock have typical features of shock (cool pale skin, sweating, and fast heart rate) in addition to features specific to the underlying event. The treatment of distributive shock is removal of the mechanical blockage to blood flow, which usually requires a surgical procedure. *See* BLOOD; CARDIOVASCULAR SYSTEM; RESPIRATION.

J. Stephan Stapczynski

Bibliography. D. Annane, E. Bellissant, and J. M. Cavaillon, Septic shock, *Lancet*, 365(9453):63–78, January 1, 2005; C. A. Graham and T. R. J. Parke, Critical care in the emergency department: Shock and circulatory support, *Emerg. Med. J.*, 22(1):17–21, January 2005; F. A. Moore, B. A. McKinley, and E. E. Moore, The next generation in shock resuscitation, *Lancet*, 363(9425):1988–1996, June 12, 2004; M. Mullner et al., Vasopressors for shock, *Cochrane Database Syst. Rev.*, 3:CD003709, 2004; T. A. Sanborn and T. Feldman, Management strategies for cardiogenic shock, *Curr. Opin. Cardiol.*, 91(6):608–612, November 2004.

Shock tube

A laboratory device for rapidly raising confined samples of fluids (primarily gases) to preselected high temperatures and densities. This is accomplished by a shock wave, generated when a partition that separates a low-pressure (driven) section from a high-pressure (driver) section is rapidly removed. In most shock tubes the two sections are separated by a diaphragm that supports the pressure differential until it is punctured (**Fig. 1***a*). Shock tubes operated by rapidly opening valves also have been developed. The tubes can be circular, square, or rectangular in cross section, with diameters that range from 0.8 to 24 in. (2 to 60 cm). The lowest gas pressure to be used in the driven section is such that the corresponding molecular mean free path is less than several hundreds of the diameter of the driven tube; the length of that section should be 40–60 times its diameter. Driver lengths vary, depending on the designed test time and the desired ultimate pressure in the shock tube. Concurrently with propagation of the high-pressure (heating) wave from the partition toward one end wall (Fig. 1*b*), an expansion (cooling) wave propagates in the opposite direction. Although shock waves can be generated under laboratory conditions by means of a variety of devices, such as ballistic pistons, electrical discharges, distributed explosives, and the impact of hypersonic flows on stationary targets, shock tubes are the simplest to construct and to operate.

In the driven section, gaseous samples can be heated to temperatures as high as 26,500°F (15,000 K) under strictly homogeneous conditions. The material in any tube cross section is thus rapidly and uniformly heated, except for a very thin layer adjacent to the walls, for a duration that depends

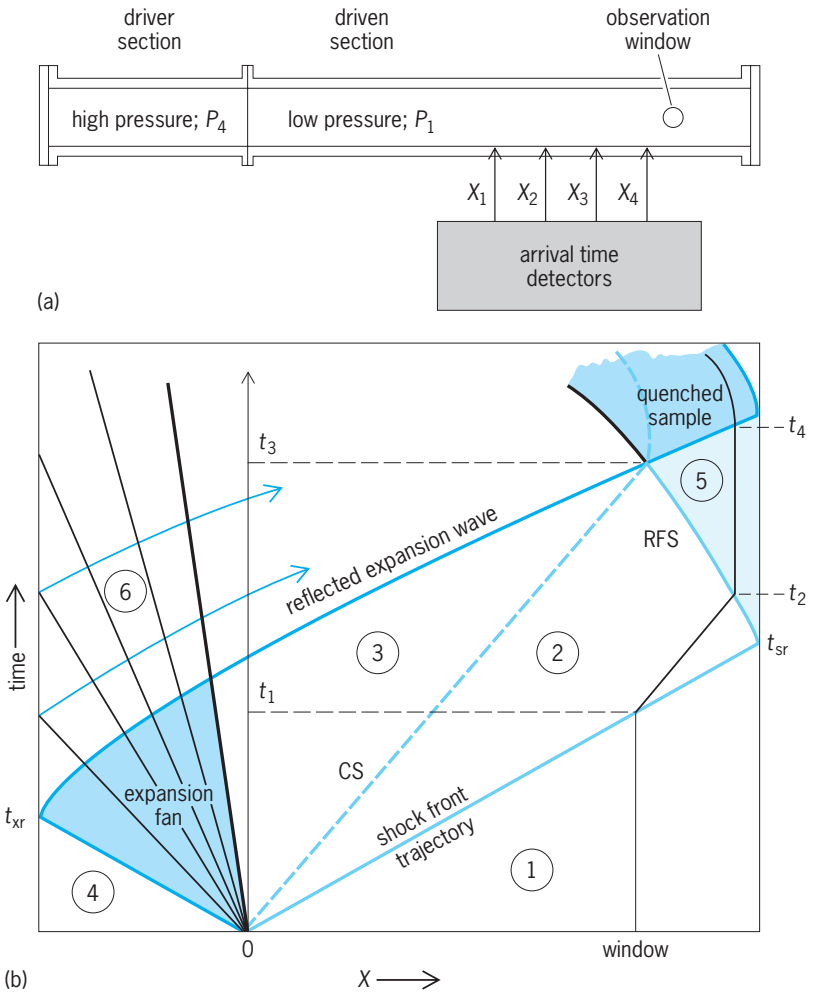


Fig. 1. Shock tube. (*a*) Schematic. (*b*) Flows in a shock tube. The special conditions in a single-pulse tube are shown, and the circled numbers identify the different regions.

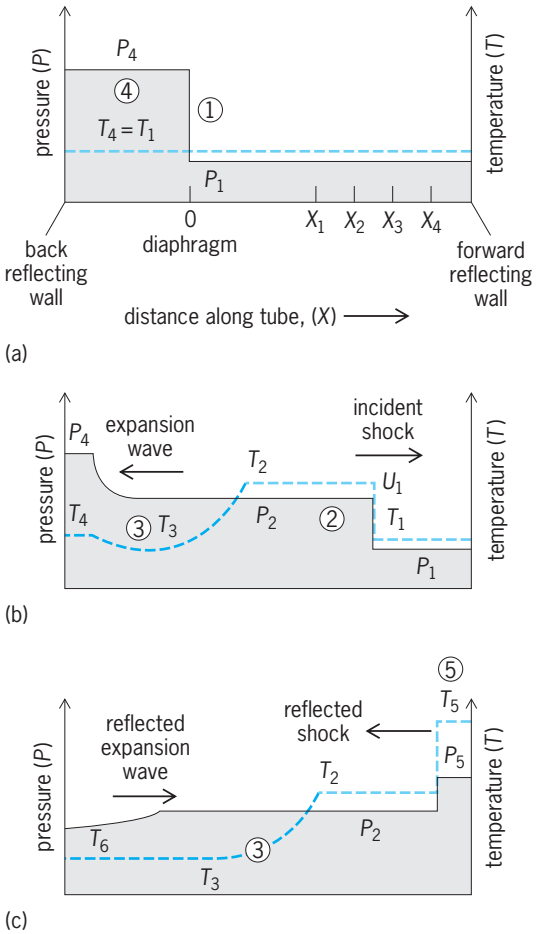


Fig. 2. Schematic profiles of pressure and temperatures along the shock tube at selected times (*t*). Circled numbers (regions) correspond to those shown in Fig. 1*b*. (*a*) At $t = 0$. (*b*) At $t = t_1$. (*c*) At $t \geq t_{sr}$.

on its location along the tube. At the shock front the heating rate is about 180°F/microsecond (100 K/μs), and the transition from the unshocked to the high-temperature condition is of short but of finite duration; with high-resolution diagnostic devices it is possible to separate, in time, sequences of relaxations. The incident shock wave is reflected at the tube terminus (t_{sr}), generating a slower-moving wave (RFS) that additionally heats the compressed gas. The elevated temperature can be maintained for times of the order of milliseconds. The expansion wave is also reflected at the (left) terminus (t_{xr}). Upon convergence of the two reflected waves (t_3), the processed material is quenched at cooling rates of 0.9–9.0°F/μs (0.5–5 K/μs). On the high-pressure side the expansion wave never attains a steep front (**Fig. 2**); the cooling rate is both time- and position-dependent. *See* SHOCK WAVE.

Operation. With the diaphragm in place, both sections of the shock tube are evacuated; in particular, the driven section must be well pumped to minimize contamination of the sample by residual gases and material desorbed from the walls. That section is then filled to a modest pressure [$P \approx 0.01$–1.0 atm or 10^3–10^5 pascals at initial temperature (T_1)] with the mixture under investigation. Most often this mixture consists of 90–99.9% argon (or a heavier inert gas), plus the gas of interest. The driver section is filled to a high pressure ($P_4 \approx 10$–100 atm or 10^6–10^7 Pa) with a low-molecular-weight gas: helium or hydrogen. When the diaphragm is rapidly ruptured, expansion of the high-pressure gas acts as a low-mass piston that generates a steepening pressure front, which moves ahead of the boundary between the driver and the test gases (trajectory CS, Fig. 1*b*). A planar shock develops in 6–10 tube diameters: it moves at a hypersonic speed (U_1), determined by the pressure ratio (P_4/P_1) and the heat capacity ratio [$\gamma_{test}/\gamma_{driver}$: γ is the ratio of the specific heat at constant pressure (C_p) to the specific heat at constant volume (C_v)]. The shock speed is measured by recording its arrival times at several precisely located stations (X_1, X_2, . . .); the transition from P_1 to P_2 occurs over a distance equivalent to about five mean free paths, measured in the unshocked gas: 10^{-8} to 10^{-7} cm for $P_1 \approx 10$ torr. The test sample is raised to temperature = T_2, density = ρ_2, and pressure = P_2, thereby initiating vibrational or electronic excitation, isomerization, bond scission, ionization, or bimolecular reactions.

Shock tubes are unique reactors: on the one hand, they may be used to investigate the gas dynamics of shocks; on the other hand, they may be used for the preparation of test samples, for equilibrium or kinetic study of the conversion of materials at elevated temperatures (or at low temperatures in the driver section). Multifold diagnostic devices are in use for tracking the changes of state that occur in the processed materials with microsecond resolution. Characteristic spectral line sources may be used to chart the evolution of specific absorbing species on a state-by-state basis. Rapid response detectors, photovoltaic and photoconducting elements, as well as photoelectron multipliers, coupled to fast digitizers and computers, permit the concurrent recording of vast amounts of data for each shock. Appropriate spectrographs and detectors permit measurement of intensities and wavelengths of the emitted light. With selected frequency sources, particularly lasers, specific absorption spectra and changes in refractive index can be recorded. Spectrometers equipped with parallel-array detectors permit the acqusition of time-dependent spectra at numerous wavelengths concurrently. In some experiments the shock-heated gas is expanded, through a tiny aperture at the terminus of the test section, into a rapid-scan mass spectrometer for gas analysis. In a single-pulse shock tube a ballast tank is attached to the tube, close to the diaphragm on the low-pressure side, with an aperture comparable to the tube diameter. This greatly attenuates subsequent reflected shock waves and permits single-pulse processing of the driven gas from region 5 (Fig. 1*b*). A sample of the heated and subsequently quenched material may then be extracted for extended chemical analysis. *See* MASS SPECTROMETRY.

Thermodynamic processes. The shock wave carries energy and momentum. Hence, when it reaches the end plate it is reflected as another shock wave, now propagating into the previously heated gas. Additional heating (to T_5) and compression (to ρ_5) occur. At the observation window, the incident shock front is seen first at t_1, followed by samples that were exposed for various times to T_2. The sample of gas initially at the observation point is swept toward the end plate. It is exposed to T_2 for the period $t_1 \rightarrow t_2$; then to T_5 for the period $t_2 \rightarrow t_4$, when it is quenched by the rarefaction wave that met the reflected shock wave at t_3. In a properly designed shock tube, such samples may be extracted for detailed chemical analysis, since they have been exposed to a single high-temperature pulse of known magnitude (T_5) and duration ($\tau = t_4 - t_2$). The production of a large number of chemical species at high temperatures, integrated over τ, can thus be ascertained. Figure 2 shows schematic profiles of pressures and temperatures along the shock tube at selected times.

The thermodynamic state of the shock-heated sample is determined by the equation of state of the test gas (over the full temperature range); the conservation conditions of mass, momentum, and energy for the material that passes through the shock front; and the velocity of the shock wave, which is equal but opposite to the mass velocity with which molecules enter the front. The measured shock speed is usually expressed as a dimensionless Mach number ($M_1 = U_1/\alpha_1$); $\alpha_1 = (\gamma\ P_1/\rho_1)^{1/2}$ is the velocity of sound in the unshocked test gas. The pressure jump across the shock front (P_2/P_1) is proportional to M_1, as in Eq. (1). For ideal gases, the ratios for density (ρ) and temperature (T) are given by Eqs. (2) and (3). The higher the value of γ, the larger these ratios.

$$\frac{P_2}{P_1} = \frac{2\gamma M_1^2 - (\gamma - 1)}{(\gamma + 1)} \tag{1}$$

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma + 1)M_1^2}{2 + (\gamma - 1)M_1^2} \tag{2}$$

$$\frac{T_2}{T_1} = \frac{\left[2\gamma M_1^2 - (\gamma - 1)\right]\left[2 + (\gamma - 1)M_1^2\right]}{(\gamma + 1)^2 M_1^2} \tag{3}$$

The Mach number M_1 increases with the initial pressure ratio across the diaphragm (P_4/P_1) and with the sound speed of the driver gas (α_4). Thus, $M_1 = 2.8$ for argon driving into argon, when $P_4/P_1 = 10^3$. The same ratio develops $M_1 = 6$ for helium into argon, and $M_1 = 8.3$ when hydrogen drives argon. Since the enthalpy and heat capacity of the material of interest progressively change during relaxation to equilibrium at T_2, argon, which is an inert diluent and has a high γ (1.67), is used as the major component in the test gas. Then, the drift in the parameters of state due to the presence of 0.1–1.0% of the component of interest is minimized. *See* ENTHALPY; FLUID FLOW.

Applications. The many gas dynamic problems that can be investigated in shock tubes include thermal boundary-layer growth, shock bifurcation, and shock-wave focusing and reflection from plane and wedge-shaped surfaces. However, shock tubes have their greatest utility as devices for preparing gases for study at very low or very high temperatures without thermal contact with extraneous surfaces. For example, if condensable vapors are commixed with the driver gas, upon expansion (Figs. 1 and 2) they attain various degrees of supersaturation prior to homogeneous condensation. Nonequilibrium condensation of methanol vapor has thus been measured at the shock-tube end wall behind reflected expansion waves.

The optical and spectroscopic parameters of samples that attain thermodynamic equilibrium after the passage through the shock front can be measured: for example, line intensities in emission and absorption, effects of collisional broadening and the presence of ions on spectral line contours, indices of refraction, and microwave attenuation. Conversely, when the indices of refraction or the spectroscopic parameters of the various species suspected of being present in the shock-heated gases are known (or can be reliably estimated), their measurement provides information on the composition and hence the thermodynamic state of the gas.

Shock-tube techniques are particularly useful for studying the sequence of states through which a sample passes after its temperature is precipitously raised from T_1 to T_2 (or from T_2 to T_5). The evolution in time of the post-shock-front samples can be followed to address a variety of questions, ranging from simple ones (such as how long does it take for the initial high-translational, low-vibrational energy partition to equilibrate to a uniform temperature) to very complex problems [such as what is the kinetic mechanism that determines ignition delays in a combustible mixture of fuel and oxidizer; specifically, the effect of functional groups on ignition delays in H_3CX/O_2 mixtures (X = CH_3, CH_2, NO, CN, Br) was found to correlate with the C-X bond dissociation energy].

The gas density in the vicinity of the shock front, $\rho(X)$, can be scanned by recording, interferometrically, the profile of the index of refraction. A more sensitive measurement may be obtained from the gradient of the index of refraction. Then, from the measured shock speed (U_1) the time-dependent change in the density immediately past the shock front can be derived. This reflects the change in the local temperature, and consequently the shuffling of enthalpy from the translational to the internal molecular degrees of freedom.

One of the first significant applications of shock tubes to chemical kinetics was the determination of diatom dissociation rates; this was followed by studies of small, then larger, polyatomic molecules. Such processes could be investigated only with difficulty in conventional high-temperature reactors, either static or flow type, because of unavoidable

concurrent complex reactions that take place at hot reactor walls. Isomerization reactions constitute another simple type of conversion; for example, cyclopropane → propene; *cis*-butene-2 → *trans*-butene-2. For these the measured rates fit well with values extrapolated from kinetic parameters obtained at lower temperatures. *See* MOLECULAR ISOMERISM; MOLECULAR STRUCTURE AND SPECTRA.

It is possible to generate mixtures of highly supersaturated metal vapors by shock-heating volatile precursors diluted with argon (Ar) [iron from $Fe(CO)_5$; bismuth from $Bi(CH_3)_3$; silicon from SiH_4]. Initial precursor pressures are selected such that complete decomposition occurs rapidly in the postshock region (T_5), thus generating a highly supersaturated mixture of metal/Ar species. Then condensation occurs to produce minute metallic clusters (up to 10–30-nanometer radius). The kinetics of such processes can be followed by recording the intensity of scattered laser light and sample turbidity. The results of these experiments proved to be inconsistent with classical nucleation theory.

The use of shock tubes as a research tool for investigations of gas dynamics, determination of parameters of state, spectroscopy, and chemical kinetics—at high temperatures—is detailed in an ongoing series, Proceedings of Shock Tube Symposia, held biannually. Aspects that focus on chemical kinetics have been summarized in two reviews (1990 and 1993). Therein are listed over 350 specific reactions for which temperature-dependent rate constants were measured over 800–2500 K; 600 references to the literature are included. In one review, by W. Tsang and A. Lifshitz (1990), chemical kinetics studies were classified into five categories, based on the diagnostic devices that were used. In one of the major techniques the rates of production (or destruction) of atomic species were measured with microsecond resolution, by recording their atomic resonance absorptions, which is both highly specific and very sensitive; for example, notation (4).

$$\begin{gathered} \underline{H} + CH_4 \longrightarrow H_2 + CH_3 \\ k = 10^{14.9} \exp(-7577/T) \\ CH_3 + Ar \longrightarrow CH_2 + \underline{H} \\ k = 10^{16.3} \exp(-46{,}100/T) \end{gathered} \tag{4}$$

Broad-band absorptions and emissions have also been used to follow the evolution (or decay) of molecules or radicals, such as notation (5).;

$$\begin{gathered} \underline{CH_3I} \rightarrow CF_3 + I \\ k = 10^{14.67} \exp(-27{,}200/T) \end{gathered} \tag{5}$$

The "single pulse" configuration that develops a rapid quench of the shock-heated gas (Fig. 1*b*) has been extensively exploited for investigations of fragmentation sequences of large molecules at high temperatures. Free-radical precursors have been pyrolized to generate initially the desired free radicals, and their reactions with commixed reagents followed either spectroscopically or by analysis of the distribution of products in the quenched samples. Kinetic parameters for many reactions, derived via three additional diagnostic procedures, were also listed: use of laser-schlieren (high-resolution laser diagnostic technique) for measuring density gradients at shock fronts (to determine enthalpy flow among the heated gas components), mass-spectrometric sampling of the shock-processed gas, and determination of induction periods for the onset of combustion when mixtures of fuel and oxidizer were shock-heated.

Simon H. Bauer

Bibliography. E. F. Greene and J. P. Toennies, *Chemical Reactions in Shock Waves*, 2d ed., 1964; J. V. Michael and K. P. Lim, *Annu. Rev. Phys. Chem.*, 44:429 1993; W. Tsang and A. Lifshitz, *Annu. Rev. Phys. Chem.*, 41:559, 1990.

Shock wave

A mechanical wave of large amplitude, propagating at supersonic velocity, across which pressure or stress, density, particle velocity, temperature, and related properties change in a nearly discontinuous manner. Unlike acoustic waves, shock waves are characterized by an amplitude-dependent wave velocity. Shock waves arise from sharp and violent disturbances generated from a lightning stroke, bomb blast, or other form of intense explosion, and from various types of supersonic flow over stationary objects. Extraterrestrial examples include supernovae expanding against neighboring intergalactic gas clouds and solar wind flowing over magnetized planets and satellites. This article discusses shock waves in gases and in condensed materials.

Gases

The abrupt nature of a shock wave in a gas can best be visualized from a schlieren photograph or shadow graph of supersonic flow over objects. Such photographs show well-defined surfaces in the flow field across which the density changes rapidly, in contrast to waves within the range of linear dynamic behavior of the fluid. Measurements of fluid density, pressure, and temperature across the surfaces show that these quantities always increase along the direction of flow, and that the rates of change are usually so rapid as to be beyond the spatial resolution of most instruments except when the shock wave is very weak or at very low gas densities. These surfaces of abrupt change in fluid properties are called shock waves or shock fronts. *See* SCHLIEREN PHOTOGRAPHY; SHADOWGRAPH; WAVE MOTION IN FLUIDS.

Shock waves in supersonic flow may be classified as normal or oblique according to whether the orientation of the surface of abrupt change is perpendicular or at an angle to the direction of flow. A schlieren photograph of a supersonic flow over a blunt object is shown in **Fig. 1**. Although this photograph was obtained from a supersonic flow over a stationary model in a shock tube, the general shape of the shock wave around the object is quite typical of those observed in a supersonic wind tunnel, or of similar objects (or projectiles) flying at supersonic speeds in a stationary atmosphere. The shock wave in this case assumes an approximately parabolic shape and

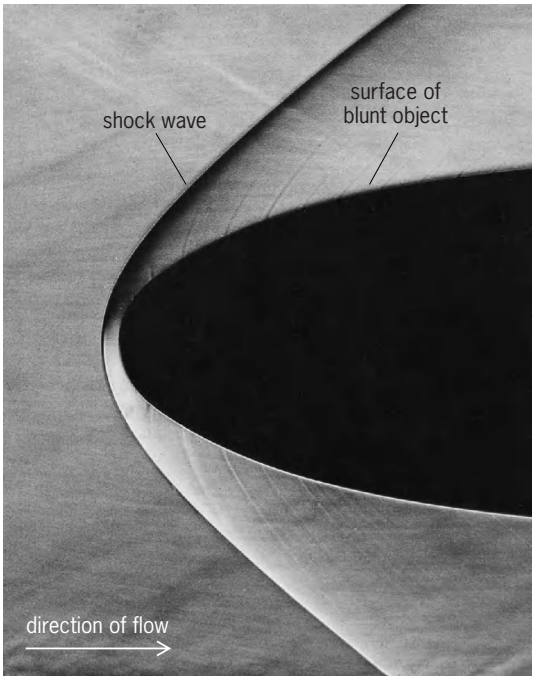


Fig. 1. Schlieren photograph of supersonic flow over blunt object. Shock wave is approximately parabolic, and detached from object. (*Avco Everett Research Laboratory, Inc.*)

is clearly detached from the blunt object. The central part of the wave, just in front of the object, may be considered an approximate model of the normal shock; the outer part of the wave is an oblique shock wave of gradually changing obliqueness and strength. The stand-off distance Δ (that is, the closest distance between the shock wave and the object) is governed by the density ratio across the normal shock, ρ_2/ρ_1, which increases with the upstream

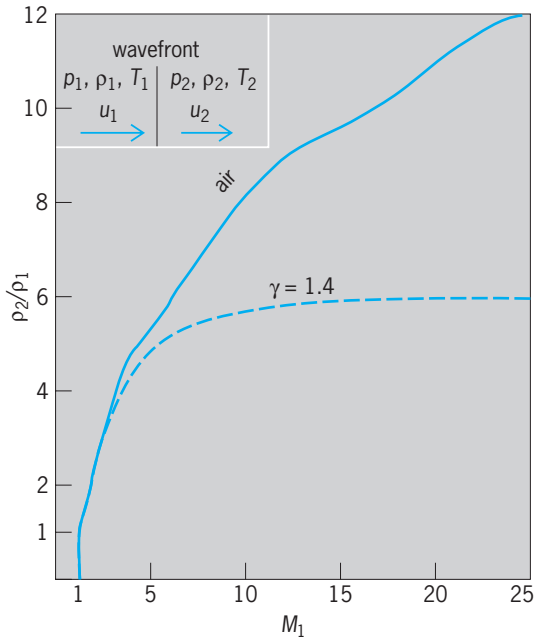


Fig. 2. Density ratio ρ_2/ρ_1 across normal shock wave in air at standard atmospheric density as a function of Mach number M_1. Inset shows variables in case of steady flow across a stationary wavefront.

Mach number, M_1, and varies with gas composition (**Fig. 2**). At high Mach numbers, the value of Δ becomes a small fraction of the nose radius of the object, R_N, such that Δ/R_N is approximately equal to ρ_1/ρ_2. For supersonic flow over sharp objects, such as slender cones and thin wedges, Δ approaches zero as the nose radius R_N vanishes. The shock waves then become oblique shock waves attached to the tip of the sharp object. For a discussion of experimental techniques for supersonic flow *see* BALLISTIC RANGE; WIND TUNNEL. *See also* SONIC BOOM; SUPERSONIC FLIGHT.

Curved shock waves of the type shown in Fig. 1 can also be found in quasi-steady supersonic flows over blunt objects with flexible skins. A good example is the solar wind flow over the Earth's magnetopause. The stand-off distance of the bow wave from the self-adjusting curved surface of the magnetopause also varies with Mach number in the same way as in terrestrial supersonic flow, provided that the Mach number is suitably redefined as the ratio between solar wind velocity and the highest wave speed for propagating small pressure and momentum flux disturbances in the upstream direction, which may or may not be the ordinary sound speed. *See* MAGNETOSPHERE; SOLAR WIND.

Normal shock wave. The changes in thermodynamic variables and flow velocity across the shock wave are governed by the laws of conservation of mass, momentum, and energy, and also by the equation of state of the fluid. For the case of normal shock, the inset in Fig. 2 illustrates a steady flow across a stationary wavefront (as in the case of a stationary model in a shock tube). The mass flow and momentum equations are the same as for an acoustic wave. However, in a shock wave, changes in pressure p and density ρ across the wavefront can no longer be considered small. As a consequence, the velocity of propagation of the shock wave relative to the undisturbed fluid is givenby Eq. (1), where the initial state

$$u_1^2 = \frac{\rho_2(p_2 - p_1)}{\rho_1(\rho_2 - \rho_1)} \tag{1}$$

of the fluid is denoted by subscript 1 and variables behind the shock front are denoted by subscript 2. In addition, conservation of thermal and kinetic energy across the shock front requires the validity of Eq. (2),

$$h_1 + {}^1/_2 u_1^2 = h_2 + {}^1/_2 u_2^2 \tag{2}$$

where h is the specific enthalpy (or total heat per unit mass) of the fluid, and u_1 and u_2 are fluid velocities relative to the shock wave. By eliminating u_2 and u_1 with the aid of Eq. (1) and the law of conservation of mass, $\rho_1 u_1 = \rho_2 u_2$, the energy equation becomes Eq. (3). If the thermodynamic properties of the fluid

$$h_2 - h_1 = {}^1/_2 \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right) (p_2 - p_1) \tag{3}$$

are known, specific enthalpy h can be expressed as a function of pressure and density, or of any other pair of thermodynamic variables. Equations (1) and (3),

together with the appropriate equation of state of the fluid, are known as the Rankine-Hugoniot equations for normal shock waves. From this set of equations, all thermodynamic variables behind the shock front (denoted by subscript 2) can be expressed as functions of the propagation velocity of the shock wave and the known initial state of the fluid (denoted by subscript 1). For example, if the fluid is a perfect gas of constant specific heats, enthalpy h can be written as Eq. (4), where C_p is the specific heat at constant

$$h = C_p T = \frac{\gamma}{\gamma - 1}\frac{p}{\rho} = \frac{a^2}{\gamma - 1} \tag{4}$$

pressure, T is the thermodynamic temperature, γ is the ratio of specific heats, and a is the adiabatic speed of sound given by $(\gamma RT)^{1/2}$, where R is the gas constant. For this case, the pressure and density ratios across the shock front are given by Eqs. (5) and (6).

$$\frac{p_2}{p_1} = \frac{2\gamma M_1^2 - (\gamma - 1)}{\gamma + 1} \tag{5}$$

$$\frac{\rho_2}{\rho_1} = \frac{u_1}{u_2} = \frac{(\gamma + 1)M_1^2}{2 + (\gamma - 1)M_1^2} \tag{6}$$

The temperature ratio, deduced from the perfect gas law, is given by Eq. (7). These expressions show that

$$\frac{T_2}{T_1} = \frac{\left[2\gamma M_1^2 - (\gamma - 1)\right]\left[2 + (\gamma - 1)M_1^2\right]}{(\gamma + 1)^2 M_1^2} \tag{7}$$

the ratios of all thermodynamic variables and flow velocities across the shock depend on only one parameter for a given gas, which is the Mach number M_1 of the flow relative to the shock front, where $M_1 = u_1/a_1$, or in other words, M_1 equals the velocity of the shock wave divided by the speed of sound for the gas into which the shock propagates. Because of this, the magnitude of M_1 is often used as a measure of the strength of the shock wave. For comparison with the amplitude of acoustic waves, sound waves correspond to values of $M_1 \cong 1.001$ or less. *See* ENTHALPY; GAS; HEAT CAPACITY; MACH NUMBER.

The results of Eqs. (5), (6), and (7) have been derived for gases of constant specific heats. From molecular and atomic physics, it is well known that, when a gas is heated to high temperatures, vibrational excitation, dissociation, and ionization take place, with accompanying changes in heat capacities of the gas. Therefore, for strong shock waves, the appropriate expression for the specific enthalpy h, and the equation of state, which takes into account these phenomena, must be used in place of Eq. (4) to obtain the shock wave solution from Eqs. (1), (2), and (3). The ratios ρ_2/ρ_1, T_2/T_1, and p_2/p_1 for normal shock waves in air at typical atmospheric pressure $p_1 = 1$ bar (75 cm Hg or 100 kPa) and temperature $T_1 = 288$ K (59°F) are plotted in Figs. 2, **3**, and **4**. The approximate solutions, as given by Eqs. (6) and (7), hold only for the weaker shock waves ($M_1 < 6$), even though the pressure ratio is relatively insensitive to the changes in heat capacities of the gas.

As the upstream Mach number M_1 increases, contributions from vibrational excitation, molecular dissociation, and ionization to the specific heats and enthalpy of the shock-heated gas mixture become important. The resultant effect is to lower the temperature T_2 and to increase the compressibility. The increased compressibility then leads to a higher value of the mass density ρ_2.

The relative insensitivity of the pressure ratio p_2/p_1 to changes in M_1, on the other hand, is due to the fact that at high Mach numbers conservation of momentum across the normal shock depends primarily on a balance between the downstream pressure p_2 and the upstream momentum flux $\rho_1 {u_1}^2$. As M_1 further increases, the other terms, p_1 and $\rho_2 {u_2}^2$, become even less important, whatever the thermal properties of the shock-heated gas.

Because the Rankine-Hugoniot equations do not impose any limit on the value of that is M_1, there remains the question of whether a shock wave can propagate into an undisturbed gas at a speed lower than the speed of sound of this gas, that is, $M_1 < 1$. Although this question cannot be answered by the first law of thermodynamics, an examination of the change in specific entropy, using Eq. (8), may provide the answer. Thus Eq. (8) holds, which, for gases of constant specific heats, becomes Eq. (9), showing

$$\Delta S_{12} = S_2 - S_1 = \int_1^2 \frac{dE + p\,dv}{T} \tag{8}$$

$$\frac{\Delta S_{12}}{R} = \frac{1}{\gamma - 1}\ln\frac{\left[2\gamma M_1^2 - (\gamma - 1)\right]\left[2 + (\gamma - 1)M_1^2\right]^{\gamma}}{(\gamma + 1)^{\gamma + 1} M_1^{2\gamma}} \tag{9}$$

that entropy change ΔS_{12} assumes a negative value when M_1 is noticeably less than unity. This violates the second law of thermodynamics, which states that the entropy accompanying any naturally occurring processes always tends to increase. Therefore, it follows that shock waves always travel at supersonic speeds relative to the fluids into which they propagate. *See* ENTROPY; THERMODYNAMIC PRINCIPLES; THERMODYNAMIC PROCESSES.

The Rankine-Hugoniot equations relating the thermodynamic variables p, ρ, and T on the two sides of a normal shock wave are based on the fundamental laws of mechanics, that is, conservation of mass, momentum, and energy during any adiabatic change of state in an isolated system. Since the laws of mechanics are invariant to the choice of inertial frame of reference, the computed values of ρ_2/ρ_1, T_2/T_1, and p_2/p_1 shown in Figs. 2, 3, and 4 for a stationary shock wave in supersonic flow, are equally applicable to a propagating shock wave generated by a violent explosion in an otherwise quiescent atmosphere.

Oblique shock wave. The changes in flow variables across an oblique shock wave are also governed by the laws of conservation of mass, momentum, and

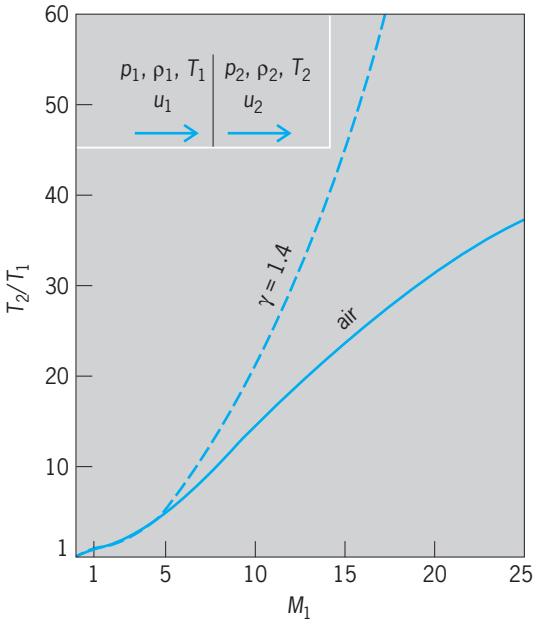


Fig. 3. Temperature ratio T_2/T_1 across normal shock wave in air at standard atmospheric density as a function of Mach number M_1. Inset shows variables in case of steady flow across a stationary wavefront.

energy in a coordinate system which is stationary with respect to the shock front. In this case, the problem is slightly complicated by the fact that the flow velocity will experience a sudden change of direction as well as magnitude in crossing the shock front. Thus, if β_1 and β_2 denote the acute angles between the initial and final flow velocity vectors and the shock surface (**Fig. 5**), then in crossing the oblique shock, the flow will be deflected by a finite amount $\theta = \beta_1 - \beta_2$.

The oblique shock solution can be obtained directly from the complete set of conservation equations. However, the solution already obtained for normal shock waves provides the following simplifying information.

The rate of mass flow per unit area across the shock wave is determined by the normal component of the flow velocity (Fig. 5). Thus, for conservation of mass across the shock, Eq. (10) holds. On

$$\rho_1 u_1 \sin \beta_1 = \rho_2 u_2 \sin \beta_2 \qquad (10)$$

the other hand, conservation of the parallel component of momentum across the shock front requires that Eq. (11) hold. Equations (10) and (11) show that Eq. (12) holds. Itis equivalent to the statement

$$\rho_1 u_1^2 \sin \beta_1 \cos \beta_1 = \rho_2 u_2^2 \sin \beta_2 \cos \beta_2 \qquad (11)$$

$$u_1 \cos \beta_1 = u_2 \cos \beta_2 \qquad (12)$$

that the tangential component of the flow velocity must remain unchanged in crossing the oblique shock wave. Therefore, the resultant flow across the oblique shock shown in Fig. 5 will be identical to what would be seen by an observer who moved at a uniform velocity $u_1 \cos \beta_1$ along the surface of a normal shock wave propagating at a velocity $u_1 \sin \beta_1$. Such a translation of the frame of reference in the direction parallel to the shock front should not change the strength of the shock wave; thus, the changes in thermodynamic variables across the shock should depend only on the velocity component normal to the shock wave. The substitution of M_1 in Eqs. (5), (6), (7), and (9) with $M_1 \sin \beta_1$ gives the corresponding expressions for oblique shock waves in gases of constant specific heats. Again, thermodynamic considerations show that the normal component of the flow velocity into the oblique shock wave must be at least sonic. Therefore, in a supersonic stream of Mach number $M_1 > 1$, the value of β_1 must lie within the range given by expression (13).

$$\sin^{-1} \frac{1}{M_1} \leq \beta_1 \leq \frac{\pi}{2} \qquad (13)$$

The lower limit corresponds to the Mach angle of acoustic waves, while the upper limit corresponds to that of the normal shock wave. At both limits, the flow deflection angle $\theta = \beta_1 - \beta_2$ will be zero. For any intermediate value of β_1, the value of β_2 can

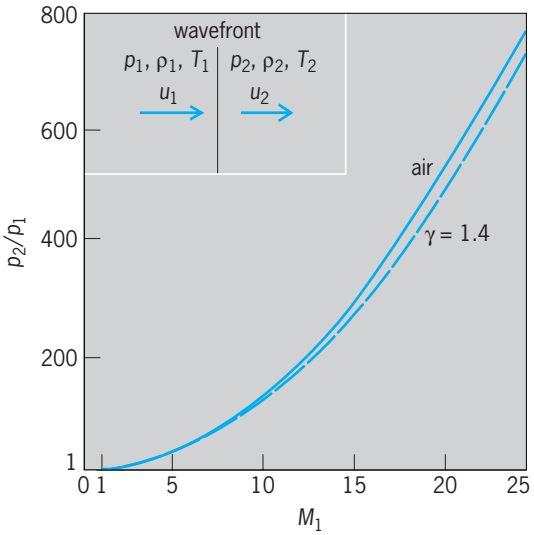


Fig. 4. Pressure ratio p_2/p_1 across normal shock wave in air at standard atmospheric density as a function of Mach number M_1. Inset shows variables in case of steady flow across a stationary wavefront.

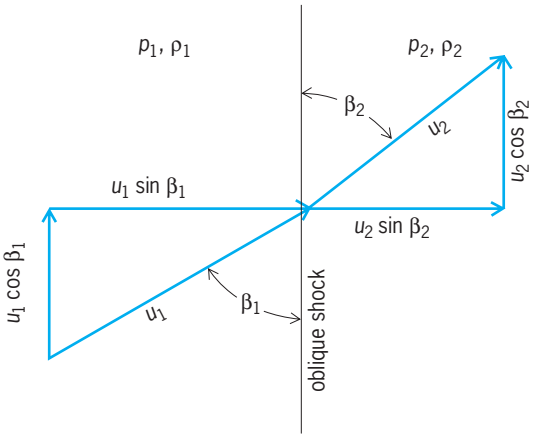


Fig. 5. Flow across an oblique shock wave.

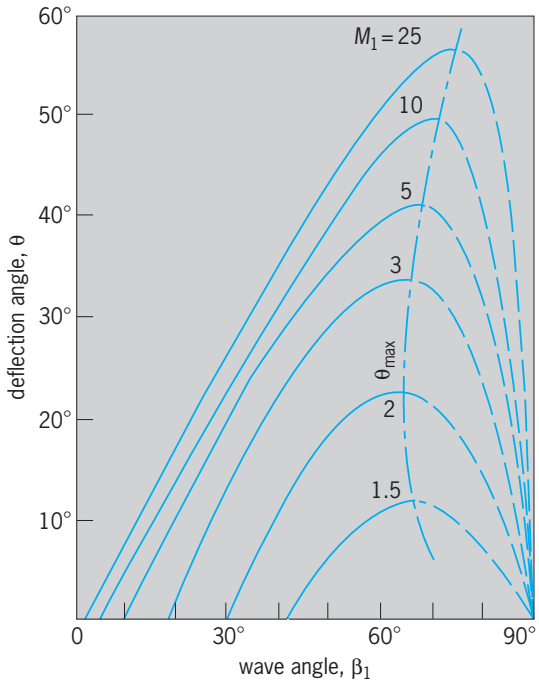


Fig. 6. Flow-detection angle for oblique shock waves in air at standard atmospheric density.

be obtained from Eqs. (10) and (12). Thus Eq. (14)

$$\beta_2 = \tan^{-1}\left(\frac{\rho_1}{\rho_2}\tan\beta_1\right) \qquad (14)$$

holds. The flow deflection angle θ for oblique shock waves in air at normal density is plotted in **Fig. 6** as a function of M_1 and β_1. For a given Mach number M_1, the flow deflection angle first increases with the wave angle β_1, reaches a maximum value θ_{max}, and then decreases toward zero again as the wave angle approaches that of the normal shock. Conversely, for any given flow deflection angle $\theta < \theta_{max}$ such as would be produced by sudden introduction of a wedge or an inclined plane surface into the initially uniform supersonic stream, there exist two possible values of β_1. The higher value of β_1 corresponds to the stronger shock. If the wedge angle or the inclination of the plane surface so introduced exceeds the value of θ_{max} for the given M_1, the shock wave either will become detached from the obstructing object or will form a more complicated pattern. The question of exactly what shock-wave pattern to expect from a given situation is complicated by interaction of the resultant shock wave and flow pattern on the overall boundary condition as well as on the local state of the flow.

Bomb blast. When energy is suddenly released into a fluid in a concentrated form, such as by a chemical or a nuclear explosion, the local temperature and pressure may rise instantly to such high values that the fluid tends to expand at supersonic speed. When this occurs, a blast wave forms and propagates the excess energy from the point of explosion to distant parts of the fluid. If the point of explosion is far from any fluid boundary, the blast wave assumes the form of an expanding spherical shock wave followed by a radially expanding fluid originating from the point of detonation. The changes in thermodynamic variables across the spherical shock are the same as those for a normal shock propagating at the same instantaneous velocity. However, because of the continuous expansion and the finite amount of energy available from the explosion, both the strength of the shock and the specific energy of the expanding fluid must decay with time. The decay of a blast wave goes through three principal stages. A strong shock period begins immediately after the formation of the blast wave, during which the shock strength decays rapidly with distance from the point of detonation. During this stage, the shock velocity decays with the inverse $^3/_2$ power of the distance, and the overpressure behind the shock decays with the inverse cube of the distance. The second stage is a transition period, during which the strong spherical shock gradually changes into an acoustic wave. During the last stage or residual acoustic decay period, the acoustic wave carries the sound of explosion great distances from the point of detonation. As characteristic of sound propagation in three dimensions, the overpressure carried by the spherical acoustic wave decays inversely with distance, and velocity of propagation is constant. *See* EXPLOSIVE.

Extraterrestrial shock waves. The formation and propagation of shock waves are natural processes that cause massive stars to end their lives in supernova explosions and spread their remnants to provide heavy elements that are essential for sustaining life on Earth. While prominent supernovae, such as the one observed by Chinese astronomers in 1054 and Kepler's supernova in 1604, are rare, in the sense that only roughly 25 (most of them obscured by dust) are believed to occur in the Milky Way Galaxy each 1000 years, this slow pace still adds up to tens of millions of supernova explosions over the cosmic time scale of billions of years before the birth of the solar system. *See* SUPERNOVA.

A supernova explosion occurs after a massive star has converted most on its available fusion energy into radiation and its fusion ash, iron, has accumulated at its core to a mass significantly greater than the solar mass, and to a size comparable to that of our Sun. As the fusion energy conversion rate slows and the iron core continues to cool and shrink, the gravitational pressure at its center will eventually overwhelm the degeneracy pressure of the atomic electrons, allowing them to be suddenly captured by the nuclear protons to form neutrons. In the process, neutrinos are released to carry away the excess energy, the specific volume of iron is suddenly reduced to that of nuclear matter, and the iron core implodes gravitationally into a proto-neutron star (p-n star). *See* GRAVITATIONAL COLLAPSE; NEUTRON STAR.

Before the p-n star stabilizes into a neutron star at the end of the core-collapse process (assuming that the star was not massive enough to collapse into a black hole), it is overcompressed momentarily by the momentum flux of the imploding stream and recoils. At that moment, a shock wave begins to form at the

p-n star interface to bounce the incoming stream outward. The formation of this shock wave and its propagation through the dense stellar layers just above the p-n star are affected by p-n star reverberation, gravitational deceleration, energy transfer from neutrinos to the dense stellar plasma behind the shock front, heavy-element production, and so forth, and hence are quite complex. The shock wave is strong enough to superheat the stellar plasma to temperatures far beyond the local precollapse values (about 10^{11} K near the core and 10^4 K in the photosphere).

As the shock wave propagates out of the stellar envelope and into the near vacuum of interstellar space, the superheated stellar plasma explodes into an extremely bright supernova. During the explosion, the stellar plasma expands rapidly in the manner of an unrestrained expansion wave. The shock wave ahead of it continues to propagate into the low-density interstellar gas. As the supernova plasma cools, it eventually condenses into a luminous cloud that becomes the remnant of the supernova. The Crab Nebula is believed to be such a remnant from the 1054 supernova. *See* CRAB NEBULA.

The supernova shock wave is generally invisible. However, it can be detected when it propagates into a gas cloud of substantial density. A good example is Supernova 1987A, whose shock wave was invisible but was detected 11 years later in 1998 by the *Hubble Space Telescope* and the *Chandra X-ray Observatory*. *See* CHANDRA X-RAY OBSERVATORY; HUBBLE SPACE TELESCOPE.

Another example of an extraterrestrial shock wave is the solar wind shock wave just outside the Earth's magnetopause, where the solar wind momentum flux is just balanced by magnetic field pressure so that there is no net flow across the equipotential surface. Space probes launched by the National Aeronautics and Space Administration (NASA) beginning in 1958 suggested the existence of such a shock wave. Plasma probe measurements onboard the *Imp 1* (*Explorer 18*) spacecraft in 1963–1964 confirmed the existence of a quasisteady solar wind of relatively high velocity (300–800 km/s or 200–500 mi/s) and low density (less than 10 ions/cm^3) outside the magnetopause transition region. Magnetometer signals also indicated the presence of plasma turbulence and magnetohydrodynamic wave structure inside the transition region.

The mean profiles of the outer and inner boundaries of the magnetopause transition region mapped out by the first 30 orbits of the *Imp 1* spacecraft seemed to resemble the typical bow shock wave shown in Fig. 1 for steady supersonic flows over blunt objects. In 2005, computer simulations using real-time data from NASA's *ACE* (*Advanced Composition Explorer*) solar wind "weather report" showed that both the shock wave and magnetopause surfaces actually change shape and position rapidly in response to the rapidly changing upstream solar wind conditions during periods of strong solar activities.

Structure, thickness, and stability. Shock wave structure and shock wave stability are topics of considerable interest in plasma physics and astrophysics, as well as in high-temperature gas dynamics. In certain engineering problems, such as the feasibility and efficiency of laser-driven fusion, the question of shock wave stability is crucial.

In gas dynamics, shock wave structure usually refers to spatial distributions of thermodynamic variables such as ρ, P, and T along the flow direction across a normal shock wave. In diatomic and polyatomic gas mixtures, it may include spatial variations of chemical composition and molecular energy level population. In plasma physics and astrophysics, it may also include spatial variations of magnetic field pressure, plasma wave spectra, energy transfer rate, and so forth. From such distributions, one may define an overall shock wave thickness as the spatial extent of the transition zone through which the working fluid changes from the initial equilibrium state 1 to the final equilibrium state 2. However, the term "shock wave thickness" or "shock front thickness" is commonly used to refer only to the length scale over which the sudden jump in gas pressure or density occurs across a normal shock wave. *See* GAS DYNAMICS.

Kinetic theory and low-density shock tube experiments show that the shock front thickness for sudden pressure or density change is only about the length l_1 of one mean free path for molecular collisions at the upstream gas density and temperature over a very wide range of values of shock strength and initial chemical composition. Thus, for a normal shock in air at normal sea level atmospheric pressure and temperature, l_1 is about 0.06 μm, which is the optical wavelength in the deep ultraviolet. This explains why the shock front thickness is difficult to measure. On the other hand, in the upper stratosphere at 50 km (30 mi) altitude, where $p_1 = 0.5$ torr (67 Pa) and $T_1 = 355$ K (179°F), the molecular mean free path l_1 is about 0.1 mm and so would be the thickness of the shock front.

The reason for such a thin shock front is that diatomic gas molecules need only one close (but not head-on) collision to exchange translational momentum and rotational energy. In contrast, only head-on or nearly head-on collisions at sufficiently high kinetic energies are effective for vibrational excitation, dissociation, and ionization, so completion of these processes takes place at much greater distances behind the shock front.

Shock wave stability refers to the ability of the shock wave to maintain its geometrical configuration in the presence of extraneous disturbances, or its ability to recover its original configuration after such disturbances. In telecommunications, knowledge about solar wind shock wave stability would be valuable as it pertains to the recovery time of the Earth's magnetosphere after a severe magnetic storm.

In laser-driven fusion, a form of inertial confinement nuclear fusion which has been proposed as an alternative to magnetic confinement fusion for solving the world's long-term energy supply problem (for a million years or more according to some estimates based on the world's current total energy consumption rate and the total amount of deuterium

and tritium in ocean water), the ability of a converging shock wave to maintain its spherical symmetry in a compressed deuterium–tritium (D–T) mixture would be crucial in the efficient conversion of the fusion reaction energy into electricity. In this case, the shock wave, driven by the reactive force (as in rocket propulsion) of a rapidly heated and outwardly expanding D–T vapor plasma generated by intense focused multiple laser beams, can be expected to behave like an ordinary shock wave in condensed matter. *See* NUCLEAR FUSION.

In early shock tube experiments, converging cylindrical shock waves appeared to be relatively stable up to certain shock strengths, but beyond that, the stability of the shock wave appeared to be at best neutral; that is, initial spatial nonuniformity was maintained but not smoothed out as the converging process continued. This, perhaps, was not too surprising since momentum transfer along the transverse direction of a curved ordinary shock wave cannot be expected to extend much beyond one shock front thickness, or one mean free path for molecular collisions in gaseous media. In plasmas and in condensed matter, the length scale for effective momentum transfer can be quite different. Nevertheless, the early two-dimensional cylindrical shock wave experiments already pointed to the need to maintain an extremely spherically symmetric distribution of focused laser beam intensity in laser-driven fusion in order to achieve efficient energy conversion.

Shao-chi Lin

Condensed Materials

Some common examples of shock waves in condensed materials are encountered in the study of underground or underwater explosions, meteorite impacts, and ballistics problems. The field of shock waves in condensed materials (solids and liquids) has grown into an important interdisciplinary area of research involving condensed matter physics, geophysics, materials science, applied mechanics, and chemical physics. The nonlinear aspect of shock waves is an important area of applied mathematics.

The discussion below will be restricted to one-dimensional plane-wave propagation, in which material properties vary with respect to only one spatial coordinate and time. This situation is the simplest to treat theoretically and is the most frequently studied experimentally. Problems involving one-dimensional cylindrical- or spherical-wave propagation have certain advantages for the study of solids and for many applications, but no satisfactory measurement techniques have been developed to investigate them.

Formation and behavior. The production of one-dimensional plane waves requires rapid, simultaneous deformation of a large flat surface. As the shock wave propagates through the sample thickness, it compresses each macroscopic (or continuum) volume element along the wave propagation direction. Due to inertial confinement, the volume element cannot be deformed laterally, and a state of one-dimensional or uniaxial strain results in the medium. The volume elements at the lateral edge of a finite sample are unconfined, and their expansion results in a rarefaction (or decompression) wave that destroys the uniaxial strain condition and also lowers the pressure in the sample. Experimental measurements are typically made in the central region of the sample before the arrival of the edge rarefactions. Hence, the macroscopic condition of uniaxial strain is satisfied exactly during this time, and the shock wave appears to be extended infinitely in the lateral direction for times of interest.

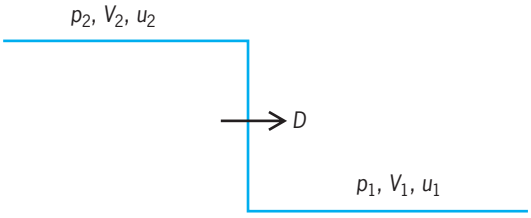


Fig. 7. Diagram of discontinuous shock front moving with a velocity *D* and connecting two uniform states. Here *p* is pressure or equilibrium stress, *V* is specific volume, *u* is particle or mass velocity, and subscripts 1 and 2 represent the initial and final states.

The formation of shock waves is a consequence of the nonlinear response of condensed materials and can be qualitatively understood by examining the effect of a piston moving into a compressible material. The piston motion and the resulting compression can be considered as a series of smaller steps. The propagation velocity for each step is given by $c + u$, where c represents the acoustic velocity corresponding to the compression for that step, and u is the particle velocity at the end of the previous step. For most materials, the propagation velocity increases with each succeeding step, and the compressive wave motion steepens with propagation distance. Eventually, a nearly discontinuous shock front propagating at a supersonic velocity is formed (**Fig. 7**). Although true mathematical discontinuities do not occur in real materials, the idealization depicted in Fig. 7 is a reasonable approximation because of the very short rise times observed in experiments. The above explanation can be reversed to show that a rarefaction or unloading wave will spread out with propagation distance. *See* NONLINEAR ACOUSTICS.

Experimentally, shock waves are produced by rapidly imparting momentum over a large flat surface. This can be accomplished in many different ways: rapid deposition of radiation using electron or photon beams (lasers or x-rays), detonation of a high explosive in contact with the material, or high-speed impact of a plate on the sample surface. The impacting plate itself can be accelerated by using explosives, electrical discharge, underground nuclear explosions, and compressed gases. The use of compressed gas to accelerate projectiles with appropriate flyer plates provides the highest precision and control as well as convenience in laboratory experiments.

By using gas guns, stresses (or pressures) ranging from 50 megapascals (0.5 kilobar) to 1 terapascal (10 megabars) have been achieved. By using lasers and underground nuclear explosions, pressures of the order of 10 TPa (100 Mbar) have been reported (the pressure at the center of the Earth is 350 gigapascals or 3.5 Mbar). Time durations range from a nanosecond to several microseconds; the upper limit is determined by sample size. Because of the short time durations, the shock process can be considered as adiabatic. The irreversible nature of shock compression can result in a large temperature rise in the sample. At the very highest pressures, temperature changes are of the order of 10^5 K, and the electronic contribution to pressure becomes significant. *See* ADIABATIC PROCESS.

Two other important characteristics of shock compression are the fast rate of loading (possibly subnanosecond) and the production of large shear stresses in solids. The rapid loading rates make it possible to examine the kinetics of atomic and molecular rearrangement due to large compressions. Although the term pressure is commonly used in shock-wave studies, the use of stress is more appropriate because the state of stress is nonhydrostatic in general. The generation of large shear stresses makes possible the investigation of various deformation and yield mechanisms in solids that could not be studied otherwise. *See* STRESS AND STRAIN.

Large-amplitude one-dimensional compression and shear waves have been studied in solids. In these experiments, a macroscopic volume element is subjected to both a compression and shear deformation. The combined deformation state is produced by impacting two parallel flyer plates that are inclined at an angle to the direction of the plate motion. Momentum conservation coupled with different wave velocities for compression and shear waves leads to a separation of these waves in the sample interior. An example of a shear wave is shown in **Fig. 8**. These experiments provide direct information about the shear response of shocked solids, and subject samples to more general loading states than the uniaxial strain state.

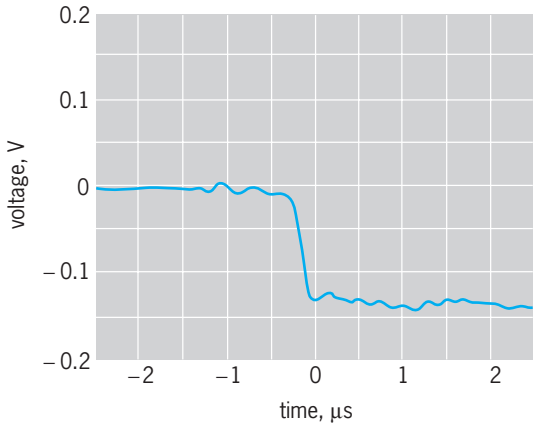


Fig. 8. Typical large-amplitude shear-wave profile. The particle velocity has been converted to a voltage signal. The negative signal is of no consequence. The small perturbation prior to the shear wave is caused by the large compression wave and imperfections in the experimental apparatus.

Theory. Equations (15)–(17) govern continuous

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} = 0 \tag{15}$$

$$\rho \frac{\partial u}{\partial t} = \frac{\partial \sigma}{\partial x} = 0 \tag{16}$$

$$\frac{dE}{dt} + \sigma \frac{\partial V}{\partial t} = 0 \tag{17}$$

one-dimensional planar compression in the absence of heat conduction. Here ρ is density, u is particle or mass velocity; σ is the compressive stress in the x direction and includes viscous and other time-dependent contributions; E is the specific internal energy; and $V (= 1/\rho)$ is the specific volume. Equations (15)–(17), written in the spatial (or eulerian) representation, are valid for both solids and fluids. In analyzing experimental data, it is often more convenient to write these equations in the material (or lagrangian) representation. In the more general case of combined compression and shear waves, the governing equations are generalized to include shear deformation.

For discontinuous jumps or steady shocks, there is no change in the wave shape, and Eqs. (15)–(17) can be integrated to give the Rankine-Hugoniot jump conditions, Eqs. (18)–(20), where D is the shock

$$\rho_1(D - u_1) = \rho_2(D - u_2) \tag{18}$$

$$\sigma_2 - \sigma_1 = \rho_1(D - u_1)(u_2 - u_1) \tag{19}$$

$$E_2 - E_1 = {}^1/_2(\sigma_2 + \sigma_1)(V_1 - V_2) \tag{20}$$

velocity. Equations (18)–(20) relate the initial (subscript 1) and final (subscript 2) states, assumed to be equilibrium states, on either side of the shock front. Hence, σ appearing in the jump conditions represents equilibrium stress (or pressure for fluids) unlike the σ in Eqs. (15)–(17). Within the shock front, deviations from equilibrium can be appreciable.

When the jump conditions are applicable, the locus of end states achieved through a shock transition is referred to as the Rankine-Hugoniot (R–H) curve (**Fig. 9**) or merely the Hugoniot. Unlike curves commonly used in quasistatic experiments, the Rankine-Hugoniot curve in Fig. 9 does not represent the pressure-volume path followed in a given experiment. The actual states that a material element follows in a shock transition lie along the chord connecting the end states and are referred to as the Rayleigh line. The slope of this chord is a measure of the shock velocity and increases with the shock amplitude.

The discussion of thermodynamics of shock compression will be restricted to fluids to permit simplification. The third jump condition, Eq. (20), refers only to thermodynamic quantities and can therefore be combined with the equation of state for a material to generate the Rankine-Hugoniot curve in the

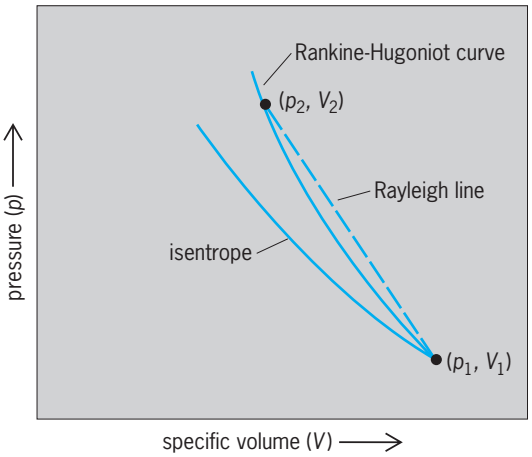


Fig. 9. Rankine-Hugoniot curve and the Rayleigh line for an initial state with pressure and volume (p_1, V_1). Different initial states will lead to different Rankine-Hugoniot curves. An isentrope through the initial state is also shown. For a solid, pressure should be replaced by equilibrium stress.

pressure-volume plane. This procedure, though simple in principle, is difficult to implement in practice because the equation of state is often not known over the pressure-temperature range of the shock experiments. Hence, it is the converse problem that is most useful: how to obtain information about the equation of state from shock-wave data. Measurement of any two quantities appearing in the jump conditions permits a complete solution to Eqs. (18)–(20). Typically, the shock velocity (D) and either pressure or particle velocity are measured in experiments. These data are then used to calculate the values of pressure, specific volume, and specific internal energy along the Hugoniot curve. Calculation of temperature along the Hugoniot and thermodynamic quantities away from the Hugoniot curve require additional assumptions. For solids, there is the additional complexity of correcting for material strength in the stress measurements.

The irreversible nature of the shock transition gives rise to entropy production, which for weak shocks (approximately 15% compression) can be written by expanding the entropy S in a Taylor's series, Eq. (21), where p is the pressure. This equation

$$S_2 - S_1 = \frac{1}{12T_1}\left(\frac{\partial^2 p}{\partial V^2}\right)_{S_1,V_1}(V_1 - V_2)^3 + \cdots \quad (21)$$

shows that the entropy increases as the third order in compression and is positive when $(\partial^2 p/\partial V^2)_S$ is positive. The pressure-volume relation for most normal materials is indeed concave upward as shown in Fig. 9. In addition, the isentrope and the Hugoniot curve have a second-order contact (a point of contact at which the second derivatives are different) at the initial state because of the result in Eq. (21). The unloading curve from a shocked state is along an isentrope and is to the right of the Hugoniot. If the sample is unloaded to zero pressure in a shock-wave compression-release experiment, the final volume will be higher than the initial volume because of thermal expansion due to a higher temperature. *See* ISENTROPIC PROCESS.

The thickness of the shock front is also closely related to the question of thermal equilibrium in these experiments. The shock front thickness is a consequence of the balance between nonlinear compressibility and dissipative mechanisms such as viscosity, heat conduction, and plastic deformation in solids. Because all real materials have some dissipation, the hypothesis of a shock discontinuity in Fig. 7 is an idealization. *See* CONDUCTION (HEAT); PLASTIC DEFORMATION OF METAL; VISCOSITY.

The Hugoniot curve shown in Fig. 9 is applicable for most so-called normal materials. However, deviations from this behavior are of considerable scientific interest and result in shock instability or multiple waves. The general theory of shock stability and its relationship to thermodynamics is complicated and is an area of ongoing research. For nonreactive materials, the stability condition can be written as inequality (22), where c_2 is the sound speed in the

$$D - u_2 < c_2 \quad (22)$$

shocked state. This condition implies that the shock front is subsonic with respect to material behind the shock front, and a disturbance from behind can catch up with the shock front.

Condensed matter properties. Shock waves subject matter to unusual conditions and therefore provide a good test of understanding of fundamental processes. The majority of the studies on condensed materials have concentrated on mechanical and thermodynamic properties. These are obtained from measurements of shock velocity, stress, and particle velocity in well-controlled experiments. Advanced techniques using electromagnetic gages, laser interferometry, piezoelectric gages, and piezoresistance gages have given continuous, time-resolved measurements at different sample thicknesses. These data have permitted improved resolution of the shock front and states behind the front. Time resolution in the 500 picosecond to 1 nanosecond range has been achieved. A large body of high-pressure data has been gathered for most elements and a large number of other materials. These experiments have achieved pressure and temperature ranges well beyond those of static high-pressure experiments, and have provided data for comparison with theoretical predictions based on quantum-mechanical calculations of the equation of state.

Two particular areas of interest in studying mechanical and thermodynamic response are dynamic yielding (plastic deformation or yielding under dynamic loading) and polymorphic phase transitions. Both of these responses lead to an unstable shock, and each can result in a two-wave structure. If both phenomena are present in the experimental range, as for iron, a three-wave structure is observed (**Fig. 10**). The first wave travels with an elastic-wave velocity and has an amplitude that is related to the dynamic strength of the material. The second wave travels with the plastic-wave velocity (close to bulk sound

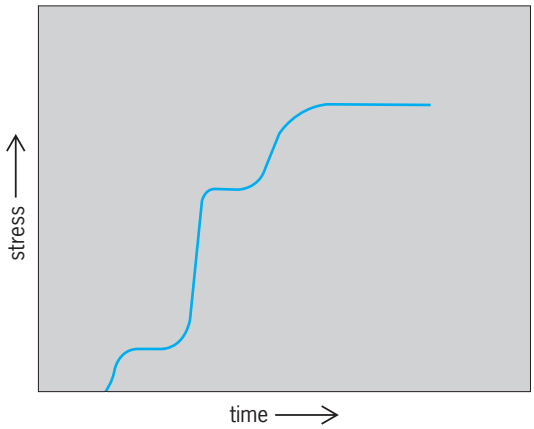


Fig. 10. Representative stress-time profile at a particular sample depth, showing the three-wave structure observed in iron. The first wave is the elastic precursor, the second-wave amplitude (13 GPa or 130 kilobars) marks the onset of the phase change, and the third wave is a measure of the peak stress for the experiment.

speed) and has an amplitude that depends on the peak stress or the phase-transition stress (if the material can undergo a phase change). In the event of a phase change, the amplitude and velocity of the third wave depend on the peak stress. Although a large number of studies have been carried out to investigate shock-induced yielding and phase transitions, the atomic mechanisms controlling these effects are not well understood. Some important unresolved questions concern the applicability of conventional dislocation theory to shock-induced yielding; the mechanisms for shear-strength loss in brittle materials; the precise mechanisms for going from one crystal structure to another in shock-induced phase changes; what controls the fast kinetics of phase changes; the role of material strength in shock-induced phase changes; and the process by which materials melt under shock loading. *See* CRYSTAL DEFECTS; CRYSTAL STRUCTURE; PHASE TRANSITIONS.

Two experimental issues fundamental to thermodynamic and mechanical studies are the complete determination of the stress state under shock loading, and the measurement of temperature (when applicable). Some progress has been made on the first issue, and a complete characterization of the stress state is expected in regions where the stress is nearly constant (the flat portions of the stress-time profile). The second problem is a major challenge for shock-wave experimentalists. Attempts to infer temperature from time-resolved emission from shocked fluids appear promising, but there are two fundamental problems with temperature measurements: the question of thermodynamic equilibrium and an independent check on the inferred temperature. Resolution of the latter problem requires comparison of results from different methods, and temperature measurements have been attempted in a range below that of the emission data. Another approach has involved the propagation of high-pressure ramp waves to approximate isentropic compression in the megabar (10^{11} Pa) range. This development should aid in ascertaining temperature effects and result in a better understanding of thermodynamic states away from the Hugoniot.

Electric, magnetic, and optical properties have not been extensively studied in shock-wave experiments, due to the difficulty in acquiring and interpreting these data. Some work has focused on shock-induced conductivity changes, demagnetization, and piezoelectric and ferroelectric studies, but a detailed understanding of these phenomena does not exist.

Technological developments in lasers and time-resolved optical diagnostics have stimulated considerable interest in spectroscopic measurements in condensed materials. The most extensive work has involved electronic spectroscopy of shocked carbon disulfide (CS_2); a typical absorption spectrum from a multiple shock experiment is shown in **Fig. 11**. Light passes through a thin layer of liquid carbon disulfide (1–300 micrometers) contained between two sapphire plates, and the intensity as a function of wavelength and time is recorded. The latter variable correlates directly with increasing pressure and temperature in the sample. The steps seen in the spectrum correspond to shock reverberation steps. The shift in absorption edge (the bottom part of the band) has been correlated with pressures and temperatures calculated from the carbon disulfide equation of state.

A clearer picture of the molecular processes governing the spectroscopic results will require numerous experiments with systematic variation of parameters and careful analysis. Despite the many challenges in both experiments and analysis, this area of research along with electrical measurements is essential to understanding the shock response of condensed materials on a molecular level.

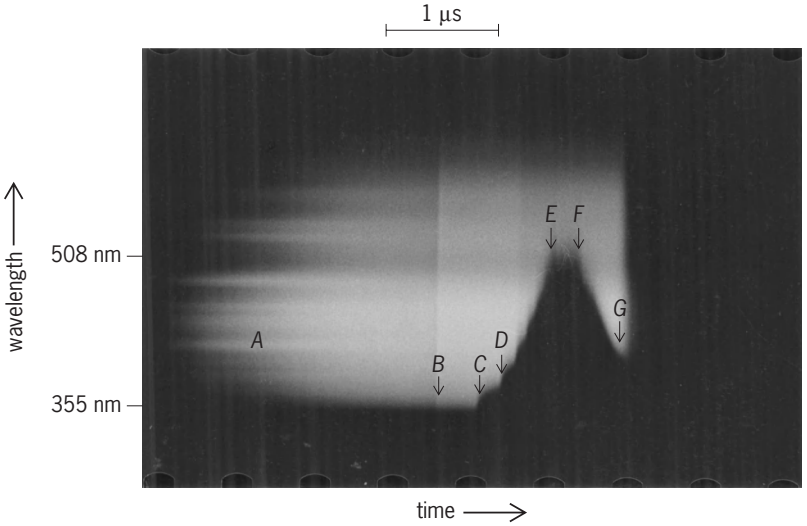


Fig. 11. Time-resolved ultraviolet-visible spectrum of liquid carbon disulfide (CS_2) compressed by shock-reverberation between two sapphire plates. Initial thickness of carbon disulfide = 0.320 mm. Peak pressure = 7.67 GPa = 76.7 kilobars. Symbols indicate events: A = xenon lamp turns on; B = impact; C = shock enters carbon disulfide; D = shock reflects from rear sapphire; E = peak pressure is reached; F = rarefaction wave decreases pressure; G = cell breaks up.

The study of residual effects, that is, the postshock examination of samples subjected to a known pulse amplitude and duration, is of considerable importance to materials science and metallurgy. The conversion of graphite to diamonds is noteworthy. Other effects that have been observed are microstructural changes, enhanced chemical activity, changes in material hardness and strength, and changes in electrical and magnetic properties. The generation of shock-induced lattice defects is thought to be important for explaining these changes in material properties. There has been growing interest in using shock methods for material synthesis and powder compaction. *See* DIAMOND; HIGH-PRESSURE CHEMISTRY; HIGH-PRESSURE MINERAL SYNTHESIS; HIGH-PRESSURE PHYSICS. Y. M. Gupta

Bibliography. J. D. Anderson, *Modern Compressible Flow: With Historical Perspective*, 3d ed., McGraw-Hill, 2003; F. N. Frenkiel and W. R. Sears (eds.), Proceedings of the International Symposium on Magneto-Fluid Dynamics, *Rev. Mod. Phys.*, 32(4):695–1033, 1960; I. I. Glass and J. P. Sislian, *Nonstationary Flows and Shock Waves*, Oxford University Press, 1994; W. D. Hayes and R. F. Probstein, *Hypersonic Flow Theory*, Academic Press, New York, 1959; W. B. Kunkel (ed.), *Plasma Physics in Theory and Application*, McGraw-Hill, 1966; M. A. Lieberman and A. L. Velikovich, *Physics of Shock Waves in Gases and Plasmas*, Springer-Verlag, 1986; M. A. Saad, *Compressible Fluid Flow*, 2d ed., Prentice Hall, 1998; A. Sawaoka (ed.), *Shock Waves in Materials Science*, Springer-Verlag, 1993; J. Smoller, *Shock Waves and Reaction-Diffusion Equations,* 2d ed., Springer-Verlag, 1994; Ya. B. Zeldovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Phenomena*, vols. 1 and 2, translated and edited by W. D. Hayes and R. F. Probstein, Academic Press, 1967.

Short circuit

An abnormal condition (including an arc) of relatively low impedance, whether made accidentally or intentionally, between two points of different potential in an electric network or system. *See* CIRCUIT (ELECTRICITY); ELECTRICAL IMPEDANCE.

Common usage of the term implies an undesirable condition arising from failure of electrical insulation, from natural causes (lightning, wind, and so forth), or from human causes (accidents, intrusion, and so forth). From an analytical viewpoint, however, short circuits represent a severe condition that the circuit designer must consider in designing an electric system that must withstand all possible operating conditions. The short circuit thus is important in dictating circuit design parameters (wire size and so on) as well as protective systems that are intended to isolate the shorted element. *See* ELECTRIC PROTECTIVE DEVICES; ELECTRICAL INSULATION; LIGHTNING AND SURGE PROTECTION.

In circuit theory the short-circuit condition represents a basic condition that is used analytically to derive important information concerning the network behavior and operating capability. Thus, along with the open-circuit voltage, the short-circuit current provides important basic information about the network at a given point. *See* ALTERNATING-CURRENT CIRCUIT THEORY; DIRECT-CURRENT CIRCUIT THEORY; NETWORK THEORY; THÉVENIN'S THEOREM (ELECTRIC NETWORKS).

The short-circuit condition is also used in network theory to describe a general condition of zero voltage. Thus the term short-circuit admittance (or impedance) is used to describe a network condition in which certain terminals have had their voltage reduced to zero, for the purpose of analysis. This leads to the terms short-circuit driving point admittance, short-circuit transfer admittance, and similar references to the zero voltage condition. *See* ADMITTANCE.

Short-circuit protection is a separate discipline dedicated to the study, analysis, application, and design of protective apparatus that are intended to minimize the effect of unintentional short circuits in power supply systems. For these analyses the short circuit is an important limiting (worst) case, and is used to compute the coordination of fuses, circuit reclosers, circuit breakers, and other devices designed to recognize and isolate short circuits. The short circuit is also an important parameter in the specification of these protective devices, which must have adequate capability for interrupting the high short-circuit current. *See* CIRCUIT BREAKER; FUSE (ELECTRICITY).

Short circuits are also important on high-frequency transmission lines where shorted stub lines, one-quarter wavelength long and shorted at the remote end, are used to design matching sections of the transmission lines which also act as tuning elements. *See* TRANSMISSION LINES. Paul M. Anderson

Bibliography. P. M. Anderson, *Analysis of Faulted Power Systems*, 1973; *IEEE Standard Dictionary of Electrical and Electronics Terms*, 5th ed., IEEE Std. 100, 1992.

Short takeoff and landing (STOL)

The term applied to heavier-than-air craft that cannot take off and land vertically but can operate within areas substantially more confined than those normally required by aircraft of the same size. A pure STOL aircraft is a fixed-wing vehicle that derives lift primarily from free-stream airflow over the wing and its high lift system, sometimes with significant augmentation from the propulsion system. Although all vertical takeoff and landing (VTOL) machines, including helicopters, can lift greater loads by developing forward speed on the ground before liftoff, they are still regarded as VTOL (or V/STOL craft), operating in the STOL mode. *See* VERTICAL TAKEOFF AND LANDING (VTOL).

Runway length. It has been customary to define STOL capability in terms of the runway length required to take off or land over a 50-ft (15-m)

obstacle, the concept of "short" length being variously defined as from 500 to 2000 ft (150 to 600 m), depending on the high-lift concept employed and on the mission of the aircraft.

In addition to being able to operate from short runways, STOL aircraft are usually expected to be able to maneuver in confined airspace so as to minimize the required size of the terminal area. Such aircraft must therefore have unusually good slow-flight stability and control characteristics, especially in turbulence and under instrument flight conditions.

Development and uses. The term STOL came into general use during the 1950s, following many efforts during the prior 20 years to develop this capability in aircraft. In these early concepts, STOL was primarily pursued by developing and applying numerous geometric means to improve the high-lift characteristics of the wing. Aircraft of this type were used in off-airport operations in Alaska, in the Canadian north, and in developing countries.

Following these early activities, propulsion system power was more fully explored in a number of schemes to augment wing lift for STOL (**Fig. 1**). Two concepts predominated in the Air Force advanced medium STOL transport competitions. The externally blown flap concept (Fig. 1*e*), used in the YC-15, employs the engine exhaust blowing through, and being turned down by, the flap system to augment the lift. The upper-surface-blowing concept (Fig. 1*f*), used in the YC-14, employs engines mounted above the wing with their exhaust flattened and spread so that it enhances flap effectiveness and is turned downward by the flap. Both of these concepts were flight-tested and demonstrated significant reduction in field length. An augmentor wing aircraft (Fig. 1*d*) was successfully flight-tested, further demonstrating the attributes of power to provide STOL capability.

Subsequently, STOL technology has been applied to explore further improvements in transport takeoff and landing performance as well as potential applications to fighter aircraft. The upper-surface-blowing concept has been employed in a number of flight-test and operational aircraft. This concept has a noise advantage because placing the engine exhaust above the wing shields people on the ground from one of the primary noise sources. The quiet short-haul research aircraft (QSRA) has been used in acoustic, aerodynamic, stability and control, and terminal area investigations. Flight tests of the ASKA (**Fig. 2**) began in Japan in 1987; the Russian Antonov An 74 (derived from the An 72) is operational.

The externally blown flap propulsive-lift concept is used in the U.S. Air Force C-17 transport aircraft. The C-17 has the capability of airlifting large payloads over long distances and delivering them into simple, rough fields. Using technology proven in the YC-15 program, the aircraft has a wing span of 165.0 ft (50.3 m) and a maximum design takeoff gross weight of 572,000 lb (259,457 kg). *See* MILITARY AIRCRAFT.

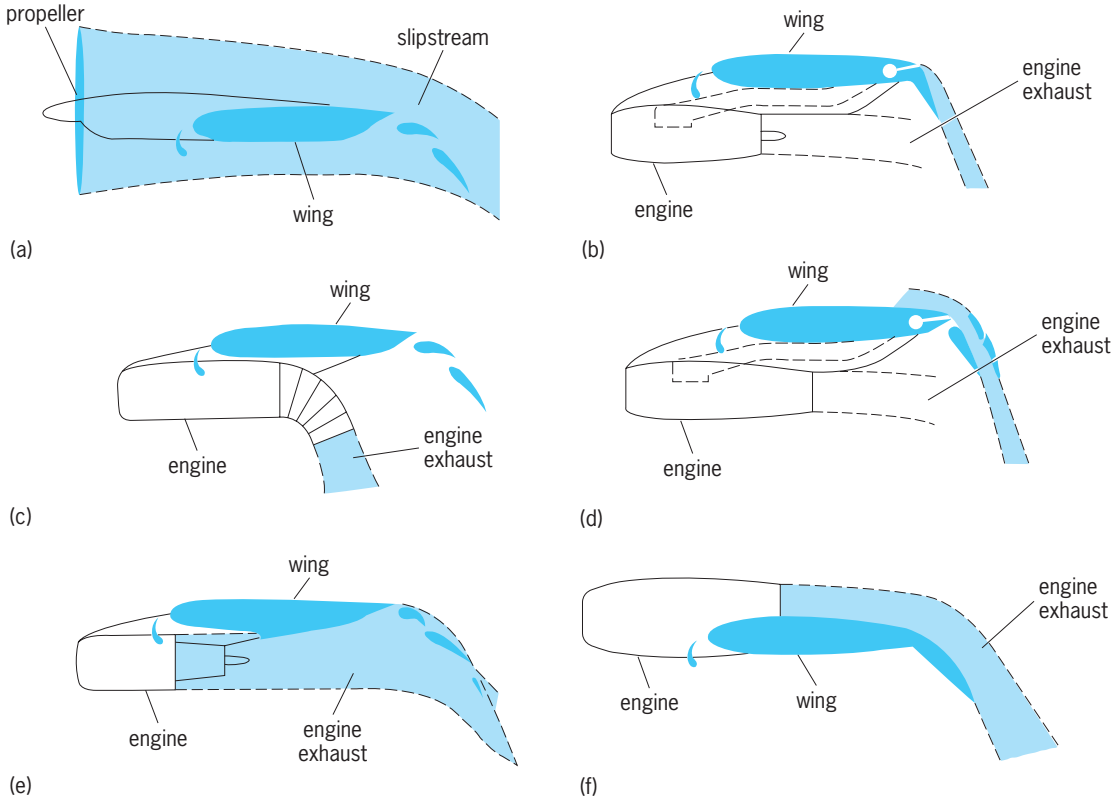


Fig. 1. Propulsive-lift concepts for STOL transports. (*a*) Propeller slipstream. (*b*) Internally blown flap. (*c*) Vectored thrust. (*d*) Augmentor wing. (*e*) Externally blown flap. (*f*) Upper-surface blowing.

Fig. 2. Japanese ASKA upper-surface-blowing transport aircraft.

Under a U.S. Air Force program, an F-15 fighter aircraft was modified to include movable canards, thrust-vectoring/thrust-reversing two-dimensional nozzles, and an advanced, four-channel digital, fly-by-wire flight/propulsion control system. Following 43 flights with standard circular engine nozzles, the aircraft first flew with the two-dimensional nozzles in 1989. The program is to flight-test technologies to provide fighter aircraft with the ability to land on wet, short, bomb-damaged runways. *See* AIRPLANE.

W. P. Nelms

Bibliography. J. A. Cochran et al., *Quiet Short-Haul Research Aircraft: The First Three Years of Flight Research*, Amer. Inst. Aeron. Astronaut. Pap. 81-2625, 1981; D. D. Few, *A Perspective on 15 Years of Proof-of-Concept Aircraft Development and Flight Research at Ames-Moffett by the Rotocraft and Powered-Lift Flight Projects Division, 1970–1985*, NASA Ref. Pub. 1187, August 1987; *Jane's All the World's Aircraft*, published periodically; *Proceedings of the International Powered Lift Conference*, Santa Clara, California, December 7–10, 1987; D. W. Riddle et al., *Quiet Short-Haul Research Aircraft: A Summary of Flight Research since 1981*, Soc. Automotive Eng. Pap. 872315, 1987.

Shrew

A mammal of the family Soricidae (previously Order Insectivora, now listed as Soricomorpha by some). Shrews are found in Asia, Africa, Europe, North America, and northern South America. The family includes 376 species in 26 genera grouped into three subfamilies: the red-toothed shrews (Soricinae), the white-toothed shrews (Crocidurinae), and the African white-toothed shrews (Myosoricini).

Subfamilies. The red-toothed shrews (Soricini) include 14 genera and 148 species and occur throughout much of the major part of the world. The largest genus is *Sorex*, with 77 species. All of the shrews of North America (five genera) are in the Soricini (see **illustration**). Three of the genera (*Sorex*, *Notiosorex*, and *Megasorex*) have long tails, and two (*Blarina* and *Cryptotis*) have short tails that are a little longer that the hind foot. Some of the better known species of red-toothed shrews are the short-tailed shrew (*Blarina brevicauda*), least shrew (*Cryptotis parva*), masked shrew (*Sorex cinereus*), water shrew (*S. palustris*), and trowbridge shrew (*S. trowbridgei*) of the New World, and the European common shrew (*S. araneus*) and European pygmy shrew (*S. minutus*) of the Old World.

White-toothed shrews (Crocidurinae) include 9 genera and 210 species. They are found only in the Old World and are the dominant Old World shrews. The two largest and best known genera are *Crocidura* (172 species) and *Suncus* (18 species).

The African white-toothed shrews (Myosoricine) include 3 genera and 18 species, with the largest genus being Myosorex (14 species).

Morphology and predation. At least two shrews are among the smallest mammals in the world; these are the European pygmy shrew (*Suncus etruscus*) and southern Indiana populations of the American pygmy shrew (*Sorex hoyi*), in which the individuals in some populations weigh about two grams. Another contender for the world's smallest mammal is the bumblebee bat of Asia. The largest shrew is the Asian house shrew, *Suncus murinus*. It weighs about 100 g.

The dental formula for most shrews is I 3/1, C1/1, P3/1, and M3/3 × 2, for a total of 32 teeth, except for *Cryptotis*, which has one less upper premolar (2/1), thus 30 teeth. Shrews have the first upper and lower incisors enlarged, thereby effectively forming a set of tweezers with which they can capture prey and extract them from small places. In addition, the teeth behind the first incisors are small and not well differentiated. There are four or five of these teeth (called unicuspids), which are helpful in shrew classification and identification. The milk teeth are resorbed prior to birth; thus shrews are born with their adult teeth. *See* DENTITION.

Shrews are small secretive animals that constantly run about searching for food. The eyes are minute,

Dusky shrew (*Sorex obscura*), native to western North America. (*Dr. Lloyd Glenn Ingles © California Academy of Sciences*)

the snout is sharp-pointed, and the ears are very small (see illustration). Musk glands on the flank region exude an odor, possibly having a protective function. However, the musk glands do not prevent shrews from being preyed upon by owls, reptiles, and various mammals, including house cats. (When house cats capture the shrews they often do not eat them, presumably because of the bad smell and taste.) The main function of the scent glands in some shrews, at least, is to mark tunnels for reproductive purposes. The glands are much more developed in males than females. Males searching for mates will enter burrows with little odor (that is, those of females) but will avoid those with male odor.

Mating, reproduction, and development. Shrews are solitary and establish distinct territories. Males and females are found together only during mating, which occurs when they are about 1 year old. The female builds a nest, usually in an underground burrow. The nest is lined with leaves and moss. She deposits her litter of about 4–10 young, born after a gestation period of approximately 3 weeks, with usually 3 or 4 litters produced during the season. The young are nearly full size when they leave the nest, mate in their second year, and die during the fall after the reproductive season, or at around 15 months. The fall and winter population is mostly of juveniles and subadults.

Diet and eating habits. Because of their small size and the fact that they lose heat rapidly, these animals have a very high metabolism. They alternate between short (3 h) stretches of feeding and sleeping throughout the 24 h, and eat approximately their own weight in food per day. During winter they do not hibernate, but continue to actively search for food. Shrews are, or were, classed in the order Insectivora, but they do not feed entirely on insects. Although they are often said to feed on whatever is available, in fact they exhibit quite a bit of selectivity. Major foods of some of the larger shrews are earthworms, snails, and centipedes, whereas small ones often feed on small insect larvae. In addition, many shrews feed on hypogenus fungi such as *Endogone*. Some shrews (such as *Blarina* and *Sorex*) have poisonous saliva which is used to overcome prey. Some shrews store food items, such as earthworms and snails. These remain alive but are paralyzed, and thus remain fresh and at the site until eaten.

Novel traits. A few shrews, such as the water shrew, *Sorex palustris*, European water shrew, *Neomys fodiens*, and Tibetian water shrew (*Nectogale elegans*), are aquatic. The latter species is the only shrew with webbed feet. Other aquatic shrews have toes lined with a fringe of hairs. At least some shrews have echolocation, which they use to negotiate their burrow systems, and many also have various high pitched squeaks and twitterings, probably used in communication. Another interesting behavior of at least some shrews is caravanning, in which the young follow the mother by each shrew grasping the one ahead by the base of the tail. Some shrews (*Sorex cinereus*, *Crocidura*) practice coprophagy—or "fecal feeding"—in which partially digested food and fat globules are extruded from the rectum and eaten; it is thought that shrews obtain nutrients from this process that otherwise would not be digested. Shrews are generally solitary and antagonistic to one another, yet the least shrew, *Cryptotis parva*, has been found in groups, with one of the groups numbering 31 individuals.

Another interesting trait of shrews is that they tend to partition their habitat such that any one habitat contains an assemblages of different sized species, which feed on different sized prey and thereby reduce competition. Overall, shrews are a poorly known, but very interesting group of animals deserving of much further study. *See* MAMMALIA.

John O. Whitaker, Jr.

Bibliography. S. Churchfield, *A Natural History of Shrews*, Cornell University Press, Ithaca, 1990; D. Macdonald (ed.), *The Encyclopedia of Mammals*, Facts on File Publications, New York, 1984; R. M. Nowak, *Walker's Mammals of the World* (3 vols.), Johns Hopkins University Press, Baltimore, 1999; J. O. Whitaker, Jr., and W. J. Hamilton, Jr., *Mammals of the Eastern United States*, Cornell University Press, Ithaca, 1998; D. E. Wilson and D. M. Reeder, *Mammal Species of the World. A Taxonomic and Geographic Reference*, 3d ed., in press; D. E. Wilson and S. Ruff, *The Smithsonian Book of North American Mammals*, American Society of Mammalogists, 1999.

Shrink fit

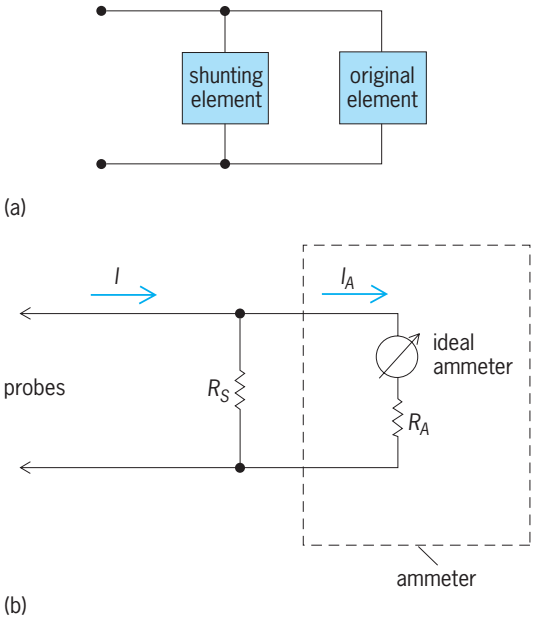
A fit that has considerable negative allowance so that the diameter of a hole is less than the diameter of a shaft that is to pass through the hole, also called a heavy force fit. Shrink fits are used for permanent assembly of steel external members, as on locomotive wheels. The difference between a shrink fit and a force fit is in method of assembly. Locomotive tires, for instance, would be difficult to assemble by force, whereas a shaft and hub assembly would be convenient for force fit by a hydraulic press. In shrink fits, the outer member is heated, or the inner part is cooled, or both, as required. The parts are then assembled and returned to the same temperature. *See* ALLOWANCE; FORCE FIT.

Paul H. Black

Shunting

The act of connecting an electrical element in parallel with (across) another element. The shunting connection is shown in **illus.** *a*.

An example of shunting involves a measuring instrument whose movement coil is designed to carry only a small current for a full-scale deflection of the meter. To protect this coil from an excessive current that would destroy it when measuring currents that exceed its rating, a shunt resistor carries the excess current.



Shunting. (*a*) Shunting connection. (*b*) Ammeter shunted by resistor R_S.

Illustration *b* shows an ammeter (a current-measuring instrument) with internal resistance R_A. It is shunted by a resistor R_S. The current through the movement coil is a fraction of the measured current, and is given by the equation below. With different

$$I_A = \frac{R_S}{R_A + R_S} I$$

choices of R_S, the measuring range for the current I can be changed. *See* AMMETER; CURRENT MEASUREMENT.

Similar connections and calculations are used in a shunt ohmmeter to measure electrical resistance. Shunt capacitors are often used for voltage correction in power transmission lines. A shunt capacitor may be used for the correction of the power factor of a load. In direct current shunt motors, the excitation (field) winding is connected in parallel with the armature. *See* DIRECT-CURRENT MOTOR; OHMMETER; POWER FACTOR; RESISTANCE MEASUREMENT.

In electronic applications, a shunt regulator is used to divert an excessive current around a particular circuit. In broadband electronic amplifiers, several techniques may be used to extend the bandwidth. For high-frequency extension, a shunt compensation is used where, typically, a capacitor is shunted across an appropriate part of the circuit. Shunt capacitors (or more complicated circuits) are often used to stabilize and prevent undesired oscillations in amplifier and feedback circuits. *See* ALTERNATING-CURRENT CIRCUIT THEORY; AMPLIFIER; DIRECT-CURRENT CIRCUIT THEORY; FEEDBACK CIRCUIT.

Shlomo Karni

Bibliography. D. G. Fink and H. W. Beaty (eds.), *Standard Handbook for Electrical Engineers*, 14th ed., 2000; M. Kaufman and A. H. Seidman, *Handbook for Electronics Engineering Technicians*, 2d ed., 1985; J. Markus and C. Weston, *Essential Circuits Reference Guide*, 1988.

Sickle cell disease

An inherited disorder of red blood cells characterized by lifelong anemia and recurrent painful episodes. The sickle cell mutation is caused by a single nucleotide effecting a change in the β-globin gene, resulting in the substitution of valine for glutamic acid as the sixth amino acid of β-globin. The short circulatory survival of red blood cells that contain sickle cell hemoglobin S results in anemia, and their abnormal rigidity contributes to painful obstruction of small blood vessels. *See* ANEMIA; GENETIC CODE; HEMOGLOBIN.

Genetics. The sickle cell gene is found most commonly among individuals of African ancestry, but also has a significant incidence in Mediterranean, Middle Eastern, and Asian Indian populations. Its evolutionary persistence in these populations is related to its ability to partially protect heterozygous individuals from death due to malaria. The geographic dispersion of this gene is related to its having arisen as a spontaneous mutation at least four different times and to its transmission along ancient trade routes.

Inheritance of one sickle gene and one normal β-globin allele results in a simple heterozygous condition known as sickle cell trait. This benign carrier condition is associated with a normal life expectancy, and it does not cause either anemia or recurrent pain. However, carriers may occasionally experience blood in their urine and be more subject to the risk of sudden death when engaging in rigorous prolonged physical conditioning. The large amounts of hemoglobin A within sickle-cell-trait red blood cells protect against the deleterious effects of hemoglobin S. The inheritance of homozygous sickle cell anemia results in sufficiently high intracellular concentration of sickle cell hemoglobin S to cause clinical disease. *See* HUMAN GENETICS.

Pathophysiology. The pathophysiology of sickle cell disease has three aspects: molecular, cellular, and rheological. The property of sickle cell hemoglobin S responsible for clinical disease is its insolubility when deoxygenated. Oxygenated sickle cell hemoglobin S is as soluble as oxygenated normal hemoglobin, but when it is deoxygenated it aggregates and forms an insoluble polymer.

The effects of sickle cell hemoglobin on sickle red blood cells is catastrophic in terms of their circulatory survival and deformability. While normal red blood cells circulate for 4 months, sickle cells survive in the circulation for only about 12–20 days. Polymerization of sickle cell hemoglobin within deoxygenated sickle cells reversibly reduces cellular deformability and distorts cells to the sickle shape (see **illus.**). Sickle cells usually return from the venous circulation to the arterial, where the hemoglobin is reoxygenated and the cells unsickle. The so-called vicious cycle of erythrostasis that has been proposed to account for the episodic vascular

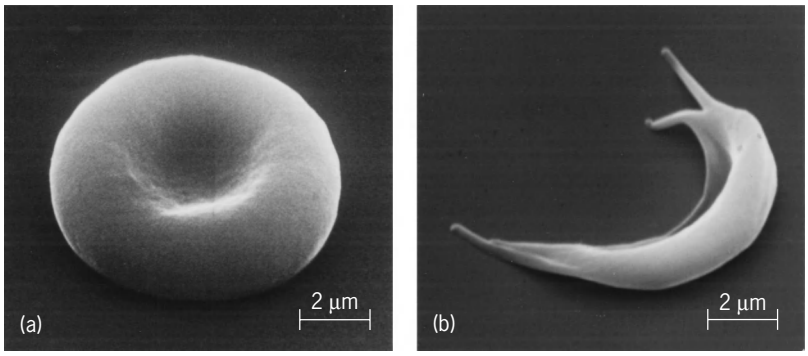
occlusion of sickle cell disease suggests that transient stasis of blood flow results in deoxygenation of sickle cells and the generation of an acidic milieu, both of which make possible the polymerization and sickling responsible for recurrent pain and organ dysfunction. Initiation of vascular occlusion, however, appears to be the result of abnormally sticky young sickle cells adhering to the vascular endothelium.

Persistent cycles of sickling and unsickling result in the generation of dehydrated, very dense sickle cells; these are irreversibly sickled cells that are incapable of resuming a normal shape when reoxygenated. Their generation is the result of abnormal cell losses of cations and water via potassium-chloride cotransport and calcium-sensitive potassium efflux, called the Gardos pathway.

As a result of the poor deformability of individual sickle red blood cells, sickle cell blood has high viscosity. The impaired rheologic properties of sickle blood are compounded by abnormal adherence of sickle red cells to endothelial cells lining the blood vessels.

Clinical manifestations. The short-lived nature of sickle red blood cells results in lifelong chronic hemolytic anemia with which accelerated red blood cell production cannot keep pace. The increased turnover of red blood cells results in elevated levels of hemoglobin degradation and bilirubin production by the liver and in very frequent formation of gallstones. Individuals with sickle cell disease are also afflicted with episodes of increased severity of anemia, usually due to cessation of erythropoiesis. Vasoocclusive complications of sickle cell disease include the episodic painful crises and both chronic and acute organ dysfunction, for example, strokes, painful necrosis of bone, loss of vision, diminished function of the liver and kidneys, blood in the urine, and poorly healing ulcers of the legs. Older individuals frequently have a degree of heart failure that may be related to both anemia and vascular occlusion. Individuals with sickle cell disease are generally regarded as having severely shortened life expectancy, but survival has improved to the extent that average life expectancy is in the fifth decade. One disease manifestation that is particularly problematic in young children is susceptibility to infections. Prior to the use of prophylactic antibiotics in infants and toddlers, septicemia was the leading cause of death in this age group.

Diagnosis. The standard method of diagnosing sickle cell syndromes is hemoglobin electrophoresis. The replacement of negatively charged glutamic acid by neutrally charged valine results in sickle cell hemoglobin S having a different electrophoretic mobility from normal hemoglobin. On electrophoresis the hemoglobin of subjects with homozygous sickle cell anemia is virtually all sickle cell hemoglobin, while that of individuals with sickle cell trait includes both normal and sickle cell hemoglobin. Another method of detecting sickle cell hemoglobin is solubility testing in a high-ionic-strength solution, a test that will be positive in sickle trait as well as in sickle cell disease. A simple diagnostic tool is to review the peripheral blood smear microscopically for the presence of irreversibly sickled cells. These cells are almost always found in individuals with sickle cell disease but are absent in those with sickle trait. *See* ELECTROPHORESIS.



Scanning electron micrographs of (*a*) an oxygenated sickle red blood cell and (*b*) a deoxygenated sickle red blood cell. The appearance of the oxygenated sickle cell is indistinguishable from a normal red blood cell. (*Electron micrographs by Dr. James White*)

There are a variety of diagnostic tests based on deoxyribonucleic acid (DNA). These tests depend on the ability to detect substitution of the single nucleotide responsible for the sickle gene in native genomic DNA or in DNA that has been amplified enzymatically from genomic DNA using the polymerase chain reaction. These DNA-based diagnostic methods are particularly useful for prenatal diagnosis of sickle cell disease, where often only small amounts of DNA for diagnosis are obtained by amniocentesis or chorionic villus sampling. *See* DEOXYRIBONUCLEIC ACID (DNA); PRENATAL DIAGNOSIS.

Therapy. Despite the profound understanding of sickle cell disease, treatment of painful episodes often consists of only symptomatic therapy, including analgesics for pain, antibiotics for infections, and transfusions for episodes of severe anemia. Previous attempts at inhibiting polymerization of sickle cell hemoglobin have been unsuccessful because of intolerable side effects or lack of beneficial effect. Genetic counseling and prenatal diagnosis remain important therapeutic approaches. *See* BLOOD; GENETIC ENGINEERING.

Stephen H. Embury

Bibliography. H. F. Bunn and B. G. Forget, *Hemoglobin: Molecular, Genetic and Clinical Aspects*, 1986; W. A. Eaton and J. Hofrichter, Sickle cell hemoglobin polymerization. *Adv. Prot. Chem.*, 40:63–279, 1990; S. H. Embury et al., *Sickle Cell Anemia: Scientific Principles and Clinical Practice*, 1994; R. P. Hebbel, Beyond hemoglobin polymerization: The red blood cell membrane sickle cell disease pathophysiology, *Blood*, 77:214–237, 1991; G. R. Serjeant, *Sickle Cell Disease*, 3d ed., 2001.

Sideband

The range of the electromagnetic spectrum located either above (the upper sideband) or below (the lower sideband) the frequency of a sinusoidal

carrier signal $c(t)$. The sidebands are produced by modulating the carrier signal in amplitude, frequency, or phase in accordance with a modulating signal $m(t)$ to produce the modulated signal $s(t)$. The resulting distribution of power in the sidebands of the modulated signal depends on the modulating signal and the particular form of modulation employed. *See* AMPLITUDE MODULATION; FREQUENCY MODULATION; PHASE MODULATION.

Amplitude-modulation systems. In conventional analog amplitude modulation (AM), changing the amplitude of a carrier signal of frequency f hertz in accordance with the amplitude of a modulating signal whose power spectrum extends from 0 to B hertz produces an upper sideband extending from f to $f + B$ hertz and a lower sideband extending down from f to $f - B$ hertz. Except for this shift in frequency, the distribution of power in the two sidebands equals that of the modulating signal, with one sideband being the mirror image of the other about the carrier frequency. The total power of the modulated signal is the sum of the power in the carrier signal at f hertz and the power in the modulating signal, the latter being evenly divided over the two sidebands.

Modifications. Several modifications of conventional analog amplitude modulation are used in practice. Since the carrier signal conveys no information about the modulating signal, it can be eliminated. This so-called double sideband–suppressed carrier (DSB-SC) modulation reduces the total amount of transmitted power in the modulated signal. In addition, either sideband contains all the information about the modulating signal, so that it suffices to transmit only one of them. This single-sideband (SSB) modulation therefore reduces the required channel bandwidth by a factor of 2. In a modification of single sideband, known as vestigial sideband (VSB), the transmitted signal consists of one complete sideband, the carrier signal, and a small fraction of the other sideband. The latter can be used to simplify the demodulation process relative to the complexity of single-sideband demodulation. *See* SINGLE SIDEBAND.

Digital AM systems. In practical terms the power spectra of digital amplitude-modulation schemes such as on-off keying (OOK) or amplitude-shift keying (ASK), in which the carrier signal is shifted in amplitude between a discrete and finite number of values, are similar to those for analog amplitude modulation. The upper and lower sidebands are usually taken to extend over the range from f to $f + R$ and from f to $f - R$, respectively, where R is the rate at which the amplitude of the carrier signal is shifted.

Angle-modulation systems. In analog angle-modulation systems the power spectral distribution of the modulated signal generally does not exhibit the same symmetry about the carrier frequency as for amplitude modulation. However, if the modulating signal is periodic, the power in the modulated signal exists at discrete and evenly spaced frequencies, with equal amounts of power at pairs of frequencies symmetrically disposed about the carrier frequency. *See* ANGLE MODULATION.

Optimization of power spectrum. In general, the power spectral distribution in the sidebands is a function of the type of modulation employed and the power spectral distribution of the modulating signal. Some modulation schemes, in particular offset quadrature phase-shift keying and minimum-shift keying, offer a distinct advantage over other forms in that most of the power is concentrated in a relatively narrow portion of the sidebands, resulting in reduced transmission bandwidth requirements. *See* MODULATION.

Hermann J. Helgert

Bibliography. B. P. Lathi, *Modern Digital and Analog Communication Systems*, 3d ed., Oxford University Press, 1998; M. Schwartz, *Information Transmission, Modulation, and Noise*, 4th ed., McGraw-Hill, 1990; D. R. Smith, *Digital Transmission Systems*, 2d ed., Van Nostrand Reinhold, 1993.

Sidereal time

One of several kinds of time scales used in astronomy, whose primary application is as part of the coordinate system to locate objects in the sky. It is also the basis for determining the solar time commonly used in everyday living.

Time measurement. Time is measured with respect to some repetitious event. The measure of time is a parameter whose scale and units are adopted for specific purposes at hand. The most common measurements of time are based on the motions of the Earth that most affect everyday life; rotation on its axis, and revolution in orbit around the Sun. Objects in the sky reflect these motions and appear to move westward, crossing the meridian each day. A particular object or point is chosen as a marker, and the interval between its successive crossings of the local meridian is defined to be a day, divided into 24 equal parts called hours. The actual length of the day for comparison between systems depends on the reference object chosen. The time of day is reckoned by the angular distance around the sky that the reference object has moved westward since it last crossed the meridian. In fact, the angular distance west of the meridian is called the hour angle. *See* MERIDIAN.

Sidereal day and year. The reference point for marking sidereal time is the vernal equinox, one of the two points where the planes of the Earth's Equator and orbit appear to intersect on the celestial sphere. The sidereal day is the interval of time required for the hour angle of the equinox to increase by 360°. One rotation of the Earth with respect to the Sun is a little longer, because the Earth has moved in its orbit as it rotates and hence must turn approximately 361° to complete a solar day. A sidereal year is the time required for the mean longitude of the Sun to increase 360°, or for the Sun to make one circuit around the sky with respect to a fixed reference point. *See* EQUINOX.

Astronomical coordinate system. Sidereal time is essential in astronomy and navigation because it is part of the most common coordinate system for

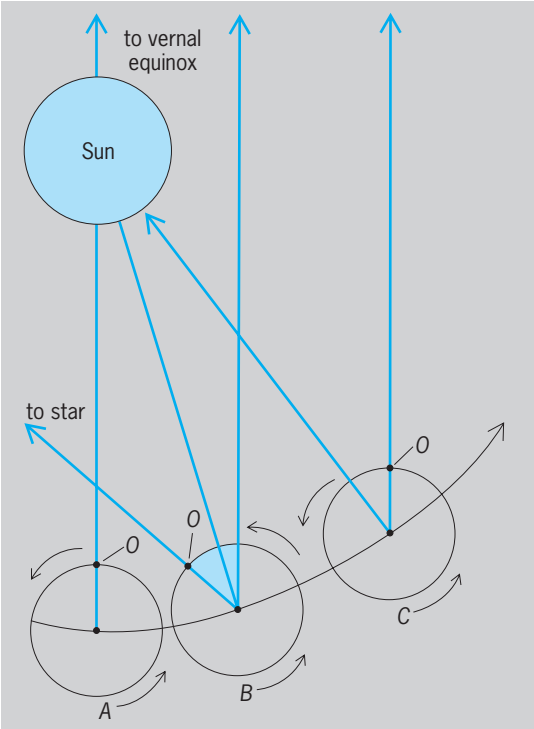


Diagram of the Earth's motions, illustrating the concepts of sidereal time, hour angle, and right ascension. Symbols are explained in text.

describing the position of objects in the sky. These coordinates are called right ascension and declination, analogous to longitude and latitude on the surface of the Earth, with the vernal equinox playing the same role as the Greenwich meridian for the zero point. The local sidereal time is the hour angle of the vernal equinox past (westward of) the local meridian, and also the right ascension of any object on the meridian. By a sidereal clock, a star rises and crosses the local meridian at the same times every day. Greenwich sidereal time is local sidereal time plus the longitude of the observer west of Greenwich. *See* ASTRONOMICAL COORDINATE SYSTEMS.

These concepts are further described in the **illustration**, in which the reader is looking down on the North Pole as the Earth moves counterclockwise in orbit around the Sun. An observer at point *O* is considered. When the Earth is at position *A*, the Sun is taken to be in line with the vernal equinox and overhead for the observer. At position *B*, the Earth has rotated counterclockwise through the shaded angle, which represents the hour angle of the vernal equinox, or the local sidereal time for the observer and also the right ascension of a star overhead for the observer. At position *C*, the Earth has made a complete rotation with respect to the vernal equinox but not with respect to the Sun. The right ascension of the Sun has increased by approximately 1°.

Determination. For a measure of time to be of practical use, determination of it by observation must be possible. Sidereal time, defined by the daily motion of the vernal equinox, is determined in practice from observations of the daily motions of stars. Solar time, which is used in the conduct of everyday life, is not directly observable. It is determined from the sidereal time by a known relationship.

Effect of motion of Earth's axis. Upon more sophisticated examination, the sidereal day is not exactly the period of rotation of the Earth, being shorter by about 0.0084 s. The Earth's axis of rotation moves very slowly but regularly in space because of the gravitational effects of other bodies in the solar system, and the Earth's crust also shifts very slowly and irregularly about the axis of rotation. As a result, neither the equinox nor the local meridian plane is absolutely fixed. Consequently, the relation of sidereal time to the measure of time defined by the rotational motion of the Earth alone is highly complex. *See* EARTH ROTATION AND ORBITAL MOTION; PRECESSION OF EQUINOXES; TIME.

Alan D. Fiala

Bibliography. J. Kaler, *The Ever-Changing Sky*, Cambridge University Press, 1996; D. Morrison, S. Wolff, and A. Franknoi, *Abell's Exploration of the Universe*, 7th ed., Saunders College Publishing, 1995; P. K. Seidelmann (ed.), *Explanatory Supplement to the Astronomical Almanac*, University Science Books, 1992; W. M. Smart and R. M. Green, *Textbook on Spherical Astronomy*, 6th ed., Cambridge University Press, 1977, reprint 1999.

Siderite

A mineral ($FeCO_3$) with the same space group and hexagonal crystal system as calcite ($CaCO_3$). Siderite has a gray, tan, brown, dark brown, or red color, has rhombohedral cleavages, and occasionally may show rhombohedral crystal terminations. It may display curved crystal faces like dolomite ($CaMg[CO_3]_2$), but more commonly is found as massive, compact, or earthy masses. *See* CARBONATE MINERALS; CRYSTAL STRUCTURE; DOLOMITE.

Siderite forms extensive solid solutions with magnesite ($MgCO_3$), rhodochrosite ($MnCO_3$), and less commonly with smithsonite ($ZnCO_3$); with calcite and ankerite ($Ca[Fe,Mg][CO_3]_2$), it has wide miscibility gaps that vary with temperature. Its physical properties are variable, depending on the amounts of magnesium (Mg), manganese (Mn), and zinc (Zn) solid solutions; pure iron(II) carbonate ($FeCO_3$) has a high specific gravity of 3.94, a medium hardness of 3.5–4, and a high index of refraction, 1.88. It is only weakly soluble in dilute hydrochloric acid (HCl) but is very soluble in hot hydrochloric acid and other hot strong acids. It may resemble some sphalerite ([Zn,Fe]S) but yields carbon dioxide (CO_2) and not hydrogen sulfide (H_2S) upon reaction with acids. *See* ANKERITE; CALCITE; MAGNESITE; RHODOCHROSITE; SOLID SOLUTION.

Siderite, a widespread mineral in near-surface sediments and ore deposits, occurs in hydrothermal veins, lead-silver ore deposits, sedimentary concretions formed in limestones and sandstones, and Precambrian banded iron formations that precipitated under acidic conditions. Upon metamorphism it may be oxidized or decarbonated to form the iron

oxide magnetite (Fe_3O_4) at 300–500°C (570–930°F). In the presence of quartz (SiO_2) it reacts to form various ferrous hydroxyl silicates upon metamorphism, including minnesotaite ($Fe_3Si_4O_{10}[OH]_2$), at higher metamorphic grades (250–400°C or 480–750°F), or greenalite ($Fe_3Si_2O_5(OH)_4$) at lower metamorphic temperatures (<250°C or 480°F). The exact metamorphic reaction involving the decomposition of siderite depends on the relative pressures of carbon dioxide, water, and oxygen as well as temperature. Siderite may also alter to various iron oxyhydroxides such as goethite during weathering or low-temperature alteration.

Famous localities for siderite are found in Styria (Austria), Westphalia (Germany), Cornwall (Britain), Wawa (Northern Ontario, Canada), Minas Gerais (Brazil), and Llallagua and Potosi (Bolivia). These and other occurrences have provided locally significant quantities of siderite as iron ore. *See* IRON; MAGNETITE; METAMORPHISM; ORE AND MINERAL DEPOSITS.

Eric J. Essene

Bibliography. L. L. Y. Chang, R. A. Howie, J. Zussman, *Rock-Forming Minerals,* vol. 5B: *Non-Silicates: Sulphates, Carbonates, Phosphates, Halides*, 1995; C. Klein and C. S. Hurlbut, Jr., *Manual of Mineralogy*, 21st ed., 1993, rev. 1999; R. J. Reeder (ed.), *Carbonates: Mineralogy and Chemistry*, Rev. Mineral. 11, 1983.

Siderophores

Low-molecular-mass (generally 500–1000 daltons) molecules that have a high specificity for chelating or binding iron. Siderophores are produced by many microorganisms, including bacteria, yeast, and fungi, to obtain iron from the environment. More than 500 different siderophores have been identified from microorganisms. Some bacteria produce more than one type of siderophore. *See* BACTERIA; BIOINORGANIC CHEMISTRY; CHELATION; FUNGI; IRON; YEAST.

Iron is required by aerobic bacteria and other living organisms for a variety of biochemical reactions in the cell, including production of adenosine triphosphate (ATP) for energy and for synthesis of deoxyribonucleic acid (DNA). Although iron is the fourth most abundant element in the Earth's crust, it is not readily available to bacteria. Iron is found in nature mostly as insoluble precipitates that are part of hydroxide polymers. Bacteria living in the soil or water must have a mechanism to solubilize iron from these precipitates in order to assimilate iron from the environment. Iron is also not freely available in humans and other mammals. Most iron is found intracellularly in heme proteins and ferritin, an iron storage compound. Iron outside cells is tightly bound to proteins such as transferrin in serum and lactoferrin on mucosal surfaces. Therefore, bacteria that grow in humans or other animals and cause infections must have a mechanism to remove iron from these proteins and use it for their own energy and growth needs. Siderophores have a very high affinity for iron and are able to solubilize and transport ferric iron (Fe^{3+}) in the environment and also to compete for iron with mammalian proteins such as transferrin and lactoferrin. The majority of bacteria and fungi use siderophores to solubilize and transport iron. Microorganisms can use either siderophores produced by themselves or siderophores produced by other microorganisms. *See* ADENOSINE TRIPHOSPHATE (ATP); DEOXYRIBONUCLEIC ACID (DNA); IRON METABOLISM.

Structure. The many different types of siderophores can generally be classified into two structural groups, hydroxamates and catecholate compounds. Despite their structural differences, all form an octahedral complex with six binding coordinates for Fe^{3+} (**Fig. 1**). Siderophores can be detected in bacterial cultures using chemical assays, bioassays, and functional assays. There are specific chemical assays for hydroxamate and catecholate groups. However, these assays do not work for all siderophores because some siderophores do not have the appropriate functional group. Bioassays use selected strains of bacteria that require addition of siderophores for growth in a low-iron medium. Cultures or extracellular fluids of siderophore-producing bacteria can often promote growth of these strains. The bioassay strains, however, do not detect all siderophores. The chrome azurol S assay is a functional assay that is often used to detect siderophores. Chrome azurol S is a blue compound that binds iron but changes to an orange color when iron is removed from it by siderophores. The chrome azurol S assay is independent of siderophore structure and has made it relatively easy to detect siderophores in many microorganisms. *See* BIOASSAY; CULTURE.

Uptake by bacteria. Gram-negative bacteria have a cell envelope consisting of an outer membrane and an inner or cytoplasmic membrane; these membranes are separated by the periplasm. The outer membrane contains proteins called porins that form

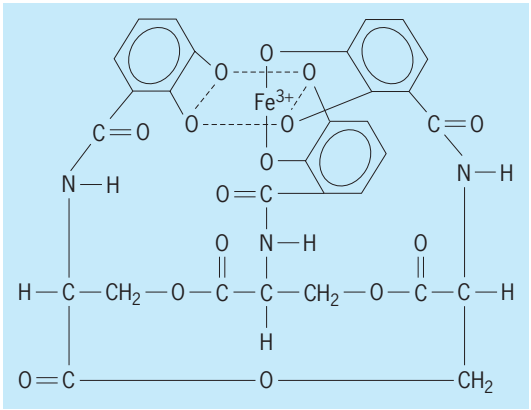


Fig. 1. Structure of ferric enterobactin, a catechol-type siderophore produced by many enteric bacteria. Three molecules of enterobactin bind each molecule of Fe^{3+} to provide the six binding coordinates. (*Adapted from J. Biol. Chem., 270:26723–26726, 1995*)

water-filled channels. Small substrates with masses of up to 600 daltons, including sugars, phosphates, and nucleosides, can diffuse through these channels in the membrane. However, iron-siderophore complexes are too large to diffuse through these channels. Therefore, gram-negative bacteria produce specific proteins in the outer membrane that function as receptors for iron-siderophore complexes. These receptor proteins have a very high affinity for specific siderophores and concentrate the iron-siderophore complex at the cell surface. These complexes are then translocated into the periplasm of the cell by an active transport mechanism. *See* BACTERIAL PHYSIOLOGY AND METABOLISM; CELL (BIOLOGY); CELL MEMBRANES; CELL PERMEABILITY; PROTEIN.

Siderophore receptors, proteins ranging in mass from 70,000 to 90,000 daltons, are produced by bacteria when iron is not readily available. Bacteria may produce receptors for several different siderophores, including those produced by themselves and those produced by other bacteria that might be present in their environment. It is not unusual for bacteria to produce six or more different siderophore receptors. These receptors enable them to compete with other microorganisms for iron and provide a competitive advantage that allows these bacteria to grow in nature or in a mammalian host. Bacteria even have receptors for siderophores produced by fungi. For example, *Escherichia coli, Pseudomonas aeruginosa*, and *Yersinia enterocolitica* are bacteria that cause diseases in humans. These bacteria all have receptors for ferrichrome siderophores produced by fungi. *See* ESCHERICHIA; PSEUDOMONAS; YERSINIA.

In addition to specific siderophore receptors, other proteins are involved in translocating iron from the surface of the bacteria into the cytoplasm, where it is needed for various biochemical functions (**Fig. 2**). Proteins known as TonB, ExbB, and ExbD form a complex in the inner membrane. TonB extends from the inner membrane through the periplasm and along with ExbB and ExbD provides energy to the siderophore receptor, located in the outer membrane, and facilitates transport of the iron-siderophore complex into the periplasm. Periplasmic iron-binding proteins carry the iron through the periplasm to the inner membrane. ATP-binding proteins and permeases then transport the iron into the cytoplasm. In the cytoplasm, ferric iron, Fe^{3+}, is reduced to ferrous iron, Fe^{2+}, for use by the cell.

Although gram-positive bacteria do not have an outer membrane or specific siderophore receptor proteins, they do have proteins that are similar to the periplasm iron-binding proteins that directly bind the iron-siderophore complex. Gram-positive bacteria also have a TonB, ExbB and ExbD complex, and the inner-membrane permease and ATP-binding proteins involved in iron transport.

Iron toxicity. Although iron is very important for bacterial metabolism, too much iron can be toxic to cells. Free Fe^{2+} in bacteria and other cells can act as a catalyst to generate hydroxyl radicals (·OH) from hydrogen peroxide (H_2O_2). Generation of hydroxyl radicals from superoxide ($O_2^{\cdot -}$) and hydrogen peroxide is known as the Haber-Weiss cycle. Free hydroxyl radicals can cause damage to the cell, including damage to DNA. Therefore, bacteria tightly control the amount of iron taken into the cells to ensure that there is sufficient iron for cell growth and replication, but not an excess amount which would be toxic and cause cell death. Bacteria control iron levels in the cell by regulating the amount of siderophores and receptors produced. *See* ENZYME; HYDROXYL.

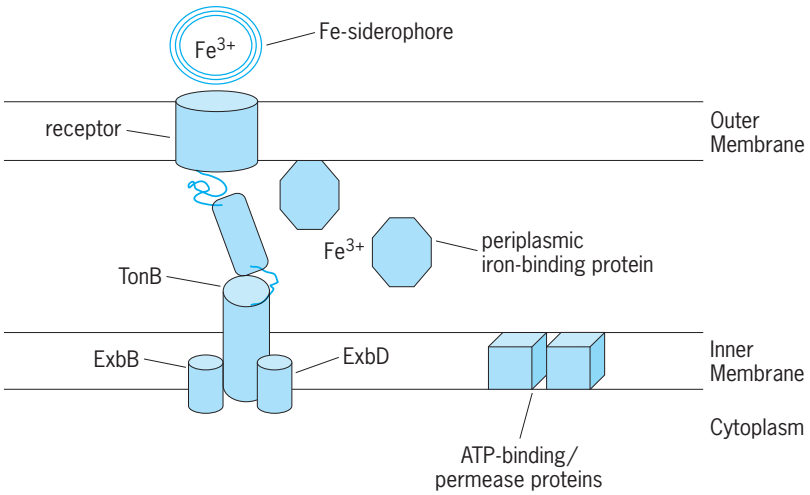


Fig. 2. Siderophore-iron complexes bind to a specific receptor in the outer membrane of gram-negative bacteria. The TonB and Exb proteins (ExbB and ExbD) form a complex that provides energy to the receptor to facilitate transport of Fe^{3+} into the periplasm. Periplasmic iron-binding proteins carry iron through the periplasm to the cytoplasmic membrane. ATP-binding proteins transfer Fe^{3+} into the cytoplasm, where it is needed for bacterial growth and metabolism.

Regulation of siderophore uptake. Siderophores and their specific receptors are produced only when bacteria sense that they are in an iron-deprived environment. Gram-negative bacteria control expression of the genes for siderophore synthesis and receptor expression with a protein called Fur (standing for ferric uptake regulation). When iron levels in the cell are sufficient for the needs of the bacteria, Fe^{2+} binds to Fur. The Fe-Fur complex binds to DNA upstream of the siderophore and receptor genes, and acts as a repressor of transcription so that these genes stop being expressed. Gram-positive bacteria also have proteins functionally similar to Fur called ideR (iron-dependent regulation) that repress siderophore production. Careful control of siderophore uptake systems guards against iron toxicity. *See* BACTERIAL GENETICS.

Some bacteria also have additional mechanisms that turn on production of specific siderophore receptors only when that siderophore is present. These regulatory processes consist of complex sensing mechanisms and transcriptional activator genes. These systems help conserve energy in bacteria that are capable of producing several different siderophore receptors, preventing the unnecessary

synthesis of receptors for siderophores not present in their environment.

Bacterial disease. Bacteria that infect humans and animals and cause disease symptoms must have a mechanism to remove iron from host iron proteins such as transferrin, lactoferrin, or heme. There are two mechanisms that bacteria use to get iron. Some bacteria, for example *Neisseria meningitidis*, bind transferrin, lactoferrin, or heme to specific receptors on their outer membrane. These receptors are similar to siderophore receptors. Iron is removed from the bound host proteins and internalized by a process that also requires the TonB complex. These bacteria cannot grow outside humans or animals because there is no transferrin or lactoferrin in soil or natural environments. However, most pathogenic bacteria do not directly bind host iron proteins, but produce siderophores and siderophore receptors that can remove iron from transferrin, lactoferrin, or precipitates in the environment. These bacteria can also live in soil and water, where host proteins are not present. *See* INFECTIOUS DISEASE; MENINGITIS.

There are many studies performed on culture media that demonstrate that siderophores promote growth of pathogenic bacteria in iron-starvation conditions. Mutant strains of bacteria that do not produce siderophores, siderophore receptors, or components of the uptake process such as TonB are not able to grow in these iron-restricted conditions. In addition, animal model systems have shown that siderophore production and uptake is required for virulence in some bacteria. *See* BACTERIAL GROWTH; BACTERIOLOGY.

Pseudomonas aeruginosa and *Burkholderia cepacia* cause chronic respiratory infections, and *P. aeruginosa* also causes burn-wound infections. In experimental animal infection studies involving these bacteria, siderophores have been shown to be required for virulence. *Yersinia pestis*, which causes bubonic plague, has been shown to require production of the siderophore yersiniabactin for virulence in mice. Other types of *Yersinia* that cause diarrhea and intestinal disease also produce this siderophore. The siderophore aerobactin has been shown to promote growth and virulence of *E. coli* and several other gram-negative pathogens involved in human and animal infections. Most of the invasive strains of *E. coli* that cause septicemia produce aerobactin. *Vibrio* species cause a variety of diseases in humans and fish and produce related siderophores. *Vibrio cholerae* causes the intestinal disease cholera and *V. vulnificus* causes septicemia and wound infections, respectively. *Vibrio anguillarium* causes fatal septicemia in salmon and is a serious problem in the fish industry. Siderophores are required for the virulence of *V. vulnificus* and *V. anguillarium*, and probably *V. cholerae*. Not as much is known about the role of siderophores in bacterial plant infections, but the siderophore chrysobactin contributes to the virulence of *Erwinia chrysanthemi*, a pathogen that causes disease in a wide range of plants. *See* CHOLERA; CLINICAL MICROBIOLOGY; FOOD POISONING; MEDICAL BACTERIOLOGY; PLAGUE; VIRULENCE; YERSINIA.

Biocontrol. Some soil and water bacteria have been found to naturally promote growth of plants. These bacteria aggressively colonize the roots of plants and prevent the growth and inhibit the virulence of many species of fungi as well as other bacterial species, including *Erwinia* and *Fusarium*. These pathogens destroy root hairs and root tips, resulting in root rot, wilting, and stunted growth. Production of siderophores by beneficial bacteria that bind iron and make it unavailable to the pathogenic bacteria or fungi is one mechanism by which bacteria can promote the growth of plants. The growth-promoting bacteria produce specific siderophores that cannot be used by the pathogenic organisms because they lack the appropriate siderophore receptor. The pathogens are therefore not able to compete for necessary iron in the environment of the plant root system. Bacterial species that promote growth of plants include *Pseudomonas*, *Rhizobium*, *Burkholderia*, *Alcaligenes*, and *Agrobacterium*. There is considerable interest in the agriculture industry in using bacteria as natural biocontrol agents instead of chemicals to inhibit growth of plant pathogens. *See* MICROBIAL ECOLOGY; PLANT PATHOLOGY; RHIZOSPHERE; ROOT (BOTANY).

Clinical uses. Siderophores have potential applications in the treatment of some human diseases and infections. Some siderophores are used therapeutically to treat chronic or acute iron overload conditions in order to prevent iron toxicity in humans. Individuals who have defects in blood cell production or who receive multiple transfusions can sometimes have too much free iron in the body. However, in order to prevent infection during treatment for iron overload, it is important to use siderophores that cannot be used by bacterial pathogens.

A second clinical application of siderophores is in antibiotic delivery to bacteria. Some gram-negative bacteria are resistant to antibiotics because they are too big to diffuse through the outer-membrane porins. However, siderophore-antibiotic combination compounds have been synthesized that can be transported into the cell using the siderophore receptor. The antibiotic is delivered into the cell, where it can act on its target to either inhibit growth or kill the bacterial pathogen. This approach to development of new antibiotics is under investigation. *See* ANTIBIOTIC; DRUG RESISTANCE.

Pamela A. Sokol

Bibliography. V. Braun and H. Killmann, Bacterial solutions to the iron-supply problem, *Trends Biochem. Sci.*, 24:104–109, 1999; J. J. Bullen and E. Griffiths (eds.), *Iron and Infection: Molecular, Physiological and Clinical Aspects*, Wiley, New York, 1999; R. B. Byers and J. E. L. Arceneaux, Microbial iron transport: Iron acquisition by pathogenic microorganisms, *Metal Ions Biol. Sys.*, 35:37–66, 1998; D. J. O'Sullivan and F. O'Gara, Traits of *Pseudomonas* spp. involved in suppression of plant root pathogens, *Microbiol. Rev.*, 56:662–676, 1992; C. Ratledge and L. G. Dover, Iron metabolism in pathogenic bacteria, *Annu. Rev. Microbiol.*, 54:881–941, 2000.

Sigma-delta converter

A class of electronic systems containing both analog and digital subsystems whose most common application is to the conversion of analog signals to digital form and vice versa. The device is also known as a delta-sigma converter. The main advantage of the sigma-delta approach to signal conversion is its minimal reliance on the quality of the analog components required. To achieve this end, the system uses pulse density modulation to create a high-rate stream of single-amplitude pulses. For analog-to-digital conversion, the rate at which the pulses are generated depends on the amplitude of the analog voltage being sensed. For digital-to-analog conversion, the pulse density depends on the numeric digital quantity applied at the converter input. *See* PULSE MODULATION.

The simplest implementation of a sigma-delta analog-to-digital converter uses an analog circuit to generate the single-valued pulse stream from an analog source, and a digital system to repeatedly sum the number of these pulses over a fixed number of pulse intervals. The summing operation converts the pulses to a numeric value, achieving analog-to-digital conversion. Conversely, in a digital-to-analog converter, a digital circuit is used to convert numeric values from a digital processor to a pulse stream, and these pulses then are low-pass-filtered by a relatively simple analog system to produce an analog waveform. This low-pass filtering effectively sums the uniform analog pulse amplitudes over a fixed interval. The circuit—analog or digital—used to generate the pulse stream is called a sigma-delta (or delta-sigma) modulator. *See* ELECTRIC FILTER.

In each case, analog information is contained not in the pulse amplitude but in the number of pulses that occur during the conversion interval. This distribution of the analog information makes the conversion process essentially independent of the amplitude of the pulses and greatly simplifies the design and fabrication of the analog portion of the converter. It does, however, require that the sampling process be rapid, since the resolution of the conversion depends on the number of pulses that can exist in the conversion interval. For the simplest sigma-delta converter implementation, for example, if a maximum of n pulses can exist in the conversion interval, only $n + 1$ (from zero to n inclusive) different numeric values can be generated. The high-rate pulse production process, where a multitude of pulses are generated in the conversion interval, is called oversampling and usually is described as the ratio of the number of high-rate pulse intervals to the number of conversion intervals in which they exist. This oversampling ratio determines the maximum resolution of the converter. Converters are characterized by their maximum resolution, and the number of pulses in the conversion interval is usually chosen to be a power of 2. Thus, a 2^n converter is called an n-bit analog-to-digital (or digital-to-analog) converter. An eight-bit converter, for example, has a resolution of one part in two to the eighth power, or 256 levels. Therefore this simplest version of an eight-bit sigma-delta converter needs 256 pulse periods in its basic conversion interval.

The reciprocal of the duration of the conversion interval over which the pulses are counted is called the sampling frequency. This sampling frequency sets the highest rate at which a signal can change and still be represented adequately by a digital data set. The most rapid signal permitted must change from high to low and back to high in exactly two sampling intervals. This highest frequency, exactly one-half the conversion rate, is called the Nyquist frequency. A digital processor in a digital-to-analog converter cannot generate signals higher than this rate. Conversely, if an analog-to-digital converter tries to sample signals higher than this rate, they will look like frequencies lower than the desired one to the converter and will be digitized as such. This undesired frequency translation effect is called aliasing. *See* ELECTRICAL COMMUNICATIONS; INFORMATION THEORY; PULSE MODULATOR.

Sigma-delta modulator. Sigma-delta modulators used to generate a pulse-density-modulated data stream can take a variety of forms, depending on their application. The simplest version, a single-loop sigma-delta modulator (**Fig. 1**), has three essential components: (1) a differencing circuit, which takes the difference between the value of the quantity being converted and a feedback signal from a quantizer; (2) a summing circuit, that adds these

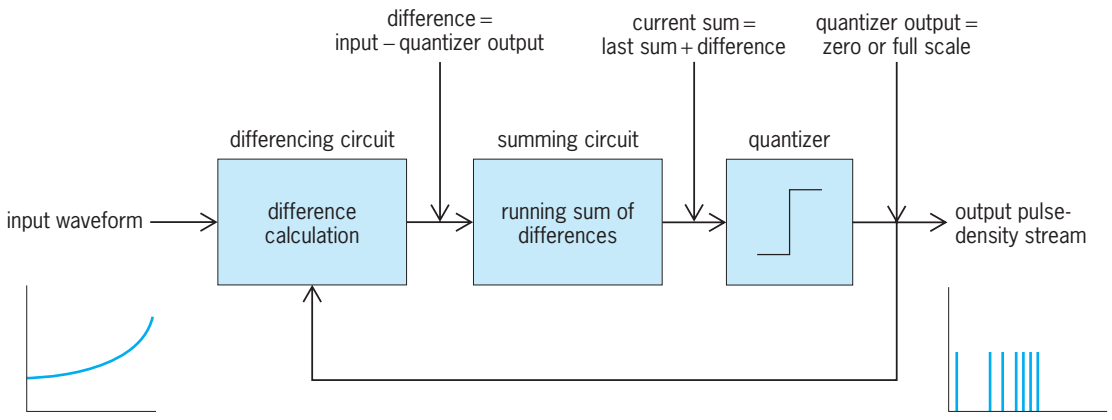


Fig. 1. First-order sigma-delta modulator block diagram.

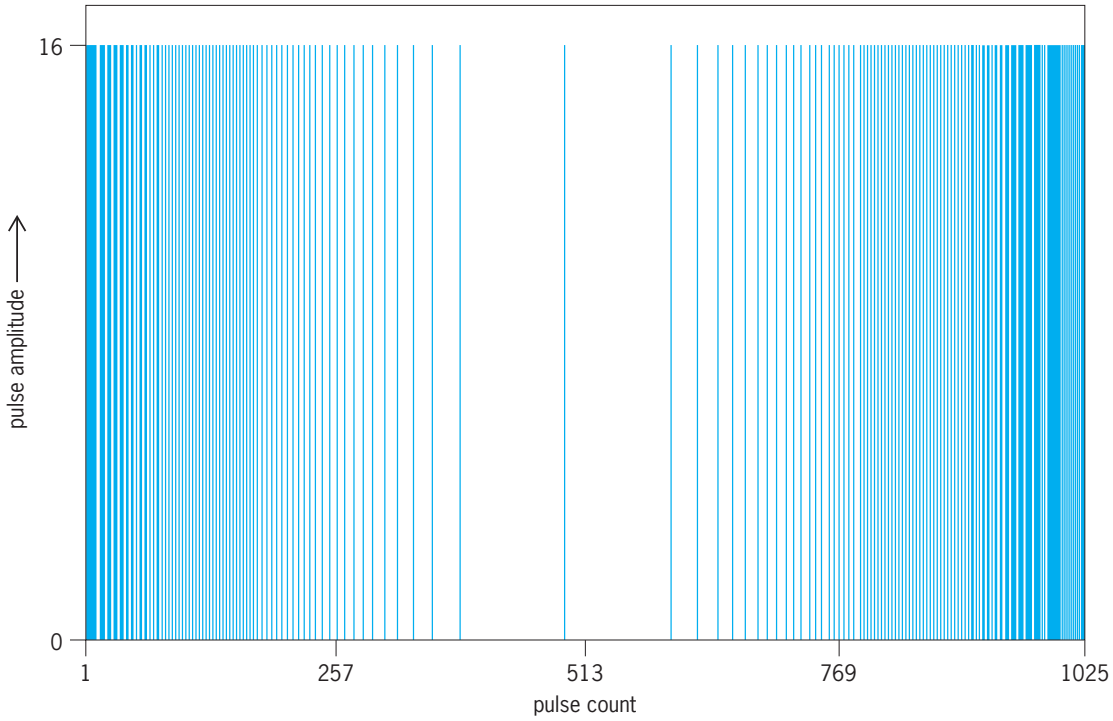


Fig. 2. Pulse density produced for a half-cycle of a periodic input of length 2048.

differences to previously calculated ones; and (3) a quantizer, which produces a signal equal to either the maximum allowable input value or zero. The quantizer output is maximum if the output of the summing circuit is greater than zero, and is zero if the summing output is negative. The input to this sigma-delta model can be considered to be either digital or analog. The modulator's name comes from the conventional mathematical symbols used to represent the two operations that it performs: differencing (delta) and summing (sigma). The 1024-pulse density pattern produced at the quantizer output for one-half cycle of a periodic input waveform is shown in **Fig. 2**.

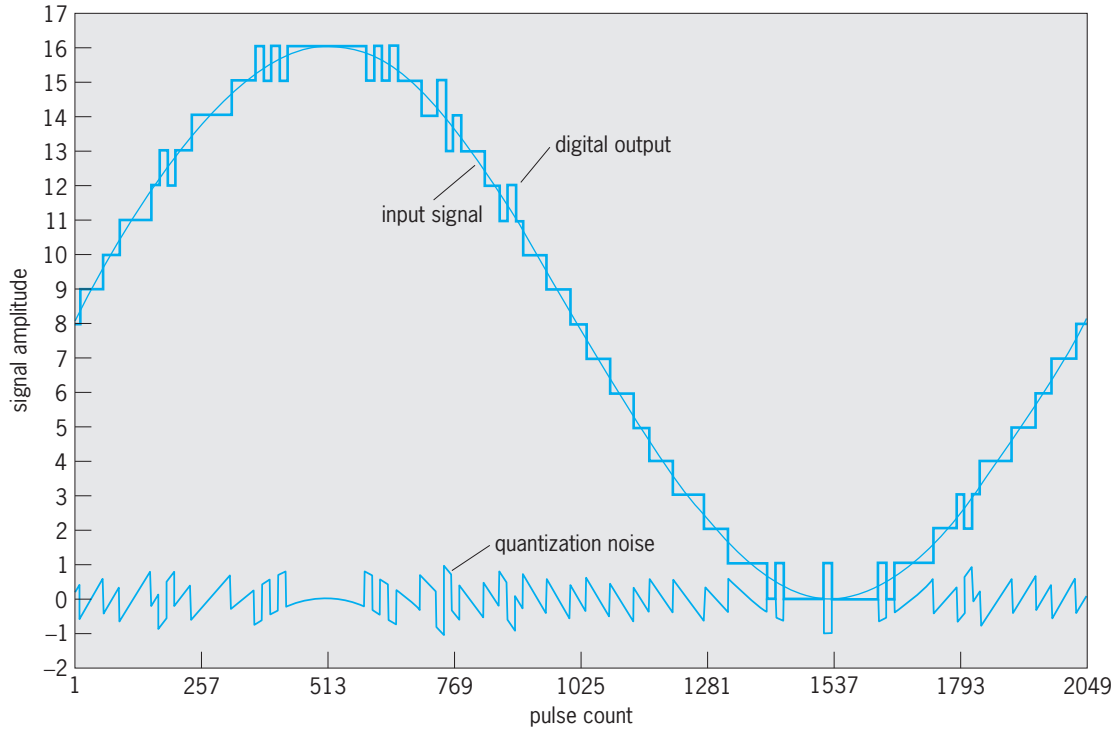


Fig. 3. Digital output produced with a 16-pulse decimation interval.

Low-pass filter, or decimator. To recover the quantized amplitude of the input signal from the pulse-density waveform, the pulses can be averaged over exactly one conversion period. For an analog-to-digital converter, this means counting the number of pulses over the conversion interval and dividing the result by the maximum number of pulses allowed in the interval. The digital filter that performs these operations automatically produces a binary number equal to the pulse count at the end of each conversion interval. The converter produces an output once every conversion interval, or once every 2^n pulses. This process, which produces data at a rate lower than the original pulse rate, is called decimation.

The output produced by a simple averaging decimation filter operating on the output of a single-loop sigma-delta modulator is shown in **Fig. 3**. The oversampling rate for this example is 16. Therefore, with simple averaging decimation, the output signal has just 17 levels (from 0 to 16) with which to represent the input. The digitized input signal is included on the plot, along with a third waveform, the quantization noise introduced by the digitizing process. Practical applications employ much higher resolution than the four-bit resolution of this example, and the quantization noise correspondingly is much lower.

Sigma-delta digital-to-analog conversion. Digital-to-analog conversion based on the pulse density concept is performed in a manner substantially identical to analog-to-digital conversion. In this case, the sigma-delta modulator is implemented as a digital process. A digital quantizer value is subtracted from the numeric value to be converted, and the resulting difference is added to preceding differences in a digital summing, or integrator, circuit. If the integrator produces a negative output, the quantizer output level becomes zero. A positive integrator output, on the other hand, sends the quantizer to its high level. The pulses produced by the modulator in this case are treated as analog signals. A suitable analog low-pass filter then processes these pulses. The net effect of this analog filtering is effectively to add the pulses over an interval that corresponds to the conversion interval used in analog-to-digital conversion.

Many variations on this basic sigma-delta configuration have been investigated and have been used in a multitude of practical applications. The one most commonly encountered in everyday use is the ubiquitous "one-bit digital-to-analog converter" found in audio compact disk players. *See* ANALOG-TO-DIGITAL CONVERTER; COMPACT DISK; DIGITAL-TO-ANALOG CONVERTER.

Philip V. Lopresti

Bibliography. J. C. Candy and G. C. Temes (eds.), *Oversampling Delta-Sigma Data Converters: Theory, Design, and Simulation*, IEEE, New York, 1992; J. A. Cherry and W. M. Snelgrove, *Continuous-Time Delta-Sigma Modulators for High-Speed A/D Conversion: Theory, Practice and Fundamental Performance Limits*, Kluwer, New York, 1999; S. R. Norsworthy, R. Schreier, and G. C. Temes (eds.), *Delta-Sigma Data Converters: Theory, Design, and Simulation*, IEEE, New York, 1996.

Signal detection theory

A theory in psychology which characterizes not only the acuity of an individual's discrimination but also the psychological factors that bias the individual's judgments. Failure to separate these two aspects of discrimination had tempered the success of theories based upon the classical concept of a sensory threshold. The theory provides a modern and more complete account of the process whereby an individual makes fine discriminations.

Modern detection theory was developed by communications engineers to help distinguish weak radar signals from the static in their background. Quite naturally, the first applications in psychology were to discriminations that taxed sensory acuity. The theory has been applied as well to cognitive discriminations, made under limitations imposed, for example, by memory or available information.

Parameters and methods. Quantitative investigations of sensory acuity fall under the heading of psychophysics, the study of relations existing between energy changes in the environment and the attributes of sensation. These attributes include intensity, quality, duration, and extent. Energy changes commonly used as signals in laboratory experiments include flashes of light, bursts of tones, gases directed to the smell receptors, solutions applied to taste buds, and vibrations applied to the skin. *See* PSYCHOPHYSICAL METHODS.

In cognitive psychology signals richer in meaning are used, for example, brief presentations of printed or spoken words. A subject might be asked whether or not the word on a card was on a list of words the subject examined earlier, whether a given statement of opinion most likely appeared in magazine A or magazine B, or whether a particular journal article, represented by its abstract, is relevant or irrelevant to a particular need for scientific information.

When the discrimination called for is between the presence and absence of a specified signal—is the light present? is a word present?—the process is called detection. Distinguishing a particular signal from among a set of possible signals—is the light red or green? which word is being flashed?—is called recognition or identification. In every instance the individual observes a stimulus, which may or may not be a signal, and attempts to choose the response alternative associated with that stimulus. The discrimination is difficult to make because the signal is weak or blurred by some random background interference (noise) or masked by another stimulus or similar to other signals that must be identified.

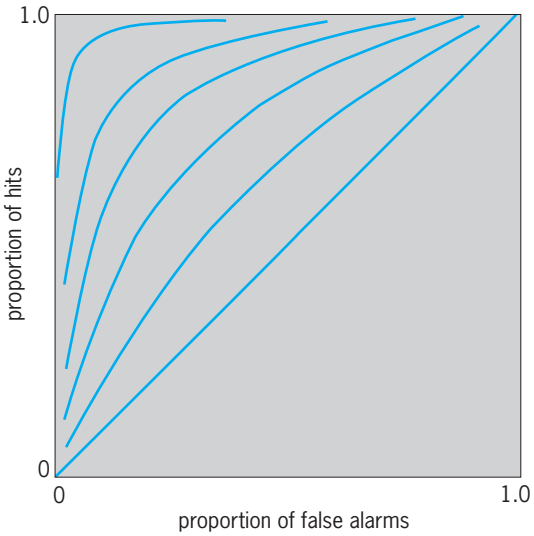
Components of the theory. The theory of signal detection has two parts of quite different origins. The first comes from mathematical statistics and is a translation of the theory of statistical decisions. The major contribution of this part of the theory is that it permits a determination of the individual's

discriminative capacity, or sensitivity, that is independent of the judgmental bias or decision criterion he may have had when he made the discrimination. The second part of the theory comes from the study of electronic communications. It provides a means of calculating for simple signals, such as tones and lights, the best discrimination that can be attained. The prediction is based upon physical measurements of the signals and their interfering noise.

This opportunity to compare the sensitivity of human observers with the sensitivity of an "ideal observer" for a variety of signals is of considerable usefulness, and of growing interest, in sensory psychology. However, the ability afforded by the first part of the theory to analyze discrimination decisions into two components, thereby separating the influences of discriminative capacity and decision criteria, is of more general relevance to psychology and is the subject of the remainder of this article.

Analysis of discrimination decisions. A simple example will illustrate the main points of the theory. To begin, it may be generally acknowledged that children tend to detect the bell of an ice cream vendor at a greater distance than do their parents. Do children hear better than their parents? That is quite possible; there is a defect of hearing associated with advancing age termed presbycousis. On the other hand, it is also possible that children are more likely to interpret any faint tinkling sound as having been produced by the vendor's bell. That is, children may have a greater bias in favor of making a response like "The ice cream truck!" than their parents do; they may have, in the language of decision theory, a more lenient criterion for making a positive judgment about an ambiguous stimulus of this kind. This is only common sense: One perceives what one wants or expects to perceive. In decision theory lingo, the value of detecting the bell when it is there and the cost of failing to detect the bell when it is there are higher for the child than for the adult. The probabilities also favor a more lenient decision criterion for the child; the child is actively expecting the bell to ring any minute, while the parent is thinking about something else. Then again, the decision criterion of an individual may vary from time to time, in a way independent of changes in his or her acuity. For the child in this example, a cool day, with a lower value of ice cream and a lower probability of the bell being in the neighborhood, may produce a more stringent criterion for a positive response.

Measurement and evolution. It is clear, according to the theory of signal detection, that recording only the proportion of times a signal is detected when it is present, called the proportion of "hits," fails to give an accurate measure of an individual's ability to discriminate between signal presence and absence. This proportion of hits will be higher or lower not solely as a result of differences in acuity, but also as a result of factors unrelated to acuity, such as expectancy and motivation. The theory indicates the need to obtain a second proportion: the proportion of times a signal is detected when it is, in fact, absent. This proportion is called the proportion of "false alarms" and tells a good deal about the individual's response bias or decision criterion. A lenient criterion, of necessity, leads to more false alarms, as well as to more hits, than a strict criterion. This is true in any discrimination problem where the discrimination is less than perfect. *See* MOTIVATION.



Receiver operating characteristics.

The procedures and analysis suggested by the theory for a simple detection problem are briefly as follows: Many trials are defined; on some a specified signal is present while on others, randomly selected, the signal is absent. From one group of trials to another the subject is encouraged to shift his or her criterion for a positive response, perhaps by changes in the values and costs assigned to the various possible decision outcomes (the payoff matrix) or by changes in the probability of a signal being present.

Plotting the proportion of hits against the proportion of false alarms, for each of several decision criteria, will lead to a curve similar to one of those shown in the **illustration**. By varying physical characteristics (for example, the energy level) of the signal or noise, a number of such curves is obtained. Each curve, called a receiver operating characteristic, shows the trading relation between the two types of response outcome for discrimination of a given degree. The degree of acuity corresponding to each curve can be represented by a single number, which gives essentially the distance of the curve from the diagonal line (no discrimination) toward the upper left-hand corner (perfect discrimination). This number, it can be seen, is independent of the decision criterion. If one wishes to represent any criterion the subject may have adopted, then one can abstract a second number from the curve, such as the slope of the curve at the point yielded by that criterion.

Other experimental procedures used in the context of detection theory typically yield measures of discrimination that agree well with those obtained from the procedure exemplified—the so-called yes-no procedure. These include the confidence-rating

procedure, in which the subject is asked to relate each observation to several criteria held simultaneously, and the forced-choice procedure, in which the task is to state in which one of a specified set of intervals, of time or space, the signal was presented. The latter procedure largely bypasses the issue of decision criteria, and it is often used in studies directed solely at questions of sensory function.

Comparison with threshold theories. The theory of signal detection is an alternative to theories based on the classical concept of a threshold. As indicated, detection theory sees the information obtained from an observation as ranging along a continuum of more or less certainty about the stimulus that gave rise to it and sees the observer as able to place a decision criterion at any point on the continuum. On the other hand, threshold theories view the sensory information available to the observer as "all or none" or, at most, as falling into one of very few categories of relative certainty, with the categories demarcated by unalterable criteria. The term threshold connotes a barrier that must be overcome: The notion is that when the observation yields enough information to exceed the barrier, then a signal is clearly seen or heard or felt, as the case may be, and that when the information does not exceed this fixed quantity, then absolutely no information about the signal exists. The older threshold theories are incompatible with much of the present evidence. The measures of discrimination they suggest are not only confounded by response biases, but the values of these measure depend too heavily upon which experimental procedure is used to determine them. *See* PSYCHOPHYSICAL METHODS.

Applications. Signal detection theory has been applied to several topics in experimental psychology in which separation of intrinsic discriminability from decision factors is desirable. Included are attention, imagery, learning, conceptual judgment, personality, reaction time, manual control, and speech. Also included are studies of animal behavior, with rats, pigeons, and monkeys producing neat curves of the sort shown in the illustration. These curves are called ROC curves, for receiver (or relative) operating characteristic.

The analytical apparatus of the theory, primarily the ROC curve, has been of value in the evaluation of the performance of systems that make decisions based on uncertain information. Such systems may involve only people, or people and machines together, or only machines. Examples come from medical diagnosis, where clinicians may base diagnostic decisions on a physical examination, or on an x-ray image, or where machines make diagnoses, perhaps by counting blood cells of various types. Further, manual and automatic systems for retrieving articles in a library have been compared for accuracy relative to cost. And one may attempt to determine the validity of systems designed for forensic purposes, for example, for identifying an unknown speaker on the basis of a recording, or for detecting lies by means of a polygraph test. *See* HEARING (HUMAN); LIE DETECTOR; PSYCHOLOGY; VISION. John A. Swets

Bibliography. J. P. Egan, *Signal Detection and ROC-Analysis*, 1975; D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, 1966, reprint 1989; J. A. Swets, The relative operating characteristic in psychology, *Science*, 182:990–1000, 1973; J. A. Swets, *Signal Detection and Recognition by Human Observers*, 1964, reprint 1989.

Signal generator

A piece of electronic test equipment that delivers a sinusoidal output of accurately calibrated frequency. The frequency may be anywhere from audio to microwave, depending upon the intended use of the instrument. The frequency and the amplitude are adjustable over a wide range. The oscillator must have excellent frequency stability, and its amplitude must remain constant over the tuning range.

The Wien-bridge oscillator is commonly used for frequencies up to about 200 kHz. For a radio-frequency signal generator up to about 200 MHz, a resonant circuit oscillator is used (such as a tuned-plate tuned-grid, Hartley, or Colpitts). Beyond this range, very high frequency and microwave oscillators are used.

Many signal generators contain circuitry that allows the output to be either amplitude- or frequency-modulated. The most common forms of amplitude modulation are sinusoidal, square-wave, and pulse. The frequency is kept constant, is sinusoidal-modulated, or is swept linearly across a band of frequencies. For example, for testing broadcast receivers, it is important to sweep the generator frequency over a range of ± 10 kHz at a low rate, say 60 times a second. *See* OSCILLATOR; PULSE GENERATOR; WAVE-SHAPING CIRCUITS. Jacob Millman

Bibliography. A. D. Helfrick and W. D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, 1990; J. Markus and C. Weston, *Essential Circuits Reference Guide*, 1988; D. L. Schilling et al., *Electronic Circuits: Discrete and Integrated*, 3d ed., 1989.

Signal processing

Operations that modify or create signals. A simple representation of a signal processing operation has an input signal A that is modified to produce an output signal B (**Fig. 1**).

For a signal to be useful, it must have one or more characteristics that vary, usually over time or distance, to represent information. Amplitude, frequency, and phase are common signal characteristics

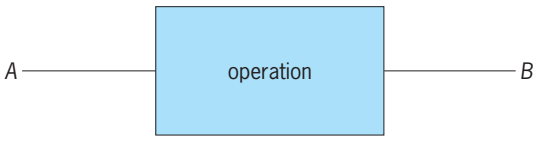


Fig. 1. Simple signal processing representation.

that can carry information. Time-varying signals are written as $A(t)$. *See* AMPLITUDE (WAVE MOTION); FREQUENCY (WAVE MOTION); PHASE (PERIODIC PHENOMENA).

A simple signal processing operation is scaling, which multiplies the input signal by a constant value to produce the output signal. Another simple operation is the addition of a constant value to the input to produce the output.

A signal processing operation may have more than one input. For example, two input signals may be added together to produce an output signal. Addition is a linear mathematical operation. An operation may also be nonlinear, as is the case where an input signal is squared or when two input signals are multiplied together.

Logical operations may also be employed in signal processing to make decisions and route signals. For example, a switch may be used to select between two signals based on the value of a third signal. *See* LOGIC CIRCUITS.

An important signal processing operation is time delay. The output signal is a delayed copy of the input signal.

Simple mathematical, logical, and time-delay operations are combined into signal processing systems that perform more complex processing operations. The simple operations, or blocks, are connected together in an appropriate manner to implement the desired complex operation. **Figure 2** shows the block diagram of a system that performs a time-averaging operation. The input signal and a delayed version of the input signal are added together and then scaled by 0.5 to produce the averaged output.

Signal processing is performed in analog and digital circuits. Analog signal processing has been used since the beginning of audio and radio electronics. Over the past few decades, digital signal processing has become an increasingly common method, especially for more complex systems. In many cases, the same operation can be performed equivalently in both the analog and digital domains, although one may have advantages depending on various factors, such as the physical domain, complexity, and required adjustability of the operation.

Because computers use numbers and perform calculations on numbers, they can be programmed to perform digital signal processing. By approximating analog signals and operations in digital form, analog signal processing can also be performed on computers using simulation programs. *See* COMPUTER; DIGITAL COMPUTER; SIMULATION.

Characteristics of operations and systems. Signal processing operations and systems can be described in different ways. The choice depends on the physical process being described, the approach to designing the processing system, the desired outcomes of the design, the modeling of the design, and the capabilities of the designer and design tools.

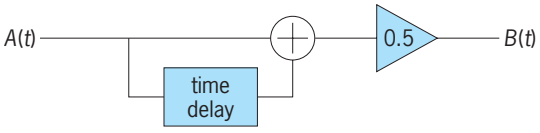


Fig. 2. Signal averager.

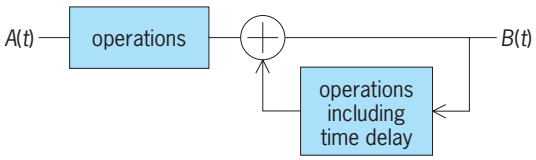


Fig. 3. System containing feeback.

Linear and nonlinear signal processing. One characteristic already discussed is whether an operation is linear or nonlinear. An operation with input x and output y is linear if Eq. (1) holds true, where m and b are

$$y(t) = m \cdot x(t) + b \tag{1}$$

constants. A nonlinear operation does not meet this requirement.

If an operation does not vary over time, as in Fig. 2 or in Eq. (1), it is said to be time-invariant. If instead the value of the scaling in Fig. 2 changes over time, the scaling operation and thus the whole system is considered time-variant. Time-invariant systems are generally easier to design, model, and analyze.

Linear time-invariant systems are by far the signal processing systems most commonly designed, analyzed, and implemented. They can be represented by simple mathematical equations called transfer functions, which describe completely the input-output relationship of the system for any input signal. *See* LINEAR SYSTEM ANALYSIS; LINEARITY.

Feedback systems. A system may contain feedback, which means at least one signal within the system returns to a point that indirectly feeds into the same signal, creating a loop. Delay must be present in the loop. The system in Fig. 2 does not have feedback. A general representation of a system with feedback is shown in **Fig. 3**. The output signal $B(t)$ is fed back through operations that include time delay, with the delayed result added to the processed input to create the output. A key attribute of a feedback system is whether or not it is stable. An unstable system will produce an output that is uncontrolled and unpredictable, making the system unusable. Methods that analyze the operations in the loop can be used to determine if a particular feedback system is stable. *See* CONTROL SYSTEM STABILITY; CONTROL SYSTEMS.

Real and complex signal processing. The representation of signals as having complex values is a very useful tool in signal processing. Here, complex does not refer to the complexity of a signal but rather to a signal having real and imaginary components. A signal that is not complex may be described as real or scalar. Operations on complex signals are also complex and have real and imaginary parts. Complex signals and operations allow certain types of signal processing tasks, like those in digital radio systems, to be performed more efficiently than if scalar signals and operations are used. They also allow signal processing implementations to follow more directly the mathematical models on which they are based.

To denote the imaginary part of a complex signal, the symbol i is used. The complex signal $C(t)$ is represented with $A(t)$ as the real part and $i{\cdot}B(t)$ as the imaginary part, as in Eq. (2). The symbol i is generally used in mathematical equations, but in signal processing the equivalent symbol j is more common. The quantities $A(t)$ and $B(t)$ are themselves real. In a signal processing system these can be signals that could for example be carried on two wires. The wires taken together are treated as a real-imaginary pair forming a complex signal. Thus, even though the complex signal has a so-called imaginary component, the signal can exist in the real world and complex operations on the signal can be performed.

$$C(t) = A(t) + i{\cdot}B(t) \tag{2}$$

The symbol j actually has the value of $\sqrt{-1}$. Although this expression cannot be calculated, it can be used in calculations describing complex operations. For example, the product $P(t)$ of two complex signals $C(t)$ and $D(t)$ is calculated in Eq. (3). In block diagram form, the product can be represented as a complex operation on complex signals or as real operations on real signals (**Fig. 4**).

$$\begin{aligned} P(t) = C(t) \cdot D(t) &= [A(t) + j \cdot B(t)] \cdot [E(t) + j \cdot F(t)] \\ &= A(t) \cdot E(t) + j \cdot B(t) \cdot j \cdot F(t) \\ &\quad + A(t) \cdot j \cdot F(t) + j \cdot B(t) \cdot E(t) \\ &= A(t) \cdot E(t) + j \cdot j \cdot B(t) \cdot F(t) \\ &\quad + j \cdot A(t) \cdot F(t) + j \cdot B(t) \cdot E(t) \\ &= A(t) \cdot E(t) - B(t) \cdot F(t) \\ &\quad + j \cdot [A(t) \cdot F(t) + B(t) \cdot E(t)] \\ &= V(t) + j \cdot W(t) \end{aligned} \tag{3}$$

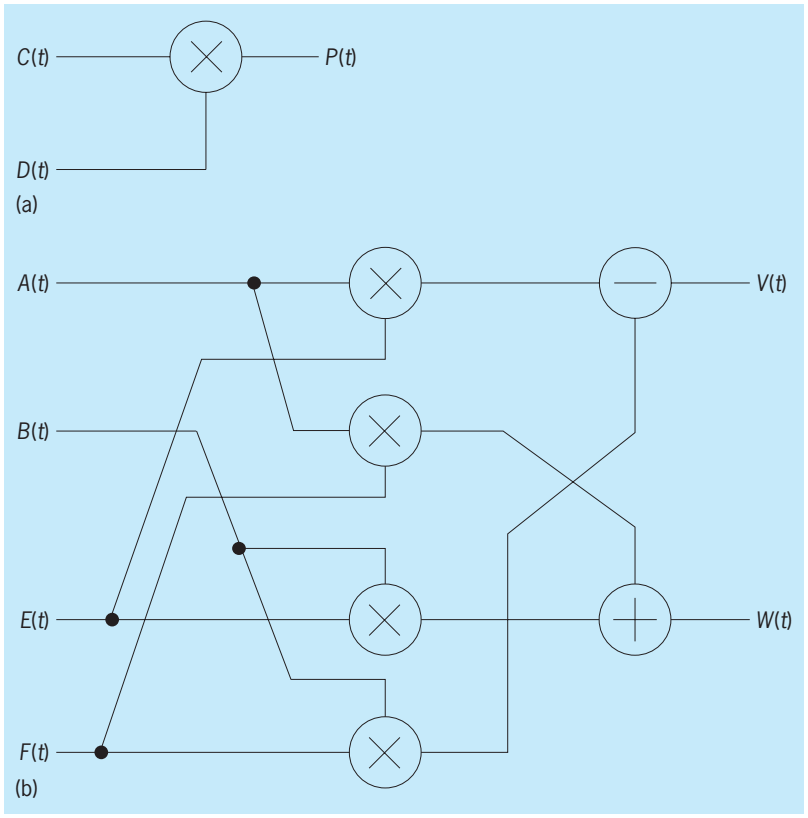


Fig. 4. Complex product operation (*a*) represented as a complex operation on complex signals, (*b*) represented as real operations on real signals.

The result is also complex. If Eq. (3) were to be implemented in an electronic circuit, all of the real signals $A(t)$ to $F(t)$ would exist as electronic signals. Similarly, if the complex operation were to be implemented in a computer program, these signals would be represented by program variables. Operations that process scalars usually have equivalent complex versions.

A complex signal $A(t) + j \cdot B(t)$ can also be represented in the form $M(t) \cdot e^{j \cdot \varphi(t)}$, where $M(t)$ and $\varphi(t)$ are the magnitude and phase of the signal. The magnitude and phase can be calculated from the real and imaginary parts, as in Eqs. (4).

$$M(t) = \sqrt{A^2(t) + B^2(t)} \tag{4a}$$

$$\varphi(t) = \arctan[B(t)/A(t)] \tag{4b}$$

Signal processing in multiple dimensions. Signal processing may be performed across a one-dimensional domain, like time, or in two dimensions. Time-based processing has already been discussed. A common application of two-dimensional processing is the generation and manipulation of visual images. The domain is space instead of time, and the space is typically measured along a horizontal and vertical axis. The image is represented as a two-dimensional array of pixels, and processing of the image involves operations on individual pixels or groups of pixels. For example, a simple way to soften an image is by averaging the intensities of its pixels. This reduces the intensity variations from pixel to pixel, and is performed along both axes, averaging short groups of adjacent pixels first in each horizontal line of the image, and then in each vertical line. Other operations can be performed to sharpen images, remove noise, change coloration, extract features, and combine images. Image-processing operations can have the various characteristics already discussed, such as nonlinearity. *See* IMAGE PROCESSING.

Processing in higher dimensions and in combinations of different domains (like time and space) are also possible. This depends only on how the signal processing designer chooses to represent a signal processing system and on the design tools available.

Creating signals. In addition to operating on input signals, signal processing is also used to create signals. For example, a sine wave oscillator is a signal source employed in radio and music synthesis systems. A signal source may have control inputs to set its output characteristics, like an oscillator's frequency and amplitude. *See* OSCILLATOR.

Analog signal processing. Whether a signal processing operation or system is implemented in analog or digital form is one of its most important characteristics. This is because the methods for implementing each form are very different. Analog processing is performed in traditional analog circuits, which are built from analog components. The signals in these circuits carry information according to

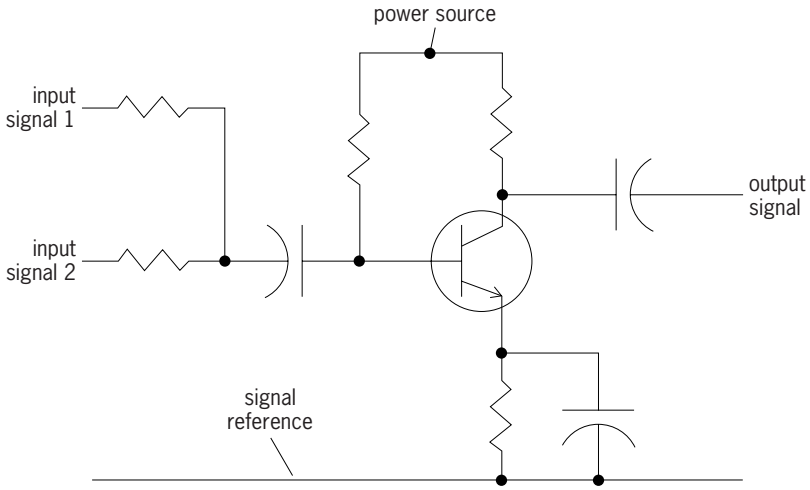


Fig. 5. Amplifier using a transistor, resistors, and capacitors.

the time-varying values of their voltage and current. Analog signals and operations are continuous in value and in time.

Analog processing components. Two kinds of analog components are used: passive and active. Resistors, capacitors, inductors, and transformers are passive components, and the operation of each can be described by a simple relationship between the voltage across the terminals of the component and the current passing through the component. Passive components do not require an external power source to operate. *See* CAPACITOR; INDUCTOR; RESISTOR; TRANSFORMER.

Active components are based on semiconductors or in older or high-power circuits on electron tubes. Transistors and diodes are common types of semiconductor components. Operations that require some type of amplification or nonlinear operation must use active components, and therefore a power source is required. To form useful signal processing operations, semiconductor components must be combined with passive components to create circuits, such as an amplifier that sums two input signals (**Fig. 5**). *See* AMPLIFIER; DIODE; ELECTRON TUBE; TRANSISTOR.

Integrated analog processing. A very useful analog device is the operational amplifier, or op amp. It is a circuit consisting of active and passive components that have been packaged, or integrated, into a single, higher-level active component. Op amps were created to make analog signal processing design easier than designing circuits based directly on semiconductors and supporting passive components. This in turn allows more complex circuits to be designed, using multiple op amps mixed with other active and passive components. *See* OPERATIONAL AMPLIFIER.

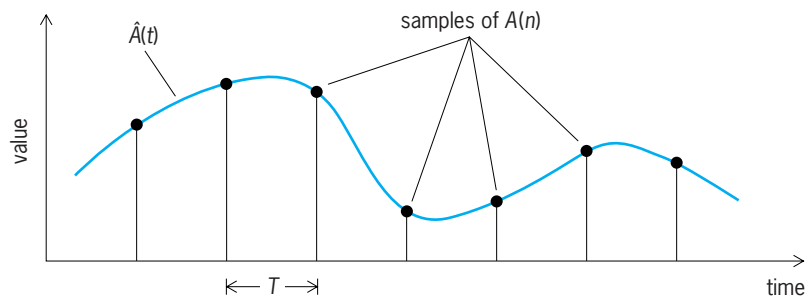


Fig. 6. Time-sampled signal.

Very complex and high-performance analog signal processing systems can be integrated into a single component, or integrated circuit. This allows, for example, most of the signal processing for a radio or television to be contained in a package the size of a quarter, helping to reduce the cost to manufacture these products. *See* INTEGRATED CIRCUITS.

Analog processing circuit applications and design. Despite the now widespread use of digital signal processing, analog signal processing remains very important. Analog circuits must be used for operations on signals that vary at high frequencies, approximately 100 MHz and higher. Certain types of processing, for example in the input stages of radio and television receivers, require analog circuits combining high-frequency signal handling with the ability to input signals at very weak and very strong levels. High-power processing, such as that of a radio transmitter, must also be performed in analog. Even in digital signal processing systems, analog processing must be employed to convert between the digital domain and the mostly analog domain of the physical world. *See* RADIO RECEIVER; RADIO TRANSMITTER; TELEVISION RECEIVER.

Digital signal processing. Up until the past few decades, most electronic products, such as telephones, radios, televisions, and sound processors (like audio mixing consoles), were implemented completely with analog processing circuits. Slowly, more of the processing in many products has been changed to digital implementations. Two important aspects of all-digital signal processing systems separate them from analog systems: sampling and quantization.

Sampling. In digital signal processing (DSP) systems, signals exist only at discrete points, usually in time or space. Such signals are called sampled signals and have a value at each of these points. Thus, a sampled signal A can also be written as $A(n)$, where n is an index of the values, or samples (**Fig. 6**). Usually the samples are arranged in time or space so that the time or distance between any two adjacent samples is the same, or uniform. The time or distance between adjacent uniform samples is the sample period, and the inverse of the sample period is the sample frequency, or sample rate. For sample period T, $A(n)$ is the sequence of samples of signal A at the discrete points in time $n \cdot T$. It is convenient to view $A(n)$ as having been created by sampling a continuous-time signal $\hat{A}(t)$, although a sampled signal does not need to originate from a continuous-time signal.

How a signal varies over time or space can be described by how fast it changes, that is, by its frequency content. A sampled signal can contain frequencies no higher than one-half its sample rate. And, for a signal with bandwidth B to be accurately represented as a sampled signal, it must be sampled

at a rate $2 \cdot B$ or higher. This minimum rate is called the Nyquist rate, and half this rate is the Nyquist frequency. In practice, the sample rate must be somewhat higher than the Nyquist rate. *See* INFORMATION THEORY.

Quantization and signal representations. In digital signal processing systems, signals are represented by binary numbers, and as such are quantized, meaning they can take on only a limited set of values. For example, an 8-bit form of a signal can take on values from 0 to 255, or −128 to 127, depending on what the signal represents or how it is processed. The quantization of a signal means that for each sample, its value must be approximated to one of the values, or quantization steps, allowed by the binary form representing the signal.

This approximation process adds quantization noise to the signal, because each sample has some amount of error compared to the exact value of the signal. With the proper design of a digital signal processing system, quantization noise can be controlled and will have a minimal adverse effect on the system's operation. *See* PULSE MODULATION.

Two forms of numbers are used to represent signals. The simpler form is fixed point, meaning the binary point, similar to a decimal point, is in a fixed position. In the second form the binary point is allowed to move, or float, to allow a wider range of values to be represented, and so this form is called floating point. Some of the bits of the number indicates the position of the binary point.

Because digital signal processing uses numbers for signals, processing operations can be represented directly by mathematical and logical equations. The general signal processing operation of the addition of two signals, for example, is performed by the actual addition of the binary values representing the signals. A delay operation is easily implemented by delaying a signal by a number of sample periods using a digital delay line. Multiplication and division are performed in the manner of long-hand arithmetic.

Digital signal processors and other implementations. As with analog signal processors, complex digital signal processing systems may be combined into single, highly integrated circuits. These are commonly called digital signal processors, and are commercially available from many sources and for many purposes. Their processing is controlled by internally or externally stored programs consisting of sequences of simple program steps, or instructions, that execute mathematical and logical operations. These processors can be programmed by designers for many different purposes, and for added versatility may include various types of interfaces for signal input and output.

A typical digital signal processor contains an arithmetic logic unit (ALU), a multiplier, memories to store signals (data) and coefficients, and a section that controls the operation of these blocks and the routing of signals between blocks and into and out of the processor (**Fig. 7**). At each program step, the ALU performs an addition, subtraction, or logic operation, according to the mode selected by the controller. Coefficients are values that do not change as often as signals, such as the gain of a gain block. Higher-performing processors may contain multiple ALUs and multipliers, and other blocks like barrel shifters and hardware accelerators, which are small nonprogrammable sections of the processor executing dedicated operations for higher speed and efficiency.

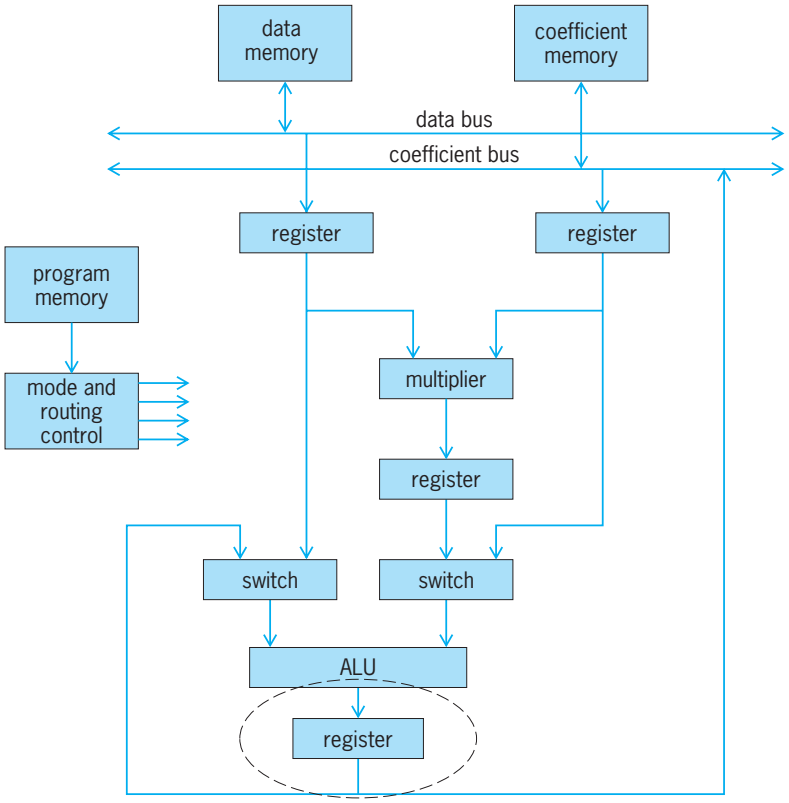


Fig. 7. Programmable digital signal processor.

Generally, a fixed sequence of program instructions is executed during each sample period. More complex programs may alter the performed sequences depending on inputs to the processor or results from previous processing steps. The sequence of executed instructions is varied using special branch instructions that cause the controller to redirect the instruction flow.

Digital signal processing can also be performed in circuits that are not programmable, but rather consist of simple processing operations hardwired together to perform a single processing task. This type of processing is sometimes more efficient to use for signal processing at higher sample rates, approximately 100 MHz and higher.

Another highly integrated device used for digital signal processing is the field-programmable gate array (FPGA). In terms of programmability and speed, it falls between programmable digital signal processors and hardwired processing. In FPGAs, logic gates are connected to create simple signal processing operations, like adders, which in turn can be connected together inside the FPGA to create more complex operations. The configuration of the connections is programmed into the FPGA device.

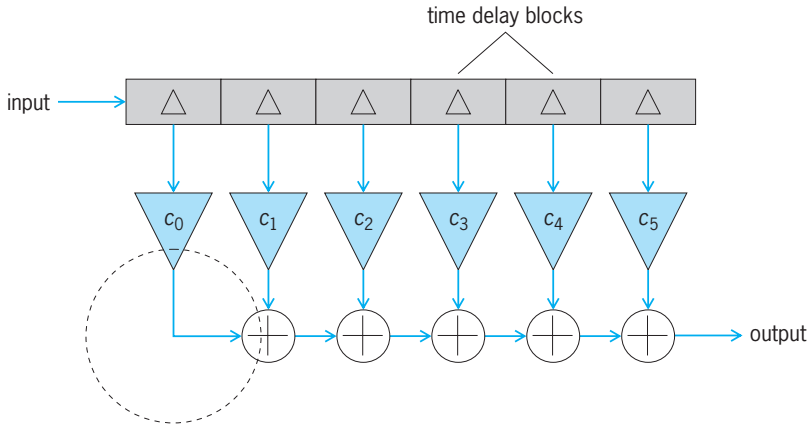


Fig. 8. Finite impulse response filter.

Hardwired, programmable, FPGA, and even analog processing may be integrated into very complex processors. These are developed for very cost sensitive products that are manufactured in high volumes, or that require a very small physical implementation, such as cellular telephones.

Analog-to-digital and digital-to-analog converters. Many applications of digital signal processors involve processing analog signals, such as the signal from a microphone. Analog-to-digital (A/D) and digital-to-analog (D/A) converters provide the interfaces between analog signals and the digital signals inside processors. Converters must sample at the Nyquist rate of the signal being converted.

Analog-to-digital converters must have sufficiently fine quantization to avoid adding excessive noise, and their input signals must usually be analog filtered to avoid a condition called aliasing. If an analog input signal has undesired frequency content above the Nyquist frequency, this content must be removed by analog filtering prior to sampling by the converter. If this antialiasing filtering is not performed, the undesired signal content will alias into the same region of frequencies as the desired content, combining with it in the converter output. When this occurs, it is impossible to separate the desired and undesired content. *See* ANALOG-TO-DIGITAL CONVERTER.

Digital-to-analog converters must usually be followed by an analog filter, called a reconstruction filter. The filter smoothes the quantization steps of the signal. In the case of an audio signal these steps would make the signal sound noisy if filtering was not used. *See* DIGITAL-TO-ANALOG CONVERTER.

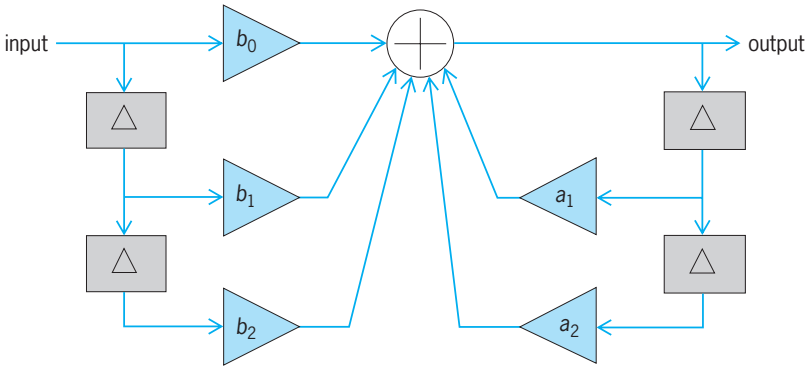


Fig. 9. Infinite impulse response filter.

Types of signal processing blocks. Signal processing systems are built from various types of signal processing blocks. Most can be implemented in analog or digital form, although the digital form is usually more directly related to a block's mathematical model.

The blocks that will be described are examples of the many types of operations used in signal processing systems. They also show how more complex blocks can be constructed from simpler ones.

Linear processing blocks and filters. Addition, scaling, and delay blocks are the most basic linear blocks, and the simplest to implement, especially using digital signal processing. From these, other linear blocks, like filters, can be created.

There are two basic types of linear filters. The first, which does not contain feedback, is the finite impulse response (FIR) filter, also called the transversal filter. Multiple delayed versions of the input signal are scaled and then added together to form the output. The FIR filter in **Fig. 8** adds six versions of the signal, and is therefore a six-tap filter. The input signal is delayed by a series of connected delay blocks, forming a delay line with the needed six taps. The scaling gains c_0 to c_5 are the filter's coefficients. The averager in Fig. 2 is an FIR filter.

The second basic filter type is the infinite impulse response (IIR) filter, which is similar to the FIR filter but contains feedback (**Fig. 9**).

Most analog signal processing filters are IIR and are formed from networks of resistors, capacitors, and inductors. Digital filters are formed directly according to the structures shown in Figs. 8 and 9.

The coefficients of an FIR or IIR filter determine the filter's frequency response, and whether the response is lowpass, highpass, bandpass, bandstop, or has a more complex shape. A filter's coefficients may also be designed to create differentiators and integrators.

If the coefficients of FIR and IIR filters are constant with time, the filters are linear time-invariant, and so can be represented by transfer functions. The transfer function H of the IIR filter in Fig 9, if implemented in digital form, is given in Eq. (5). The

$$H(z) = \frac{b_0 \cdot z^2 + b_1 \cdot z + b_2}{z^2 - a_1 \cdot z - a_2} \tag{5}$$

symbol z is an operator related to frequency, and allows the response of the filter to be conveniently evaluated for any input signal. In the analog domain, similar transfer functions are used, but with the Laplace operator s. *See* ELECTRIC FILTER; LAPLACE TRANSFORM; Z TRANSFORM.

Nonlinear processing blocks. Nonlinear blocks for multiplication, division, squares and higher powers, square roots, and logarithms can all be performed in signal processing. Analog implementations require the use of active components. The digital forms of some of these operations, like multiplication and division, are usually executed in the same way long-hand arithmetic is performed. Others, like

logarithms, use approximations based on polynomial equations (themselves consisting of multiplication and addition operations). Look-up tables may also be used to calculate logarithms, just as they were calculated from tables in books before the advent of slide-rules and calculators. Actually, any nonlinear operation for a block may be approximated using a look-up table.

Modulator blocks process audio, image, and data signals so they can be transmitted from one point to another, as in a radio broadcast, or stored into physical media, like a compact disk. An amplitude modulator of a radio transmitter varies the amplitude of a high-frequency carrier in proportion to the level of the relatively low-frequency audio signal to be transmitted (**Fig. 10**). A demodulator performs the

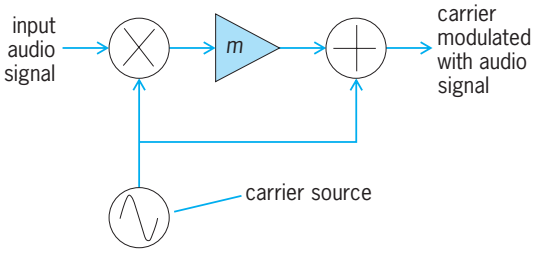


Fig. 10. Amplitude modulator.

opposite operation of its corresponding modulator to recover the original modulated signal. A related block is a frequency translator, which shifts a carrier signal from one frequency to another while preserving the modulation on the carrier. Combining a frequency translator block and a filter block allows signals to be selected, or tuned, from within a frequency space containing many signals, such as all of the radio stations in a city. *See* AMPLITUDE MODULATOR; DEMODULATOR; MODULATION; MODULATOR.

A level detector block outputs an estimate of the level of its input signal. A common type of detector measures the root-mean-square (RMS) level of a signal (**Fig. 11**). It can be used with other blocks to construct automatic gain controllers (AGC) and level compressors. *See* AUTOMATIC GAIN CONTROL (AGC).

Logical blocks. Logic signals can be created in signal processing systems and used to control other signals. For example, in a comparator block an input signal is compared to a threshold to produce a logical 0 or 1 output. A switch block outputs one of its two inputs according to a logic selection input to the block. Powerful signal processing systems can be created from combining logic-based blocks with other types of blocks. An example of this is a neural network, which is capable of complex tasks like pattern recognition in images. *See* NEURAL NETWORK.

Transform blocks. Mathematical transforms can be implemented in signal processing systems. The Fourier transform is very useful in communication and signal analysis systems, and is usually implemented in an efficient digital form called the fast Fourier transform (FFT). *See* FOURIER SERIES AND TRANSFORMS.

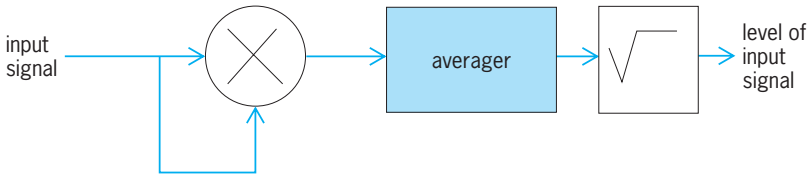


Fig. 11. RMS level detector.

Signal source blocks. Signals may be created for various purposes, such as for the carrier signal in Fig. 10. A sine wave is a common and useful signal, and may be created by designing a block in the form of an IIR filter so that its internal feedback is on the verge of instability, causing the output to vary periodically, or oscillate, in a controlled manner to produce the sine wave. Another method to create a sine wave, and other signals of any arbitrary waveform shape, is a look-up table. Samples of the waveform are precalculated and stored in the table. A counter creates an index to select the table entries in sequence, thus creating the desired waveform (**Fig. 12**). *See* SIGNAL GENERATOR; WAVEFORM GENERATOR.

Examples of signal processing systems. Radio signal processing is a key element of many electronic systems, including cellular telephones, radar, point-to-point radio communication, and transmitters and receivers for AM, FM, television, and more recently satellite broadcast systems. These systems usually employ a combination of analog and digital processing, and consist of many processing blocks. The main blocks in radio systems are frequency translators, filters, modulators, demodulators, AGC and amplifiers, A/D and D/A converters, and other blocks for manipulating data. These blocks appear in the diagram of a cellular telephone in **Fig. 13**. *See* MOBILE RADIO.

Point-to-point communication is used for voice and data signals in military and space applications and for many other purposes. Signal processing on space probes allows large amounts of telemetry data, like images of other planets, to be transmitted to radio receivers on Earth, where more processing converts the data back into pictures and other usable forms. Military radios may include processing to encrypt and decrypt signals. *See* TELEMETERING.

Devices for producing and reproducing music make heavy use of signal processing, especially digital processing. Electronic keyboards contain processing to create multiple sound sources capable

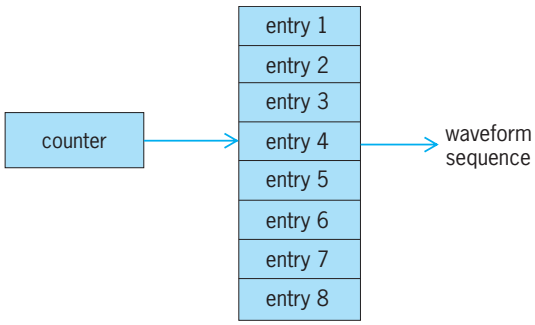


Fig. 12. Signal source using a look-up table.

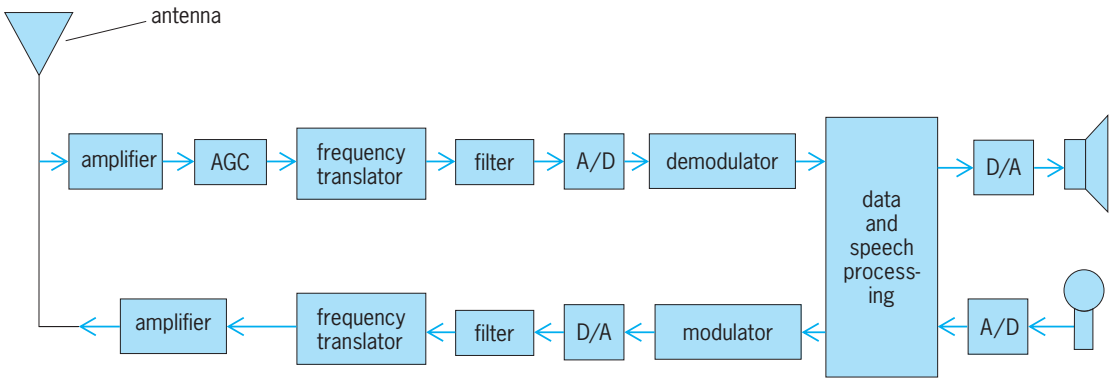


Fig. 13. Cellular telephone block diagram.

of producing many timbres, and they employ additional processing to manipulate and combine these sources into audio outputs. An example of sound manipulation is reverberation, which uses delay and gain blocks arranged into feedback loops to create echoes. Portable MP3 music players employ signal processing to convert highly compressed stored sound data into music output. *See* MUSICAL INSTRUMENTS; SOUND-REINFORCEMENT SYSTEM.

Automated systems, including industrial robots, are commonly used in factories to control machinery and position materials. The physical parts of this type of system, such as a robot arm, are moved by actuators controlled by a signal processing program, with sensors in the system providing feedback to the program regarding the positions and movement of parts of the system. The program, actuators, physical parts, and sensors form a feedback loop that allows the program to finely control the system's movements. *See* ROBOTICS.

A few more examples of the many systems that employ signal processing are medical-imaging devices, sonar, DVD recorders, engine controllers in cars, multichannel sound systems, night-vision goggles, graphics design programs, power plants, and guidance systems for rockets and missiles.

J. William Whikehart

Bibliography. J. J. Carr, *Secrets of RF Circuit Design*, 3d ed., McGraw-Hill, 2001; R. J. Higgins, *Digital Signal Processing in VLSI*, Prentice-Hall, 1990; J. H. McClellan, R. W. Schafer, and M. A. Toder, *DSP First: A Multimedia Approach*, Prentice-Hall, 1997; S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997.

Signal-to-noise ratio

The quantity that measures the relationship between the strength of an information-carrying signal in an electrical communications system and the random fluctuations in amplitude, phase, and frequency superimposed on that signal and collectively referred to as noise. For analog signals, the ratio, denoted S/N, is usually stated in terms of the relative amounts of electrical power contained in the signal and noise. For digital signals the ratio is defined as the amount of energy in the signal per bit of information carried by the signal, relative to the amount of noise power per hertz of signal bandwidth (the noise power spectral density), and is denoted E_b/N_0. Since both signal and noise fluctuate randomly with time, S/N and E_b/N_0 are specified in terms of statistical or time averages of these quantities.

An information signal usually accumulates noise at all stages of its generation, transmission, and reception. The noise contributed by the transmitters and receivers tends to have a power spectral density that is constant over a wide range of frequencies and is thus referred to as white noise, in analogy to the spectral distribution of white light. Furthermore, its level is proportional to the temperature of the various electrical components such as oscillators, amplifiers, filters, and detectors that constitute the signal generation and detection circuitry. Noise of this type, referred to as thermal noise, is caused in large part by the random motion of thermally excited electrons in conducting media. Its power spectral density is given by $N_0 = kT/2$ watts per hertz, where T is the medium's absolute temperature in kelvins and $k = 1.38 \times 10^{-23}$ joule/kelvin is Boltzmann's constant. In contrast, noise accumulated in the transmission path typically derives from external electromagnetic sources such as microwave systems, or radio and television stations. This type of noise is known as colored noise by virtue of the fact that its power spectral density varies with frequency. In most cases the total noise carried by the signal is in part derived from external sources and in part from the transmitter and receiver hardware. *See* BOLTZMANN CONSTANT.

The magnitude of the signal-to-noise ratio in a communications systems is an important factor in how well a receiver can recover the information-carrying signal from its corrupted version and hence how reliably information can be communicated. Generally speaking, for a given value of S/N the performance depends on how the information quantities are encoded into the signal parameters and on the method of recovering them from the received signal. The more complex encoding methods such as phase-shift keying or quadrature amplitude-shift

keying usually result in better performance than simpler schemes such as amplitude- or frequency-shift keying. As an example, a digital communication system operating at a bit error rate of 10^{-5} requires as much as 7 dB less for E_b/N_0 when employing binary phase-shift keying as when using binary amplitude-shift keying. *See* ELECTRICAL COMMUNICATIONS; ELECTRICAL NOISE; INFORMATION THEORY; MODULATION.

Hermann J. Helgert

Bibliography. B. P. Lathi, *Modern Digital and Analog Communication Systems*, 3d ed., Oxford University Press, 1998; M. Schwartz, *Information Transmission, Modulation, and Noise*, 4th ed., McGraw-Hill, 1990.

Signal transduction

The transmission of molecular signals from a cell's exterior to its interior. Molecular signals are transmitted between cells by the secretion of hormones and other chemical factors, which are then picked up by different cells. Sensory signals are also received from the environment, in the form of light, taste, sound, smell, and touch. The ability of an organism to function normally is dependent on all the cells of its different organs communicating effectively with their surroundings. Once a cell picks up a hormonal or sensory signal, it must transmit this information from the surface to the interior parts of the cell—for example, to the nucleus. The cell then uses this information to respond in an appropriate manner.

Information is transmitted from the cell surface to the interior via signal transduction pathways that are very specific, both in their activation and in their downstream actions. Thus, the various organs in the body respond in an appropriate manner and only to relevant signals. *See* CELL (BIOLOGY).

Cell signals. All signals received by cells first interact with specialized proteins in the cells, called receptors, which are very specific to the signals that they receive. Signals received by the cell can be in various forms. The most common are chemical signals, which include all the hormones and neurotransmitters secreted within the body as well as the sensory (external) signals of taste and smell. The internal hormonal signals include steroid and peptide hormones, neurotransmitters, and biogenic amines, all of which are released from specialized cells within the various organs. The external signals of smell, which enter the nasal compartment as gaseous chemicals, are dissolved in liquid and then picked up by specialized receptors. Other external stimuli are first received by specialized receptors (for example, light receptors in the eye and touch receptors in the skin), which then convert the environmental signals into chemical ones, which are then passed on to the brain.

Receptors. Receptors are specialized proteins that perform unique functions, the most important of which are to recognize external signals and convert them into biochemical or physiological functions. Once a receptor has received a signal, it must transmit this information effectively into the cell. This is accomplished either by a series of biochemical changes within the cell or by modifying membrane voltage potential by the movement of ions into or out of the cell. Receptors that initiate biochemical changes can do so either directly via intrinsic enzymatic activities within the receptor or by activating intracellular messenger molecules. Receptors may be broadly classified into four classes: G-protein–coupled receptors, receptor tyrosine kinases, ion channels, and steroid receptors. These types of receptors differ in their mode of action and in the molecules that activate them (**Fig. 1**).

G-protein–coupled receptors. The largest family of receptors is the G-protein–coupled receptors (GPCRs), dependent on guanosine triphosphate (GTP) for their function. Many neurotransmitters, hormones, and small molecules bind to and activate specific G-protein–coupled receptors. The G-protein–coupled receptor is located at the cell membrane; it is made up of an extracellular ligand-binding domain, seven

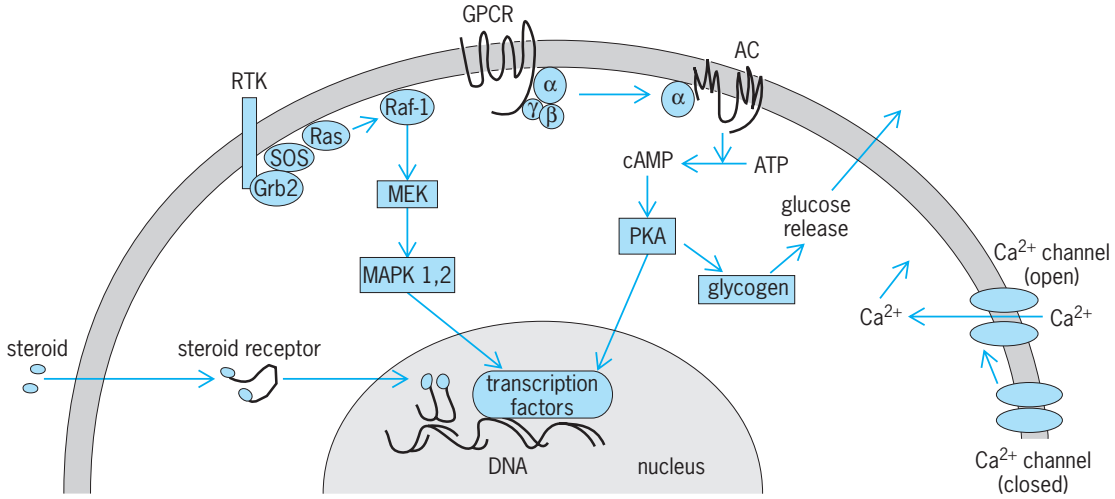


Fig. 1. Different receptors and their mechanism of action: G-protein–coupled receptor (GPCR), receptor tyrosine kinase (RTK), adenylyl cyclase (AC).

membrane-spanning domains, and an intracellular domain. In the nonstimulated state the receptor is bound within the cell to a complex of three different proteins (a heterotrimer), named alpha (α), beta (β), and gamma (γ), which together constitute the G-protein. The α subunit is also bound to guanosine diphosphate (GDP) in the inactive stage. Upon activation of the receptor, the α subunit binds GTP, the heterotrimer disassociates from the receptor, and the α subunit dissociates from the β and γ subunits. The activated α and, $\beta\gamma$ subunits (which are bound together) then bind to and regulate their effector molecules. Once the signal has been passed on, the α subunit hydrolyzes the bound GTP to GDP, which makes the α subunit inactive, and it then binds the $\beta\gamma$ subunits. At this point, the subunit is back to its resting state and capable of being activated again by the receptor.

Receptor tyrosine kinases. A second family of membrane-bound receptors is the receptor tyrosine kinases (RTKs). They function by phosphorylating themselves and recruiting downstream signaling components. Upon activation the receptor becomes phosphorylated on tyrosine residues, allowing the receptor to engage intracellular pathways.

Ion channels. These proteins form channels that open upon activation, thereby allowing the passage of ions across the membrane. Ion channels are responsive to either ligands or to voltage changes across the membrane, depending on the type of channel. The movement of ions changes the membrane potential, which in turn changes cellular function.

Steroid receptors. Steroid receptors are not located at the cell membrane but are found within the cell. They bind cell-permeable molecules such as steroids, thyroid hormone, and vitamin D, all of which are lipophilic and can thus pass through the cell membrane into the cell. Once these receptors are activated by ligand, they translocate to the nucleus, where they bind specific deoxyribonucleic acid (DNA) sequences to modulate gene expression. *See* STEROID.

Signal flow within cells. Once the receptor has received its signal, it must effectively propagate this message to the cell. The intracellular component of signal propagation, also known as signal transduction, is receptor-specific. A given receptor will activate only very specific sets of downstream signaling components, thereby maintaining the specificity of the incoming signal inside the cell. In addition to maintaining specificity, signal transduction pathways amplify the incoming signal. Signal amplification is achieved by a signaling cascade, in which molecule A activates several molecule B's, which in turn activate several molecule C's. In this way the activation of a few receptors eventually leads to the activation of numerous signaling molecules, resulting in an appropriate physiological response by the cell.

Signaling by receptors. The signaling cascade is initiated at the level of the receptor. The various receptors activate different signaling molecules and cascades. Activation of G-protein–coupled receptors results in Gα and G$\beta\gamma$ subunit activation. Gα subunits can activate or inhibit signaling molecules, depending on the Gα subtype that is bound to the G-protein–coupled receptor. Gαs stimulates adenylyl cyclases (AC) to produce cyclic adenosine monophosphate (cAMP), whereas Gαi inhibits adenylyl cyclases and the production of cAMP. Thus, an additional level of connection specificity is introduced in that different G-protein–coupled receptors bind to specific Gα subunits. Once the Gα subunit is ON (activated), it uses adenylyl cyclases to amplify its signal. The Gα subunit has a built-in OFF (inactivated) mechanism (the hydrolysis of bound GTP to GDP) which autoregulates the duration of the signal.

Initiation of the signaling cascade by receptors that have intrinsic enzyme activities, such as receptor tyrosine kinases, proceeds in a different manner. Upon binding hormone, the receptors become enzymatically competent and activate downstream signaling components. These activated messenger components then activate signaling components farther downstream, bringing into play a signaling cascade. The basic design here is that of a "bucket brigade" (**Figs. 2** and **3**). Ion channels function in yet another way to initiate signaling cascades. The receptors themselves are channels that can open and close to permit the movement of ions across the cell membrane. Once the ion channels are activated they undergo a change in their structural conformation, allowing ions to pass into the cell. When activated the channel opens, allowing an ion such as calcium (Ca^{2+}) into the cell. These ions can then trigger the activation of signaling pathways.

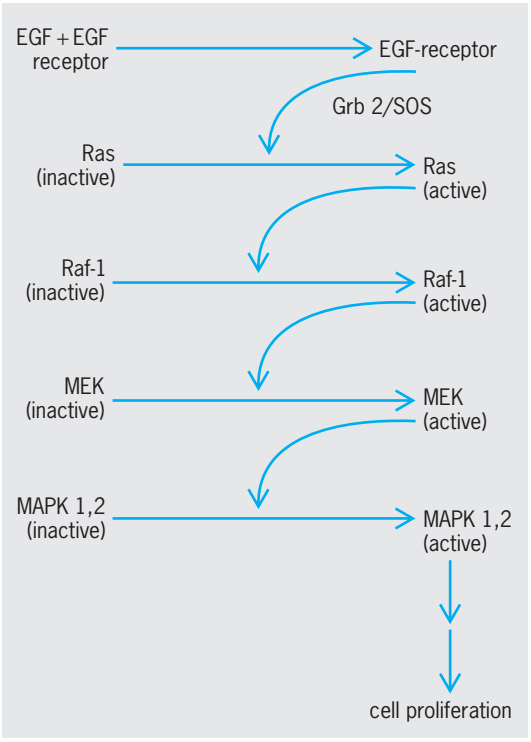


Fig. 2. Steps in the activation of the MAPK pathway leading to cell proliferation.

Signaling by small molecules. Several small molecules within the cell act as intracellular messengers. These include cAMP, cyclic guanosine monophosphate (cGMP), nitric oxide (NO), and Ca^{2+} ions. Increased levels of Ca^{2+} in the cell can trigger several changes, including activation of signaling pathways, changes in cell contraction and motility, or secretion of hormones or other factors, depending on the cell type. Increased levels of nitric oxide cause relaxation of smooth muscle cells and vasodilation by increasing cGMP levels within the cell. Increasing cAMP levels can modulate signaling pathways by activating the enzyme protein kinase A (PKA).

Signal transduction cascades. Signals often flow within the cell as cascades. As the signal is transmitted downstream from one signaling molecule to another, it is amplified. Two of the most well-defined cascades are the mitogen-activated protein kinase (MAPK) pathway leading to cell proliferation, and the glucose 6-phosphate (G6P) pathway leading to the metabolism of glycogen (Fig. 2).

In the case of the MAPK pathway, extracellular growth factors such as epidermal growth factor (EGF) bind to and activate the EGF receptor. Activation of the receptor causes autophosphorylation of the receptor and recruitment of signaling components Grb2 and SOS. Once these components are recruited to the membrane, they activate Ras, a G protein, by promoting the exchange of GDP to GTP. Ras-GTP binds and activates Raf-1, which then phosphorylates and activates mitogen-activated protein or extracellular signal-regulated kinase kinase (MEK). Activated MEK in turn phosphorylates and activates MAPK 1,2. Activated MAPK 1,2 is capable of phosphorylating and activating a number of proteins that are involved in transcriptional and translational functions. Several of the transcription factors activated by MAPK modulate genes that are important in regulating the cell cycle, thereby promoting cell division and proliferation.

The glucose 6-phosphate pathway works in a manner similar to the MAPK pathway except that one of the signaling components is the molecule cAMP. In the case of the glucose 6-phosphate pathway, glucagon binds to and activates the glucagon receptor. The activated glucagon receptor, which is coupled to a G protein, activates its G protein. The GTP-bound activated Gα subunit binds and activates adenylyl cyclase, which promotes the production of cAMP from ATP. Increased levels of cAMP activate cAMP-dependent protein kinase (PKA). Activated PKA phosphorylates and activates the enzyme phosphorylase kinase, which activates the enzyme glycogen-phosphorylase A. Activated phosphorylase A phosphorylates glycogen, leading to its breakdown to glucose, which is then released from the cell. In addition to this metabolic function, PKA can regulate gene expression (Fig. 1).

Regulation of physiological function. One of the most important functions of cell signaling is to control and maintain normal physiological balance within the body. Activation of different signaling pathways leads to diverse physiological responses (Fig. 2). For example, activation of the MAPK pathway results in cell proliferation, whereas activation of the cAMP pathway results in the release of glucose. In this manner all the normal physiological functions such as cell proliferation, death, differentiation, and metabolism are maintained. Signaling pathways in cells play more than the role of a simple transmitter. Several of the signaling pathways interact with each other and can serve as signal integrators. This added level of complexity is necessary in order to maintain normal physiology. An average cell in the body is constantly exposed to several different factors and hormones; therefore the final physiological outcome is dependent on the signal integration capabilities of the cell. Signal integration can occur in a number of ways: negative and positive feedback loops in pathways can modulate signals within a pathway; positive interactions between two signaling pathways can increase duration of signals; and negative interactions between pathways can block signals. Therefore, signaling pathways in cells play a vital role in maintaining physiological function. *See* CELL NUCLEUS; CELL ORGANIZATION; ENDOCRINE SYSTEM (VERTEBRATE); NORADRENERGIC SYSTEM.

Prahlad T. Ram; Ravi Iyengar

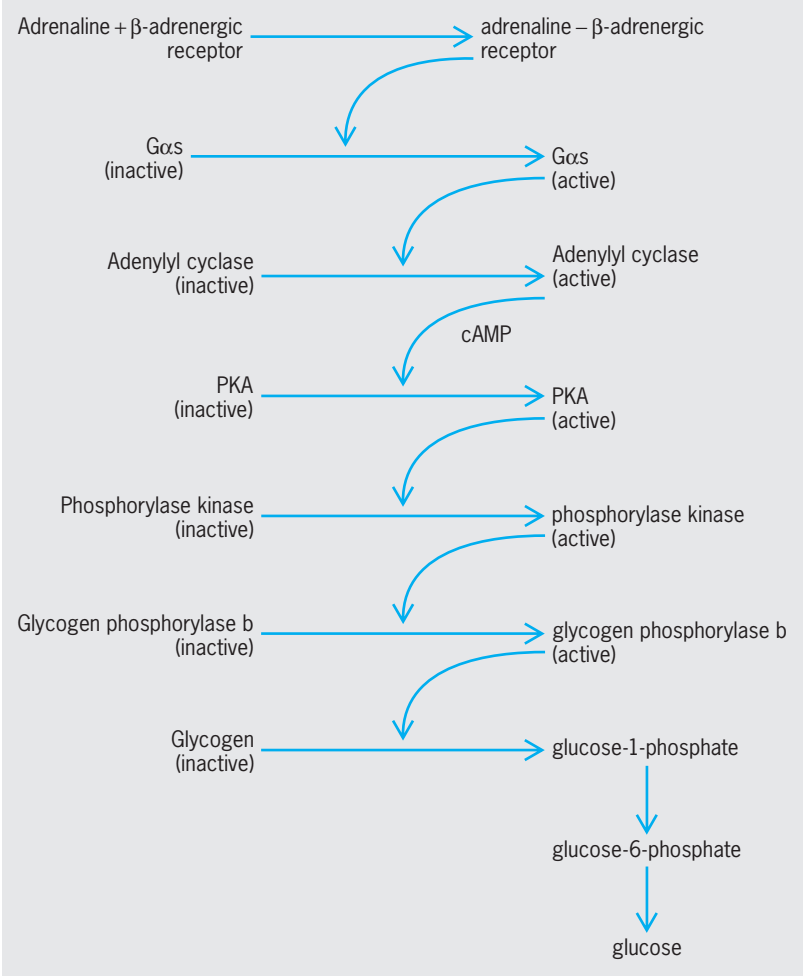


Fig. 3. Steps in the glucagon-stimulated release of glucose stores from cells.

Bibliography. H. E. Hamm, The many faces of G protein signaling, *J. Biol. Chem.*, 273:669–672, 1998; T. Hunter, Signaling: 2000 and beyond, *Cell*, 100:113–127, 2000; R. Iyengar and L. Birnbaumer (eds.), *G Proteins*, Academic Press, New York, 1989; J. D. Jordan and R. Iyengar, Modes of interactions between signaling pathways, *Biochem. Pharmacol.*, 55:1347–1352, 1998.

Significant figures

Digits that show the number of units in a measurement expressed in decimal notation.

Scientific notation is useful in showing which digits are significant. In scientific notation, a number is expressed as the product of a number 1 to 10 and a power of 10, or the product of 1 and a power of 10. Thus, the number 123,000 is 1.23×10^5 (see **table**).

The precision of a measurement is based on the size of the unit of measurement. The smaller the unit, the more precise is the measurement.

All the numbers expressing measurements are approximate. Care must be exercised not to increase the precision of a measurement simply by affixing zeros. For example, a measurement of 14.2 g, indicating precision to the nearest tenth, should not be written as 14.20 because this indicates precision to hundredths.

The absolute error in a measurement is one-half the unit. For example, the error in 480 m, measured to the nearest 10 m, is one-half of 10 m, or 5 m. The relative error, usually expressed as a percent, is the ratio of the absolute error to the measurement. The relative error for 480 m, measured to the nearest 10 m, is 5/480, or approximately 1%. *See* PERCENT.

Computations cannot improve the precision of the measurement. To add measures, they should all be rounded to the unit of the least precise measurement. The sum 8.6 cm + 0.14 cm + 2.75 cm is found by rounding each to tenths: 8.6 cm + 0.1 cm + 2.8 cm = 11.5 cm. Even by doing this, the absolute error might be as large as 0.05 + 0.005 + 0.005 or 0.06, and affect the result by as much as 0.1.

In multiplying and dividing approximate numbers, the product or quotient is rounded to the number of significant digits in the number with the fewest significant digits. For example, 6.2 m × 8.75 m by computation is 54.25. The product needs to be rounded to 54 m^2 so as to show two significant digits. *See* NUMBERING SYSTEMS; NUMERICAL ANALYSIS; STATISTICS.

Joseph N. Payne

Examples of significant digits and scientific notation

Measurement	Unit of measurement	Significant digits	Scientific notation
480 m	10 m	4 and 8	4.8×10^2
37.53 m	0.01 m	3, 7, 5, and 3	3.753×10^1
14.2 g	0.1 g	1, 4, and 2	1.42×10^1
193,000,000 mi	1,000,000 mi	1, 9, and 3	1.93×10^8
270.0 kg	0.1 kg	2, 7, 0, and 0	2.700×10^2

Bibliography. L. Elden, *Numerical Analysis: An Introduction,* 1990.

Silica minerals

Silica (SiO_2) occurs naturally in at least nine different varieties (polymorphs), which include tridymite (high-, middle-, and low-temperature forms), cristobalite (high- and low-temperature forms), coesite, and stishovite, in addition to high (β) and low (α) quartz. These forms have distinctive crystallography, optical characteristics, physical properties, pressure-temperature stability ranges, and occurrences (see **table**).

The transformation between the various forms are of two types. Displacive transformations, such as inversions between high-temperature (β) and low-temperature (α) forms, result in a displacement or change in bond direction but involve no breakage of existing bonds between silicon and oxygen atoms. These transformations take place rapidly over a small temperature interval and are reversible. Reconstructive transformations, in contrast, involve disruption of existing bonds and subsequent formation of new ones. These changes are sluggish, thereby permitting a species to exist metastably outside its defined pressure-temperature stability field. Two examples

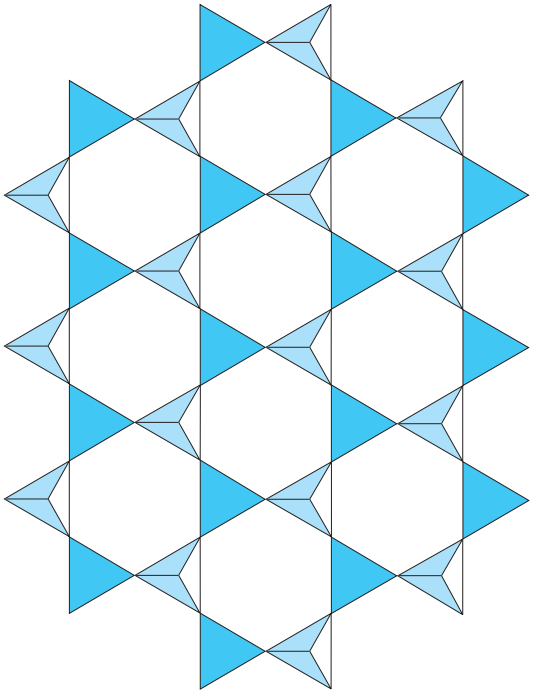


Fig. 1. Portion of an idealized sheet of tetrahedrally coordinated silicon atoms similar to that found in tridymite and cristobalite. Sharing of apical oxygens (which point in alternate directions) between silicons in adjacent sheets generates a continuous framework. (*After J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, Rev. Geophys. Space Phys., 14:37–80, 1976*)

Optical and physical properties of silica minerals

Mineral	Specific gravity	Optic sign	Mean index	Nα or Nω*	N*	Nγ or N*	Cleavage	Color	Hardness (Mohs scale)	Twinning
Stishovite	4.35	Uniaxial (+)	1.81	1.799		1.826	N.R.†	Pale gray	N.R.	N.R.
Coesite	3.01	Biaxial (+)	1.59	1.594	1.596	1.597	None	Colorless	8	{021}
Low (α) quartz	2.65	Uniaxial (+)	1.55	1.544		1.553	{1011}, {0111}, {1010}‡	Colorless, white to variable	7	*c* axis {1120}, {1122}, and others
Low (α) cristobalite	2.32	Uniaxial (−)	1.48	1.487		1.484	None	Colorless, white, or yellowish	6½	{111}
Low (α) tridymite	2.26	Biaxial (+)	1.47	1.471–1.479	1.472–1.48	1.474–1.483	None to poor prismatic	Colorless, white	7	{110}

*Nα, Nβ, and Nγ are the three principal refractive indices for biaxial crystals; Nω and Nε are the principal refractive indices for uniaxial crystals.
†N.R. = not reported.
‡Cleavages not usually observed so that quartz is frequently described as having no cleavage.

of reconstructive transformations are tridymite ⇌ quartz and quartz ⇌ stishovite.

Crystal structure. The crystal structures of all silica polymorphs except stishovite contain silicon atoms surrounded by four oxygens, thus producing tetrahedral coordination polyhedra. Each oxygen is bonded to two silicons, creating an electrically neutral framework. Stishovite differs from the other silica minerals in having silicon atoms surrounded by six oxygens (octahedral coordination). *See* CRYSTAL STRUCTURE.

Ideal high tridymite is composed of sheets of SiO_4 tetrahedra oriented perpendicular to the *c* crystallographic axis (**Fig. 1**) with adjacent tetrahedra in these sheets pointing in opposite directions. The apical oxygens of the tetrahedra are bonded to silicons in neighboring sheets, thus generating a continuous framework with hexagonal ($P6_3/mmc$) symmetry. Naturally occurring meteoritic high tridymite deviates somewhat from the ideal structure because adjacent sheets are slightly offset, producing orthorhombic ($C222_1$) symmetry (**Fig. 2**). Low tridymite has a similar structure, but displacive transformations alter the geometry of the sheets, producing a lower symmetry. Terrestrial low tridymite contains three types of sheets in a complex stacking sequence resulting in triclinic (pseudoorthorhombic) symmetry. Meteoritic low tridymite has a monoclinic (Cc) cell caused by a distortion of the hexagonal rings shown in Fig. 1. Little is known about the structure of middle tridymite, but it is believed to be hexagonal.

High cristobalite, like tridymite, is composed of parallel sheets of SiO_4 tetrahedra with neighboring tetrahedra pointing in opposite directions. However, the hexagonal rings are distorted and adjacent sheets are rotated 60° with respect to one another, resulting in the geometry shown in **Fig. 3**. The structure is cubic ($Fd\bar{3}m$), with the layers of tetrahedra parallel to (111). Further distortion of these sheets at low temperature causes an inversion to tetragonal ($P4_12_12$ or $P4_32_12$) low crystobalite.

Coesite also contains silicon atoms tetrahedrally coordinated by oxygen. These polyhedra share corners to form chains composed of four-membered rings. Bonding between chains creates a continuous framework of monoclinic symmetry (C2/c) in which each oxygen is shared between two silicons. The structure, part of which is shown in **Fig. 4**, is significantly more dense than tridymite and cristobalite.

Silicon in stishovite is octahedrally coordinated by oxygen (**Fig. 5**). These coordination polyhedra share edges and corners to form chains of octahedra parallel to the *c* crystallographic axis. The resultant structure is tetragonal ($P4_2/mnm$) and is similar to the structure of rutile (TiO_2).

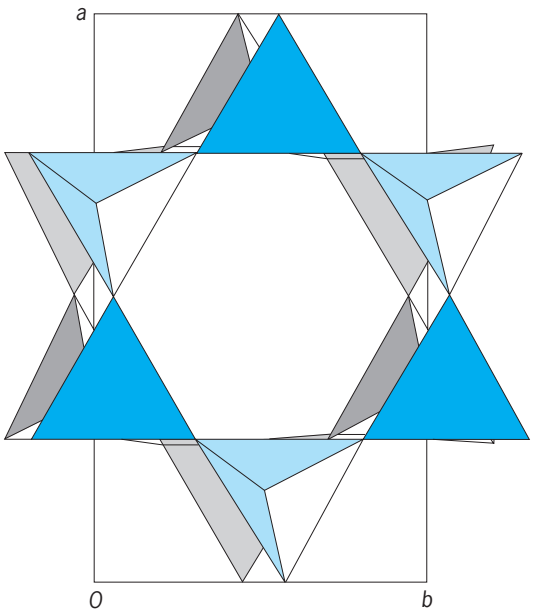


Fig. 2. Portion of the orthorhombic high tridymite structure viewed down the *c* axis showing the slight displacement between adjacent sheets. In ideal hexagonal high tridymite, neighboring sheets would be superimposed. (*After J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, Rev. Geophys. Space Phys., 14:37–80, 1976*)

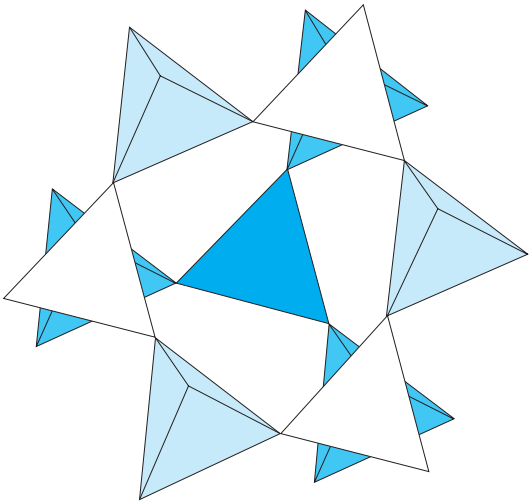


Fig. 3. Portion of cubic high cristobalite illustrating the distortion of tetrahedral sheets, which are oriented parallel to (111), and the 607 rotation of adjacent sheets. (*After J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, Rev. Geophys. Space Phys., 14:37–80, 1976*)

Properties. The physical, optical, and chemical characteristics of the silica polymorphs vary in relation to their crystal structures. Stishovite, which has the most dense packing of constituent atoms, has the highest specific gravity ($G = 4.35$), whereas tridymite and cristobalite, which have relatively open structures, are the silica polymorphs with the lowest specific gravities ($G_{Tr} = 2.26$, $G_{Cr} = 2.32$). In all instances the β forms of the polymorphs have less densely packed structures, hence lower specific gravities, than α counterparts. The refractive indices of the polymorphs are also related to crystal structure, and the mean refractive index is proportional to specific gravity, ranging from 1.47 for tridymite to 1.81 for stishovite. Specific optical and physical properties for the various silica minerals are given in the table. Twinning is common in most of these minerals and, as in the case of quartz, frequently results from the inversion of a high-temperature, high-symmetry form to a low-temperature, low-symmetry form.

Chemically, all silica polymorphs are ideally 100% SiO_2. However, unlike quartz which commonly contains few impurities, the compositions of tridymite and cristobalite generally deviate significantly from pure silica. This usually occurs because of a coupled substitution in which a trivalent ion such as Al^{3+} or Fe^{3+} substitutes for Si^{4+}, with electrical neutrality being maintained by monovalent or divalent cations occupying interstices in the relatively open structures of these two minerals. Such substitutions may decrease the SiO_2 content to as little as 95 wt %.

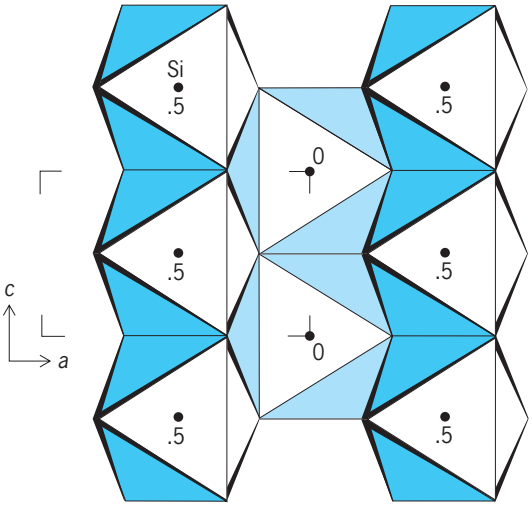


Fig. 5. Octahedral oxygen coordination polyhedra surrounding central silicons in the stishovite structure. The numbers (0, 0.5) indicate the relative height of the silicon atoms on the *a* crystallographic axis. The chains of octahedra formed by sharing polyhedral edges run parallel to the *c* axis. Chains at different levels are interconnected by silicon-oxygen bonds. (*After J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, Rev. Geophys. Space Phys., 14:37–80, 1976*)

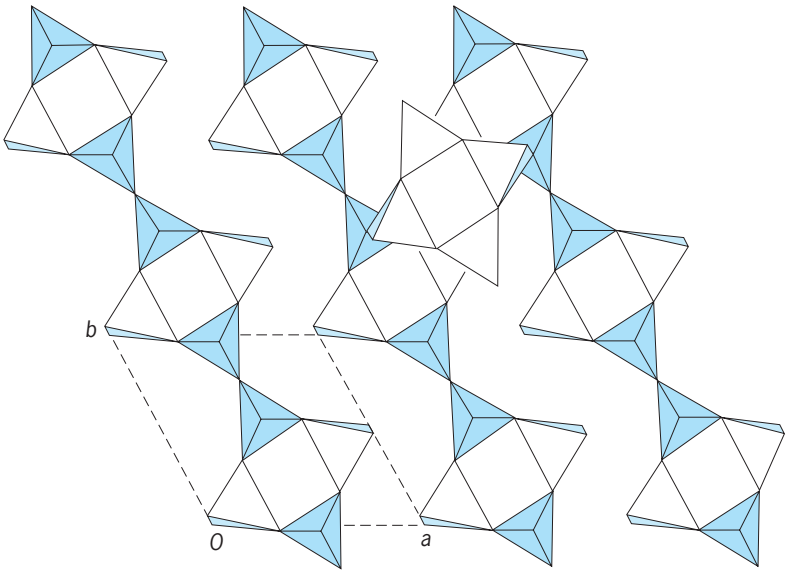


Fig. 4. Four-membered tetrahedral rings in coesite as seen in the plane perpendicular to the twofold axis of symmetry. Bonding between rings in this plane and adjacent planes (as indicated by the single ring at an elevated level) produces the dense coesite structure. (*After J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, Rev. Geophys. Space Phys., 14:37–80, 1976*)

Phase relationships and occurrences. The pressure-temperature stability fields for the silica minerals are shown in **Fig. 6**. Low quartz is the stable form at low temperatures and pressures. At 1 atm (10^2 kilopascals) high quartz is stable from 573 to 870°C (1063 to 1600°F), at which temperature tridymite becomes stable. The stability field for cristobalite ranges from 1470 to 1720°C (2680 to 3130°F), at which temperature SiO_2 melts. Coesite and stishovite are stable only at exceedingly high pressures. However, because of the sluggishness with which reconstructive changes take place, various polymorphs can exist and occasionally crystallize metastably outside their defined limits of stability. Thus high tridymite (β_2) can persist to 163°C (325°F), at which temperature it transforms displacively to middle tridymite (β_1), which exists metastably to 117°C (243°F), at which

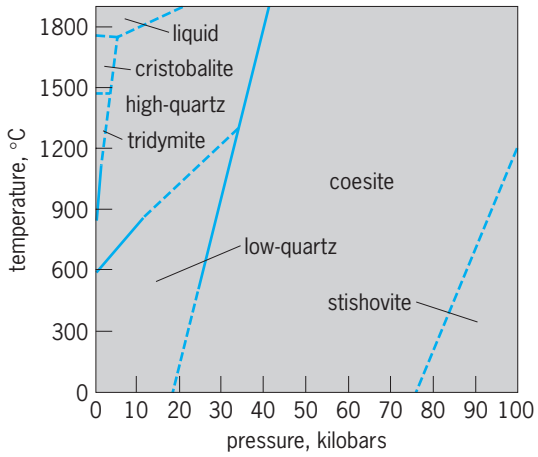


Fig. 6. Stability field of the silica polymorphs. °F = (°C × 1.8) + 32.1 kilobar = 10^2 MPa. (*After C. Klein and C. S. Hurlbut, Jr., Manual of Mineralogy, 21st ed., John Wiley and Sons, 1993*)

low (α) tridymite forms. Impurities or disordering of the stacking sequence can cause these transformation temperatures to vary and possibly merge into a single high/low transformation. High cristobalite can exist metastably below 1450 to 268°C (or 2640 to 514°F; or lower), at which temperature it transforms displacively to low cristobalite. Variation in the temperature of inversion is caused by compositional variability and stacking disorder.

Tridymite and cristobalite crystallize as primary minerals in siliceous volcanic rocks such as rhyolites, trachytes, and andesites in which both minerals may occur in vesicles (cavities) or in the groundmass. Less frequently they are found in basaltic igneous rocks. They have been identified in siliceous sedimentary rocks, such as sandstones and arkoses, subjected to high temperatures during contact thermal metamorphism. Both minerals also occur in the silicate portions of meteorites and in some lunar igneous rocks. Cristobalite, and to a much lesser extent tridymite, is a common intermediate product in the transformation of amorphous biogenic silica (opal A) in marine sediments to quartz during the process of diagenesis (the lithification and alteration of unconsolidated sediments at low temperatures and pressures). The diagenetic sequence is as follows: opal A → opal CT (= porcellanite = cristobalite containing nonessential water in which the stacking sequence of silica sheets is disordered) → chalcedony (microcrystalline quartz) → quartz.

Naturally occurring coesite and stishovite were both first discovered in the Coconino sandstone adjacent to Meteor Crater, Arizona. These minerals are considered mineralogical indicators of the high pressures associated with the shock wave that resulted from the crater-forming meteorite impact. Since their discovery at Meteor Crater, they have both been found in the highly shocked rocks surrounding Ries Crater in Germany. Coesite has also been identified in rocks associated with several other impact or suspected impact craters, in tektites, in material ejected from human-caused explosion craters, and as inclusions in diamonds from kimberlite originating at great depth (hence at great pressure) within the Earth. Stishovite and coesite, therefore, are of great geological significance as indicators of very high pressure due either to depth of formation within the Earth or to shock processes such as those occurring during meteorite impact. *See* COESITE; QUARTZ; STISHOVITE.

John C. Drake

Bibliography. W. H. Blackburn and W. H. Dennen, *Principles of Mineralogy*, 2d ed., 1993; W. A. Deer, R. A. Howie, and J. Zussman, *Rock Forming Minerals*, vol. 4: *Framework Silicates*, 1963; C. Frondel, *Dana's The System of Mineralogy*, 7th ed., vol. 3: *Silica Minerals*, 1962; K. Frye (ed.), *The Encyclopedia of Mineralogy*, 1982; J. J. Papike and M. Cameron, Crystal chemistry of silicate minerals of geophysical interest, *Rev. Geophys. Space Phys.*, 14:37–80, 1976; D. Stöffler, Coesite and stishovite in shocked crystalline rocks, *J. Geophys. Res.*, 76:5474–5488, 1971.

Silicate minerals

All silicates are built of a fundamental structural unit, the so-called SiO_4 tetrahedron. The crystal structure may be based on isolated SiO_4 groups or, since each of the four oxygen ions can bond to either one or two silicon (Si) ions, on SiO_4 groups shared to form complex isolated groups or indefinitely extending chains, sheets, or three-dimensional networks. Mixed structures in which more than one type of shared tetrahedra are present are also known. *See* SILICON.

Classification. Silicates are classified according to the nature of the sharing mechanism, as revealed by x-ray diffraction study. An abbreviated form of such a classification is given in the **table**. The sharing mechanism gives rise to a characteristic ratio of Si to O, but it is possible for oxygen ions that are not bonded to Si to be present in the structure, and sometimes some or all of any aluminum present must be counted as equivalent to Si. The constitution and classification of the silicates were controversial before the advent of x-ray structure analysis methods in 1912. The silicates were then usually considered as salts of silicic acids, many of them hypothetical, and a chemical classification, such as orthosilicates, metasilicates, and the like, was applied.

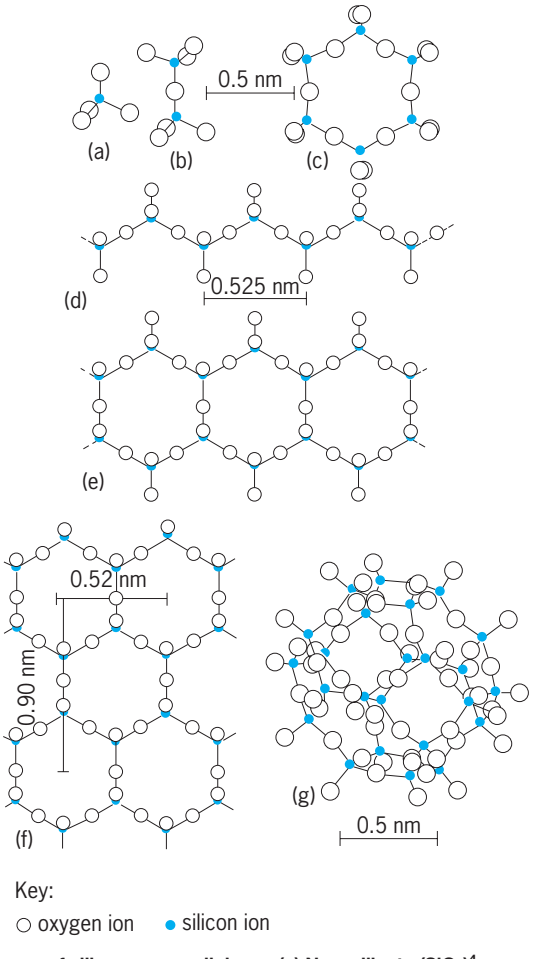
Structure. A dominant feature of the crystal chemistry of the silicates that in large part determines the chemical complexity of these species is the dual role played by aluminum in the crystal structure. The radius ratio of this ion with oxygen is near the critical value between four-coordination and six-coordination, and the aluminum (Al) ion can occur in one or the other or both roles simultaneously. When in four-coordination, the trivalent Al ion substitutes for the quadrivalent Si ion, introducing a valence deficiency of one unit and requiring a concomitant substitution of another cation elsewhere in the crystal structure to provide valence compensation. This mechanism usually involves the coupled substitution of a divalent for a monovalent cation, as of calcium (Ca) for sodium (Na), less frequently of a trivalent

Silicate structures and their characteristic Si/O ratios

Type	Nature of Si-O linkage	Si/O ratio	Examples
Nesosilicates	Isolated SiO_4 groups	1:4	Olivine, garnet
Sorosilicates	Isolated compound groups, Si_2O_7, Si_6O_{18}, and so on	2:7, 6:18, and so on	Thortveitite, beryl
Inosilicates	One-dimensional extended chains and bands	1:3, 4:11, and so on	Amphiboles, pyroxenes
Phyllosilicates	Two-dimensional extended sheets	2:5	Mica, clays, talc, chlorite
Tectosilicates	Three-dimensional networks	1:2	Feldspars, feldspathoids, zeolites

for a divalent ion. Other less common mechanisms involve coupled omissions or substitutions among the anionic units of structure. In some silicates, such as those with a silicon-oxygen framework based on a polymorph of silica, the serial substitution of Al for Si is compensated by the entrance of a cation, such as Na, into vacant interstices of the crystal structure.

The detailed crystallographic and physical properties of the various silicates are broadly related to the type of silicate framework that they possess (see **illus.**). Thus, the phyllosilicates as a group typically have a platy crystal habit, with a cleavage parallel to the plane of layering of the structure, and are optically negative with rather high birefringence. The inosilicates, based on an extended one-dimensional rather than two-dimensional linkage of the SiO_4 tetrahedra, generally form crystals of prismatic habit; if cleavage is present, it will be parallel to the direction of elongation. The tectosilicates commonly are equant in habit, without marked preference for cleavage direction, and tend to have a relatively low birefringence.



Forms of silicon-oxygen linkage. (*a*) Nesosilicate $(SiO_4)^{4-}$. (*b*) Sorosilicate $(Si_2O_7)^{6-}$. (*c*) Sorosilicate, or cyclosilicate $(Si_6O_{18})^{12-}$. (*d*) Inosilicate $(SiO_3)^{2-}$, showing chain structure of pyroxenes. (*e*) Inosilicate $(Si_4O_{11})^{6-}$, showing band structure of amphiboles. (*f*) Phyllosilicate $(Si_2O_5)^{2-}$, showing extended sheets. (*g*) Tectosilicate, showing three-dimensional structure of lazurite. (*After W. L. Bragg, Atomic Structure of Minerals, Cornell, 1937*)

Important minerals. Silicate minerals make up the bulk of the outer crust of the Earth and form in a wide range of geologic environments. Many silicates are of economic importance. Among the clays, feldspars, and refractory minerals, andalusite and wollastonite are used in the ceramic industries, mica as an electrical insulating agent, asbestos and exfoliated vermiculite as thermal-insulating agents, and garnet as an abrasive. Talc is a constituent of facial powder. Other silicates are important as ore minerals, beryllium being obtained from beryl, zirconium and hafnium from zircon, and thorium from thorite. Some silicates such as jadeite and nephrite are prized as ornamental materials, and peridot, garnet, tourmaline, and aquamarine are well-known gemstones. *See* CLAY, COMMERCIAL; CLAY MINERALS; GLASS; SILICATE PHASE EQUILIBRIA.

For discussions of certain silicate mineral groups *See* AMPHIBOLE; ANDALUSITE; CHLORITE; CHLORITOID; EPIDOTE; FELDSPAR; FELDSPATHOID; GARNET; HUMITE; MICA; OLIVINE; PYROXENE; SCAPOLITE; SERPENTINE; ZEOLITE.

Clifford Frondel

Silicate phase equilibria

Silicate phase equilibria studies define the conditions of temperature, composition, and pressure at which silicates can stably coexist. Silicate phase equilibria relations are used by geologists, ceramists, and cement manufacturers to explain the variation of composition of silica-bearing minerals, as well as their number and order of appearance in rocks, slags, glasses, and cements. Silicate phase equilibria relations are also useful to interpret the chemistry of refractories, boiler scale deposits, and welding fluxes.

Silica itself makes up nearly 60% by weight of the Earth's crust. The next most abundant oxides, in decreasing order, are Al_2O_3, CaO, Na_2O, FeO, MgO, K_2O, and Fe_2O_3; all of these occur principally combined with silica as silicates. Free silica and the hundreds of silicate minerals make up nearly 97% of the Earth's crust. The study of silicate phase equilibria was initiated by geologists seeking to apply the phase rule of J. Willard Gibbs to these abundant natural substances. Much of the work since 1907 has been done at the Geophysical Laboratory of the Carnegie Institution of Washington. *See* PHASE EQUILIBRIUM; SILICATE MINERALS.

Silicate phase equilibria are determined in systems generally specified by naming the components involved. Components are the smallest number of independently variable chemical constituents necessary to express the composition of each phase present. A binary system has two components, which may be simple oxides such as Na_2O and SiO_2, or complex compounds such as Na_2SiO_3 and $Na_2Si_2O_5$. In either case, the preferred order of writing the components or oxides follows the rule that the oxides are grouped according to increasing valence of the cation, and then in alphabetical order, as K_2O–Na_2O–CaO–Al_2O_3–SiO_2 for a quinary system and $K_2O \cdot Al_2O_3 \cdot 6SiO_2$ or $KAlSi_3O_8$ for a complex compound.

Equilibrium relations between molten silicate liquids and crystalline silicates, called liquidus equilibria, are most often determined. Equilibria between crystalline silicates in the absence of any liquid silicate melt, called subsolidus equilibria, are determined only infrequently.

Historically, research has been from binary or ternary systems toward more complicated systems. More than 500 systems have been studied to date. The most complicated studies involve limited portions of a system with six components. Specific portions of multicomponent systems are determined because of their importance. For example, in 1913, the first complete silicate phase equilibrium study, the binary system $NaAlSi_3O_8$ [albite (Ab)]–$CaAl_2Si_2O_8$ [anorhite (An)], representing the very abundant rock-forming plagioclase feldspars, was determined by N. L. Bowen, one of the most famous pioneering experimental petrologists. However, this binary system was merely a subsystem from the very complicated quaternary system Na_2O–CaO–Al_2O_3–SiO_2, which has not yet been completely determined.

Silicate crystal structures in general require at least 30% by weight of silica; hence the less siliceous portions of systems are often not determined. The limited number of possible silicate crystal structures determines the possible ratios of oxides to silica in silicates. Certain elements substitute for each other in silicate crystal structures; this reduces the possible varieties of silicates so that those making up the overwhelming bulk of rocks fall into less than a dozen major groups, called rock-forming silicates. Typical of such groups are the feldspars and feldspathoids in which alkali ions substitute for one another and aluminum and ferric iron ions substitute in part for silicon ions, and the olivine and pyroxene groups in which divalent iron, magnesium, manganese, and calcium ions substitute for each other. Most experiments on silicate phase equilibria have been and continue to be on the relations within and between these important groups of minerals. *See* FELDSPAR; FELDSPATHOID; OLIVINE; PYROXENE.

Dynamic and static methods. Both methods are used to determine equilibrium in silicate systems. Dynamic methods like those used for metals and alloys, such as measuring the heat effects of phase changes on heating or cooling, are not particularly satisfactory with silicates. These methods require large samples of silicates that are difficult to prepare, and require also that equilibrium be reached quickly. Silicates in general are slow to react, and supercooling or superheating of hundreds of degrees before reaction occurs is common. Many silicates react sluggishly at temperatures of 1000°C (1830°F) or higher. Crystals in liquid may persist metastably for weeks, nucleation of new phases may require months, and certain crystallographic processes, such as the ordering of aluminum and silicon atoms into different structural sites, may take a lifetime. Subsolidus reactions are extremely sluggish and may require geologic time before equilibrium is attained or even approached. It is this fact that permits the geologist and petrologist to determine the past thermal history of a rock; the higher-temperature products are preserved for study, even through very long cooling intervals, by the slow rate of attaining equilibrium.

Most silicate phase equilibria are determined by the static method of holding a sample under controlled conditions until equilibrium is attained, then quenching the sample for examination. Charges of 10–100 mg of pure components or glass, gel, or crystals of known composition are wrapped in platinum foil and held in air in furnaces in which the temperature is carefully regulated. Controlled partial pressures of oxygen or inert gases are required to study systems with components, such as FeO, MnO, Fe_2O_3, and TiO_2 that are oxidized or reduced by contact with air, or that alloy with platinum. The samples are quenched rapidly to room temperature by dropping them into water or mercury. Most silicate liquids with a ratio of oxygen to aluminum plus silicon of 2:1 or 3:1 are viscous and quench to glass. Crystals grown from liquid during the quench are usually fine-grained and fibrous and are easily recognized. Certain displacive polymorphic transitions in silicate crystals cannot be quenched and are studied by special techniques, but in general the crystals stable at high temperature can be quenched and studied at room temperature. The composition, amount, and number of glassy and crystalline phases are determined by optical microscopy, x-ray diffraction, and thermal analysis.

Establishment of equilibrium. Equilibrium is established when the products obtained by heating a sample to a given temperature are identical with the products of cooling a sample to that temperature; no requirement is made regarding the texture, shape, or grain size of the products. Another criterion used

in recognizing equilibrium is that no change of the sample can be observed after holding the charge at a given temperature for very long periods of time. Particularly useful for this purpose are thin, sharp, crystal fragments that sinter and become rounded on melting, or splinters of glass that devitrify to crystals before larger fragments are affected. When the same products are obtained regardless of starting materials, equilibrium is thought to be attained. At equilibrium the products are usually uniformly distributed in the sample, and there is consistency between the equilibrium results of the system studied and those portions common to other equilibrium systems. Finally, the results of equilibrium are consistent with the trends shown by natural occurrences of the same minerals.

Diagrammatic representation. The use of diagrams to present silicate phase equilibria is customary, since such diagrams express quantitatively the amount and composition of each phase present at any bulk composition in the system at any temperature. Even when molten, silicates have very low vapor pressures, and the vapor pressure and composition of the vapor are practically unaffected by large variations of temperature. Therefore the vapor phase is not considered, and the phase diagrams are drawn isobarically at a total pressure of 1 atm (10^2 kilopascals). However, diagrams involving only a single component are usually drawn with temperature plotted against total pressure.

Binary systems. Binary systems are plotted with composition (usually in percentage by weight) against temperature at constant pressure. An example of such a binary system with complete solid solution without a maximum or minimum is the system $NaAlSi_3O_8$–$CaAl_2Si_2O_8$ (**Fig. 1**). The region labeled liquid plus feldspar is a two-phase field where feldspar, whose composition lies along the lower curve, coexists with liquid, whose composition lies on the upper curve. A line connecting the liquid and feldspar that coexist at a given temperature is called a tie line. The curve separating the liquid field from the liquid-plus-feldspar field is called the liquidus. The curve separating the liquid-plus-feldspar field from the feldspar field is called the solidus.

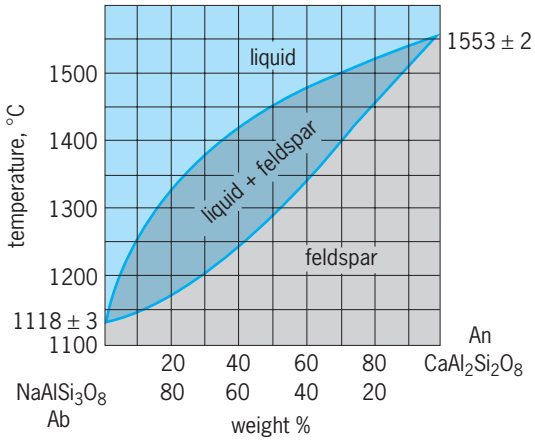


Fig. 1. System $NaAlSi_3O_8$–$CaAl_2Si_2O_8$. °F = (°C × 1.8) + 32. (*After N. L. Bowen, Amer. J. Sci., ser. 4, 35:577–599, 1913*)

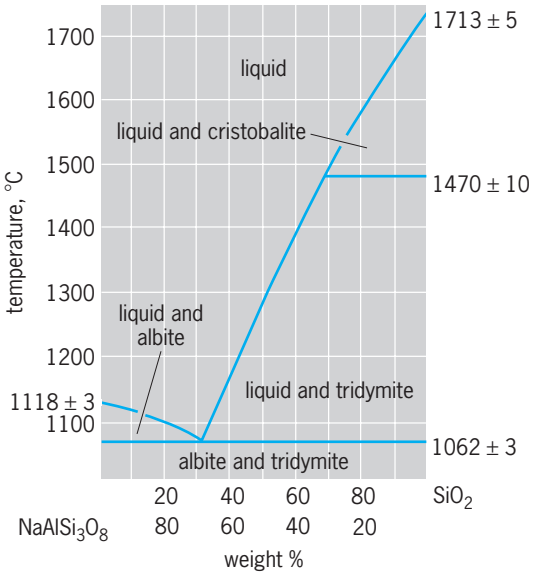


Fig. 2. System $NaAlSi_3O_8$–SiO_2. (*After J. F. Schairer and N. L. Bowen, Amer. J. Sci. 254:161, 1956*)

A composition such as 50% by weight $NaAlSi_3O_8$ and 50% by weight $CaAl_2Si_2O_8$, or more simply $Ab_{50}An_{50}$, is all liquid at temperatures above 1450°C (2640°F). The first crystals to appear in it on cooling would be of composition $Ab_{18}An_{82}$ on reaching the liquidus at 1450°C (2640°F). If equilibrium is maintained during further cooling, these crystals will react continuously with the liquid and change in composition toward $Ab_{50}An_{50}$ along the solidus and at the same time increase in amount. Simultaneously the composition of the liquid will change along the liquidus toward $Ab_{86}An_{14}$. The ratio of liquid to crystals can be found at any temperature by the lever rule, which states that the amounts of the two phases are inversely proportional to the distances of their compositions from the bulk composition. At 1285°C (2345°F), the last liquid of composition $Ab_{86}An_{14}$ will react with crystals and the entire mass will consist of feldspar of $Ab_{50}An_{50}$ composition.

Fractional crystallization occurs when crystals are removed from contact with the liquid for any reason, such as by crystal settling or by protection by armoring of the core by a rim of different composition. The magnitude of the change produced in the remaining liquid can be determined from the diagram if the composition and amount of crystals isolated are known. Thus, the equilibrium diagram can be used to determine the consequences of disequilibrium.

The diagram for the binary system $NaAlSi_3O_8$–SiO_2 shows no solid solution in the crystalline phases (**Fig. 2**). In most systems of this type a solid solution in the crystalline phases does exist, but it is too limited to be observed by ordinary techniques or to be shown on the scale of the diagram.

In the two-phase region of this system, a crystal of composition of the appropriate components exists with liquid until, upon heating, it dissolves in the

liquid at the temperature appropriate for the particular bulk composition chosen. When cooled to the temperature of the eutectic, the first crystal of the composition of the second component appears, and the temperature remains constant until all the liquid has crystallized. The bulk composition is then represented as an aggregate of crystals of both components. Only at the eutectic composition and when the components are pure do crystals melt directly to a liquid without an intervening stage of crystals and liquid. Contrary to the behavior of systems of metals, eutectic mixtures in silicate systems rarely form intimately intergrown aggregates with so-called eutectoid texture. Discrete crystals of both phases appear instead.

Many other types of binary systems are known, and one additional type is described below.

Ternary systems. Ternary systems are plotted in triangular coordinates with the three components at the apexes of an equilateral triangle; temperatures on some surface such as the liquidus are indicated by contours projected onto the diagram, as in the system $KAlSiO_4$–$NaAlSiO_4$–SiO_2 (**Fig. 3**). The fields of primary crystallization of the various solid phases are indicated, as are the compositions and melting temperatures of the end-members of solid solutions, intermediate compounds, and pure components of the system. Binary systems appear as the lines on ternary diagrams; the system $NaAlSi_3O_8$–SiO_2 discussed above appears as a segment of the boundary $NaAlSiO_4$–SiO_2.

Isothermal sections. Ternary systems may also be graphically represented by the use of isothermal sections. Phase relations and compositions are shown in detail at a specified temperature. An isothermal section at 600°C (1110°F) for the system $KAlSiO_4$–$NaAlSiO_4$–SiO_2 is a typical example (**Fig. 4**). More complex systems are represented by tetrahedrons or by various projections.

Pseudobinary system. A pseudobinary system is shown in the temperature-composition diagram for $KAlSi_3O_8$–$NaAlSi_3O_8$ (**Fig. 5**). These components make up the abundant alkali feldspars. The system is pseudobinary; the equilibria represented by the heavy lines are truly binary, but the equilibria represented by the light lines are ternary and indicate a section through regions in which the compositions of the phases do not lie in the plane of the diagram. The ternary nature of this system results from the incongruent melting of solid solutions at and near $KAlSi_3O_8$ composition to leucite, $KAlSi_2O_6$, and a liquid whose composition is indicated by the intersection of the liquidus surface and the prolongation of the line drawn from leucite composition through the bulk composition of the feldspar under consideration. These relations are apparent in the illustration of the system $KAlSiO_4$–$NaAlSiO_4$–SiO_2 (Fig. 3).

The truly binary portion of the system $KAlSi_3O_8$–$NaAlSi_3O_8$ is an example of another major group of binary systems with a minimum in the liquidus and solidus, or with a maximum. The one crystal in the solid solution series at the composition of the minimum melts directly to a liquid without an intervening region of liquid and crystals.

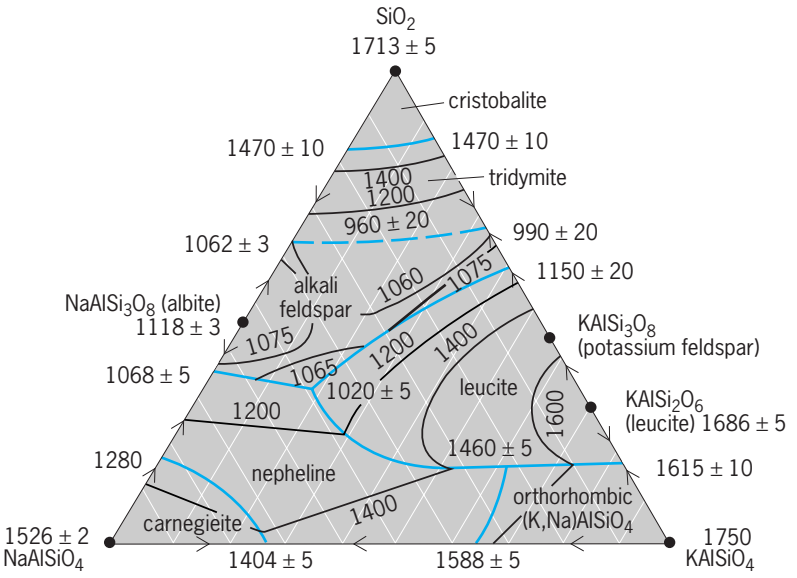


Fig. 3. Ternary system $KAlSiO_4$–$NaAlSiO_4$–SiO_2. The primary phases are indicated, and their primary fields of crystallization are outlined. Arrows indicate direction of falling temperature. Temperatures are given in °C [°F = (°C × 1.8) + 32]. (*After J. F. Schairer, J. Geol., 58:514, 1950*)

The approximate limits of a two-phase subsolidus region are shown, within which a homogeneous alkali feldspar of intermediate composition will separate into two alkali feldspar solid solutions. The curve defining this region is the locus of points representing pairs of compositions that coexist; it is called a solvus.

A subsolidus, nonquenchable polymorphic transition from monoclinic to triclinic symmetry is also shown.

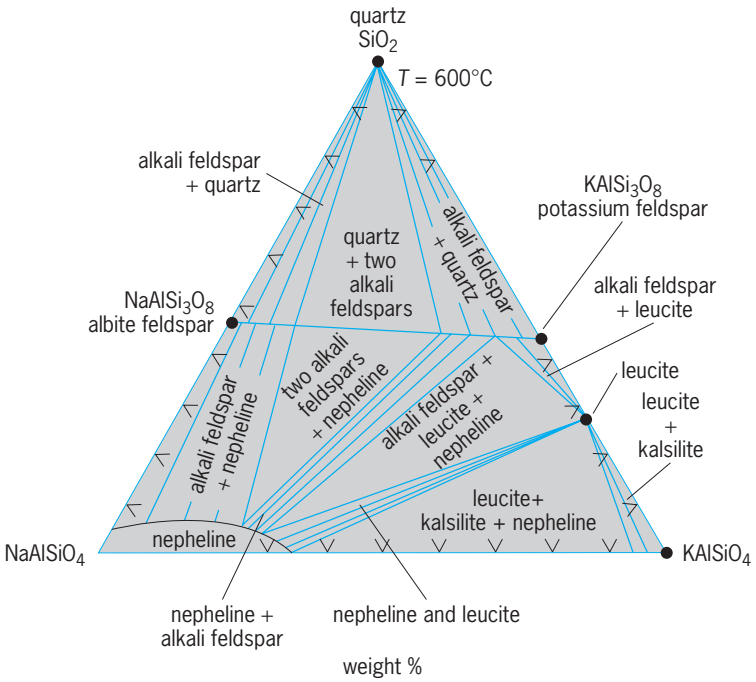


Fig. 4. Isothermal section, in part schematic, showing system, $KAlSiO_4$–$NaAlSiO_4$–SiO_2, at 600°C (1110°F). (*After O. F. Tuttle and J. V. Smith, Amer. J. Sci., 256:587, 1958*)

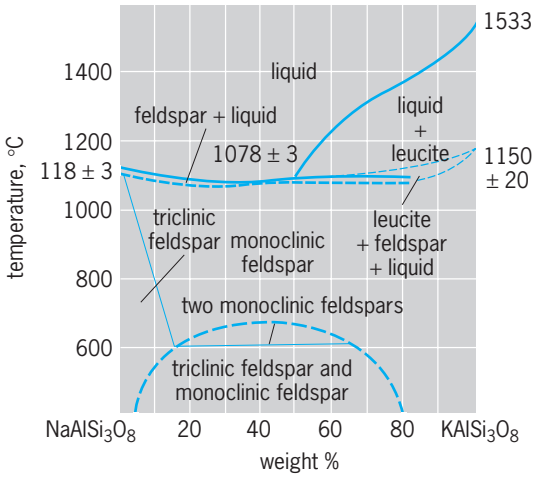


Fig. 5. System $KAlSi_3O_8$–$NaAlSi_3O_8$. Heavy solid or broken curves refer to binary equilibrium. Light solid or broken curves refer to ternary equilibrium. $°F = (°C \times 1.8) + 32$. (*After J. F. Schairer, J. Geol., 58:515, 1950; N. L. Bowen and O. F. Tuttle, J. Geol., 58:497, 1950; W. S. MacKenzie, Amer. J. Sci., Bowen vol., p. 331, 1952*)

Applications. Silicate phase diagrams permit the user to trace the paths of crystallization of a given composition both when equilibrium is maintained and when disequilibrium occurs. These paths place limits on possible relations and thereby restrict interpretation of natural occurrences.

There are several important generalizations which result from the determination of hundreds of phase diagrams for silicates. The most important of these is the generalization that solid solutions are very prevalent among silicates, and wide variations of composition are possible within a given mineral group. An important consequence is that the composition of the crystal or crystals growing from silicate melt changes as the temperature in the system varies. The liquids also change composition during this process and, if liquids are separated from crystals, fractionation and differentiation of silicate magma can proceed very extensively. Furthermore, careful study of the compositions of coexisting mineral phases can, after comparison with experimental results, indicate details of the temperatures and pressures at which the minerals were formed.

The extent of solid solution in a mineral group varies with the temperature and pressure; the limits are indicated by the appearance of another crystalline phase. High temperature favors greater solid solution in general, whereas high pressure decreases the amount of solid solution. The extent of solid solution may be used, in certain cases where the pressure of formation is known or can be approximated, to determine the temperature of formation. The method requires that two minerals with a common component coexist at equilibrium. If they are cooled so that they do not react, then their compositions, by comparison with the appropriate solvus, will indicate the temperature at which the assemblage formed. Solvi are known in the pyroxene, feldspar, nepheline, and other abundant mineral groups, so this method has wide geologic application. *See* GEOLOGIC THERMOMETRY.

Since 1950 a few experimental studies of the distribution of trace elements between coexisting silicates have been made. The results suggest that the minor elements may also be useful to evaluate the temperature and pressure of formation of mineral assemblages.

Another generalization is that there are frequently sharp changes in the number and relative amounts of the phases present as a given composition is cooled from an all-liquid to an all-crystalline state. Other sharp changes occur even in the crystalline state. A great many points have been established, most of which are peritectic (reaction) points rather than eutectic points. A decrease in temperature may cause an assemblage of minerals and liquid to transform completely into other minerals plus liquid, to several new minerals, or to a single mineral of intermediate composition. For example, siliceous liquid and leucite, $KAlSi_2O_6$, react with each other on cooling to form potassium feldspar, $KAlSi_3O_8$, at 1150°C (2100°F).

Phase equilibria studies of aluminosilicates have shown that, when incongruent melting occurs, the liquid formed is richer in silica than the newly formed crystal in most cases. This generalization is of considerable significance in petrogenesis, as it indicates a reason for the progressive enrichment in silica of late magmatic differentiates. *See* MAGMA.

Phase equilibria studies of silicate systems have also shown that the hypothesis of the immiscibility of granitic and basaltic magmas is not useful geologically. Liquid immiscibility has been shown to occur experimentally, but only in compositions that are not found in nature.

The available phase diagrams show that the silicate liquids that crystallize last and at the lower temperatures, either by equilibrium or disequilibrium crystallization, tend to be enriched in ferrous iron, silica, and alkali aluminosilicates with a molecular ratio of alkalies to alumina of unity. Further crystallization tends to remove the ferrous iron, and the last liquid remaining very closely approaches the system $KAlSiO_4$–$NaAlSiO_4$–SiO_2 shown above. This is known as petrogeny's residua system. Two major types of rocks are represented by this system. The most abundant are the granites, represented by the subsystem $KAlSi_3O_8$–$NaAlSi_3O_8$–SiO_2. Analyses of hundreds of granites plot in or close to the low-temperature "valley" on the liquidus of this system, running from the $KAlSi_3O_8$–$NaAlSi_3O_8$ binary toward SiO_2. This constitutes a major reason for believing that most granites originate from silicate magmas.

The second major type of rocks represented are the silica-deficient feldspathoidal syenites of the subsystem $KAlSiO_4$–$NaAlSiO_4$–$KAlSi_3O_8$–$NaAlSi_3O_8$. More experimental detail is required to understand all of the geologic consequences of this system, but there is excellent agreement between synthetic and natural assemblages of phases. A major problem remains in finding the reason why certain magmas differentiate toward the silica-deficient subsystem. The

solution to this problem undoubtedly involves many additional components. It is apparent, however, that the join $KAlSi_3O_8$–$NaAlSi_3O_8$, where binary, is a barrier or hump across which liquids do not move as a consequence of equilibrium crystallization.

The concept of incompatible mineral assemblages is another major consequence of studies in silicate phase equilibria. Nepheline, $NaAlSiO_4$, is incompatible at equilibrium with quartz, SiO_2, as they react to form albite feldspar, $NaAlSi_3O_8$. Similarly leucite, $KAlSi_2O_6$, and quartz are incompatible and react to form potassium feldspar, $KAlSi_3O_8$. Tens of incompatible assemblages are now known to petrologists. The recognition of any of these assemblages in the field indicates unusual conditions in which equilibrium was not attained.

Since 1948, silicate phase equilibria in systems with oxidizable components have been intensively studied to determine the effect of variations of oxygen pressure upon the phase relations. The results show that, when the oxygen partial pressure is changed, the temperature at which minerals containing an oxidizable component can exist is changed. The magnitude of the changes is quite remarkable; a small fraction of an atmosphere of oxygen may change the stability limits of an iron-bearing silicate by as much as 300°C (570°F). This principle, of course, is fundamental to the operation of blast furnaces. In nature, changes of partial pressure of oxygen are expected as gases are expelled from crystallizing magma, and these changes are now believed to affect profoundly the course of differentiation of the magma. The trend in silicate research is toward more detailed study of this effect, particularly as it affects the silica content of the residual liquid.

Much geological field evidence indicates that the temperatures required to obtain silicate liquids in experimental work are hundreds of degrees higher than those determined by various methods to have existed in magmatic assemblages or observed in volcanic eruptions. Silicate phase equilibria studies of increased complexity indicate that, as the number of components is increased, the temperatures required for liquid to form are lowered. The magnitude of this effect is inadequate to explain the geological discrepancy. Geologists have suggested that volatile components, notably H_2O, cause the great lowering of liquidus temperatures. This has indeed been shown to be the case, but experimentally the volatile components must be contained by high external pressure. Pressures of H_2O of 15,000 lb/in.^{2} (103 megapascals) lower the temperatures of the liquidus as much as 700°C (1290°F); simultaneously, several percent of H_2O dissolve in the molten silicate. Reactions are greatly accelerated, and important groups of hydrous minerals become stable. So spectacular have been the results that most current research in silicate phase equilibria for geological applications is conducted under high pressures of volatile components. *See* HIGH-PRESSURE MINERAL SYNTHESIS.

Thermodynamic theory is sufficiently developed to permit silicate phase equilibria to be calculated when values of the necessary physical-chemical parameters are available and ideal solution occurs. Accurate measurements of all these parameters are difficult, and the results of most phase equilibria studies are not precise enough to allow calculation of thermodynamic parameters. Some simple subsolidus and liquidus relations have been rigorously calculated. In general, however, the quality of calculations of complex phenomena is poorer than rough experimental work, the usual difficulties of establishing equilibrium notwithstanding. *See* CEMENT; CERAMICS; GLASS; LITHOSPHERE; PETROLOGY; PORCELAIN; REFRACTORY.

David B. Stewart

Bibliography. F. A. Hummel, *Introduction to Phase Equilibria in Ceramic Systems*, 1984; Y.-H. Li, *A Compendium of Geochemistry*, 2000; B. Mysen and P. Richet, *Silicate Glasses and Melts: Properties and Structure: Developments in Geochemistry*, 2005; R. S. Roth (ed.), *Phase Equilibria Diagrams: Phase Diagrams for Ceramists: Oxides*, 1995; J. V. Walther, *Essentials of Geochemistry*, 2005; S. L. Webb, *Silicate Melts: With 33 Figures: Lecture Notes in Earth Science*, 1997.

Silicate systems

Assemblages of silicate minerals. Studies of the nature of these minerals provide information on their origin and their role in rock formation. This article discusses kinetic processes in silicate systems, oxidation-reduction equilibria in high-temperature systems, and volatile equilibria in systems at high temperatures and pressures. *See* SILICATE MINERALS.

Kinetic Processes

The minerals that form by crystallization in a magma (molten rock) or by recrystallization in a solid rock are determined by phase equilibria and kinetics. Phase equilibria are a measure of which minerals are the most stable for a given temperature, pressure, and bulk composition. In general, minerals with lower free energies form at the expense of minerals with higher free energies. The mineral assemblage with the lowest free energy for a given set of conditions is the most stable and is known as the equilibrium mineral assemblage. Kinetics determines the rate at which chemical reactions occur; it includes such factors as nucleation rates, crystal growth rates, and diffusion rates. The rate at which a rock forms, relative to the kinetics associated with particular reactions, will determine whether the equilibrium mineral assemblage for a given set of conditions is formed. *See* PHASE EQUILIBRIUM.

Metastable equilibrium. An equilibrium mineral assemblage formed at one set of conditions of temperature, pressure, and composition may no longer be the lowest free-energy assemblage if the conditions change. Minerals that persist at conditions where they are not part of the equilibrium assemblage are said to exist metastably. Such metastable assemblages are common in igneous and metamorphic rocks, and they may persist almost indefinitely if the kinetic factors associated with the nucleation and growth

of the stable assemblage are extremely slow. Examples of metastablity include zoned minerals, volcanic glasses, and high-grade metamorphic rocks at room temperature and pressure. Minerals may also form metastably if the nucleation and growth rates of the equilibrium assemblage are slow relative to the nucleation and growth rates of a metastable assemblage. In such situations, minerals will nucleate and grow spontaneously if the total free energy of the products of the reaction is less than that of the reactants; it is not necessary that the products have the lowest possible free energy. An example of metastable crystallization occurs in volcanic rocks when cristobalite or tridymite forms under conditions where quartz has a lower free energy. Although all three minerals have the identical composition, silicon dioxide (SiO_2), tridymite, and cristobalite nucleate much more readily than does quartz because their structures are less complex.

Nucleation. The first step in the formation of a new mineral is the nucleation of seed crystals or nuclei of that mineral. Nucleation of a new crystal may occur on the surface of another crystal or on a plane of disruption within a crystal or melt, in which case the process is known as heterogeneous nucleation. Nucleation that occurs spontaneously within a liquid or solid, independent of local inhomogeneities, is known as homogeneous nucleation. The formation of individual nuclei is controlled by the free-energy change (ΔG) associated with the growth of small particles of that crystal, as in Eq. (1), where

$$\Delta G_{\text{total}} = \Delta G_V + \Delta G_S \qquad (1)$$

ΔG_V = volume free-energy change and ΔG_S = surface free-energy change. Thus this relationship includes the free energy associated with transforming a unit volume of material from the reactants to the products (volume free-energy change) and the free energy associated with the formation of a new crystal interface (surface free-energy change). The surface free-energy term for the growth of a new particle is always positive and is directly proportional to the surface area. The volume free-energy term may be either positive or negative: if it is positive, no nuclei will form; if it is negative, nuclei will grow spontaneously when the decrease in ΔG_V with increasing size exceeds the increase in ΔG_S, so that ΔG_{total} decreases as the nucleus grows. For very small particles, ΔG_{total} increases with increasing radius and, as a result, small nuclei are rapidly resorbed. Particles having a radius smaller than some critical radius r_c can grow only by random collisions. If ΔG_V is negative, however, there is a critical size beyond which ΔG_{total} decreases with increasing radius and nuclei grow spontaneously (**Fig. 1**). The critical radius for heterogeneous nucleation is less than that for homogeneous nucleation, because the ΔG_S term in heterogeneous nucleation is partially offset by the free energy associated with local inhomogeneities.

Systems at equilibrium are said to be saturated in the minerals associated with that equilibrium reaction, but no nucleation will occur at saturation be-

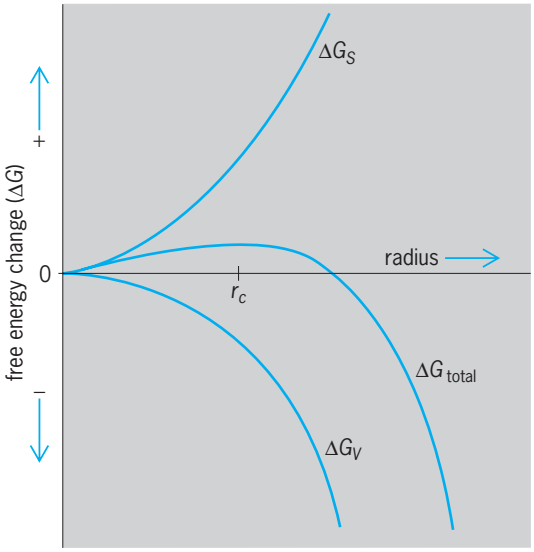


Fig. 1. For growth of spherical nuclei, surface free-energy change (ΔG_S) increases as a function of $4\pi r^2$ (the area of a sphere), while volume free-energy change (ΔG_V) decreases as a function of $^4/_3\pi r^3$ (the volume of a sphere). Nuclei that exceed the critical radius (r_c) will grow spontaneously; nuclei smaller than the critical radius will spontaneously disaggregate.

cause ΔG_V is 0. As systems get further out of equilibrium and become increasingly supersaturated in a mineral, ΔG_V increases and the critical radius decreases. The nucleation rate is a measure of the number of nuclei per unit time that exceed the critical radius. Experimental studies show that, in general, nucleation rates increase with increasing supersaturation (**Fig. 2**). *See* FREE ENERGY; NUCLEATION.

Crystal growth. The rate at which crystals grow, once stable nuclei have formed, is dependent on the rate at which chemical components attach to the surface of a crystal and the rate at which included components diffuse toward, and excluded components diffuse away from, a growing crystal. The crystal morphology that develops is dependent upon the

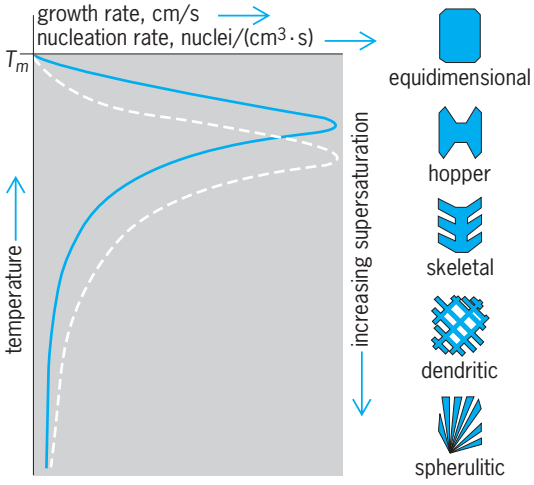


Fig. 2. Growth rate (solid curve) and nucleation rate (broken curve) as a function of temperature for a hypothetical silicate melt and resulting crystal morphology. T_m is the melting temperature of the crystal.

degree of supersaturation when nucleation first occurs. Equidimensional crystals result from interface-controlled growth in which the rate-limiting factor is the attachment of new components at the crystal interface. Cellular morphologies (hopper, skeletal, dendritic, or spherulitic crystal forms) result from diffusion-controlled growth in which the rate-controlling process is the diffusion of components to and from the crystal interface. Experimental studies indicate that cellular morphologies result when the growth rate (cm/s) exceeds the pertinent diffusion coefficients (cm^2/s). Studies also show that after an initial increase, the growth rate reaches a maximum and then decreases with increasing supersaturation. The fall-off in the nucleation and growth rates is the result of a decrease in molecular motion and diffusion rates at lower temperatures. Like nucleation rates, growth rates are equal to 0 at saturation and increase with increasing supersaturation (Fig. 2). In igneous rocks, the relative crystal morphologies are controlled by the rate and order in which minerals grow. In metamorphic rocks, the relative crystal morphologies are controlled by the surface free energies of the competing minerals; the best-developed crystal faces occur in minerals with the highest surface free energies. *See* CRYSTAL GROWTH.

Diffusion. The rate at which chemical components move through solids and liquids is known as the diffusion rate, and is an important control on the rates of crystal nucleation and growth. Diffusion coefficients are a measure of the rate at which components move across a plane perpendicular to the direction of diffusion, and they can be used to compare relative diffusion rates. Diffusion without a compositional gradient is known as intrinsic diffusion and results in no net flux of material; diffusion along a compositional gradient is known as chemical diffusion and results in a net flux of material from areas of high concentration to areas of low concentration. In general, diffusion rates in magmas increase with increasing temperature, increasing water content, and decreasing silica content. Diffusion rates are several orders of magnitude lower in solids than in melts, but can be greatly increased if a fluid phase is present along grain boundaries. *See* CHEMICAL EQUILIBRIUM; DIFFUSION; MAGMA; METAMORPHISM; MINERAL. H. R. Naslund

Redox Equilibria in High-Temperature Systems

The term "oxidation-reduction" (redox) refers to the gain and loss of electrons by atoms or ions. A redox equilibrium is one in which both processes occur.

Iron. Iron (Fe) is one of the most abundant of the elements that form terrestrial planets; it has the property of ready adoption of altervalent states (Fe^0, Fe^{2+}, Fe^{3+}), making redox equilibria particularly important in terms of geochemical processes.

For example, it is believed that all of the terrestrial planets possess a central metallic core, which on cosmochemical grounds most likely is predominantly metallic iron (Fe^0). The formation and gravitational segregation of Fe^0 in a differentiating planet has important consequences for the thermal history (including melting), volatile degassing and evolution, and initial redox state of the overlying silicate mantle of the planet. *See* COSMOCHEMISTRY.

Iron is also generally the most abundant altervalent element presently in the Earth's crust and mantle. Consequently, redox equilibria such as that shown in reaction (2), are important in terms of the min-

$$2FeO + {}^1/_2O_2 \rightleftharpoons Fe_2O_3 \qquad (2)$$

eral assemblages developed, the nature of the volatile species present in multicomponent geologic fluids through equilibria such as shown in reaction (3), and

$$8Fe_3O_4 + CO_2 + 2H_2O \rightleftharpoons 12Fe_2O_3 + CH_4 \qquad (3)$$

the character of silicate melts produced during genesis of magma.

Other redox controls. Within the range of redox states commonly found within the Earth, ranging from equilibria involving Fe^0 to air, there are other less abundant altervalent elements that are important in high-temperature silicate systems. These include the couples of nickel (Ni^0-Ni^{2+}), europium (Eu^{2+}-Eu^{3+}), chromium (Cr^{2+}-Cr^{3+}), titanium (Ti^{3+}-Ti^{4+}), and manganese (Mn^{3+}-Mn^{4+}). Furthermore, in the absence of metal redox equilibria involving silicates, carbon polymorphs and carbonates may be locally the most important redox controls defining the oxygen fugacity (f_{O_2}) in the Earth's upper

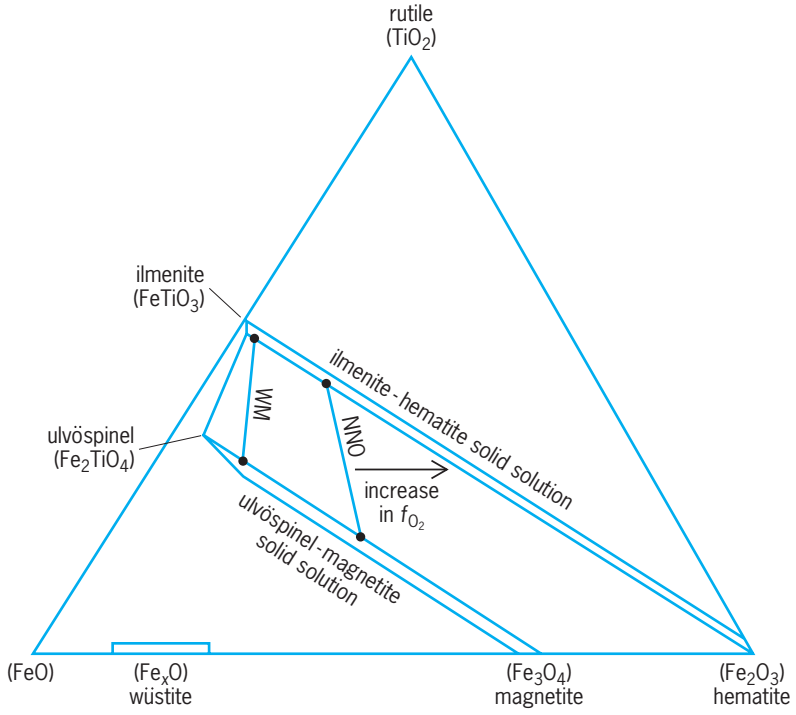


Fig. 3. A portion of the system Fe-Ti-O at 1000°C (1800°F) and 1 bar (10^5) pascals. The double lines between ilmenite-hematite and ulvöspinel-magnetite represent continuous solid solutions, and the width of these zones reflects the extent of Fe-Ti nonstoichiometry. Lines WM and NNO link coexisting phases of solid solutions of ilmenite-hematite and ulvöspinel that are stable under the conditions of the wustite-magnetite and nickel–nickel oxide oxygen buffers, respectively, at the temperature and pressure indicated.

mantle, as shown in reaction (4).

$$\underset{\text{Enstatite}}{MgSiO_3} + \underset{\text{Magnesite}}{MgCO_3} \rightleftharpoons \underset{\text{Forsterite}}{Mg_2SiO_4} + \underset{\text{Diamond or graphite}}{C} + O_2 \quad (4)$$

Experimental measurement of oxygen fugacity. Determination of the oxygen fugacity prevailing in high-temperature silicate systems has been pursued with a variety of methods. These are based on thermodynamic calculations involving mineral assemblages at equilibrium, electrochemical cell equilibria, and comparison of the measured ratio Fe^{2+}/Fe^{3+} of magmas with laboratory calibrations of the temperature–oxygen fugacity–compositional dependence of this ratio.

Thermodynamic calculations. As an example of the thermodynamic approach, the coexistence of different phases that are sensitive to temperature and oxygen fugacity in the system Fe-Ti-O has proved particularly important (**Fig. 3**). Precise compositions of coexisting Fe-Ti oxides are monitors of the temperature-redox conditions under which a given rock formed, and these two "tie lines" are examples of the many coexisting compositions that can be found within the natural range of redox states.

The precise compositions of coexisting solid solutions of ilmenite-hematite (i-h.ss) and ulvöspinel-magnetite (u-m.ss) found in a wide variety of igneous and metamorphic rocks are a function of the absolute temperature (T) and f_{O_2} of equilibration. The f_{O_2}–T dependence may be described by reaction (5), from

$$\underset{\text{in u-m.ss}}{4Fe_3O_4} + O_2 \rightleftharpoons \underset{\text{in i-h.ss}}{6Fe_2O_3} \quad (5)$$

which Eq. (6) can be derived, where ΔG° = standard Gibbs free energy of formation or reaction, T = absolute temperature, R = gas constant, and the notation a^b_c means the activity of component c in phase b. Thermodynamic data (enthalpies, entropies, and cell volumes) are available for the pure end-member components of these solid solution series for derivation of ΔG°. Adoption of appropriate solution models for the activity terms in Eq. (6) allows calculation of the variation of f_{O_2} as a function of T. *See* ACTIVITY (THERMODYNAMICS); SOLID SOLUTION.

$$\ln f_{O_2} = \frac{\ln\left(a^{\text{i-h.ss}}_{\text{hematite }(Fe_2O_3)}\right)^{6-}}{\ln\left(a^{\text{u-m.ss}}_{\text{magnetite }(Fe_3O_4)}\right)^{4}} + \frac{\Delta G^\circ}{RT} \quad (6)$$

Electrochemical cell equilibria. In the second experimental approach, an electrochemical cell is constructed with a high-temperature, O^{2-}-specific solid electrolyte [typically stabilized zirconium oxide (ZrO_2)]. One electrode is formed from the geological material of interest containing at least one redox pair (for example, Fe^{2+}-Fe^{3+}) in one or more phases. The other electrode is a reference material of known oxygen potential. At a given temperature, the electromotive force (emf; E) measured across the cell with a high-impedance voltmeter is a function of the difference in oxygen potential between the two electrodes, and is given by the Nernst equation (7), which is the expression for the activity dependence of emf; F is the Faraday constant.

$$E = \frac{RT}{4F} \ln \frac{(f_{O_2\text{sample}})}{(f_{O_2\text{reference}})} \quad (7)$$

Knowing the variation of the oxygen fugacity of the reference material as a function of T and measuring E allows calculation of the oxygen fugacity of the sample. *See* ELECTROMOTIVE FORCE (EMF); ENTHALPY; ENTROPY.

Comparison of ratios with laboratory calibrations. The theoretical basis for the third approach to experimental measurements is founded in reaction (2), which can be developed as Eq. (8). Empirical calibrations

$$\ln (f_{O_2})^{1/2} = \frac{\Delta G^\circ}{RT} + \frac{\ln\left(a^{\text{magma}}_{Fe_2O_3}\right)}{\left(a^{\text{magma}}_{FeO}\right)^2} \quad (8)$$

of this equilibrium with a wide range of natural and synthetic silicate melts have been achieved.

Interpretation. Application of the three experimental approaches either singly or in combination has shown that the majority of igneous rocks that have

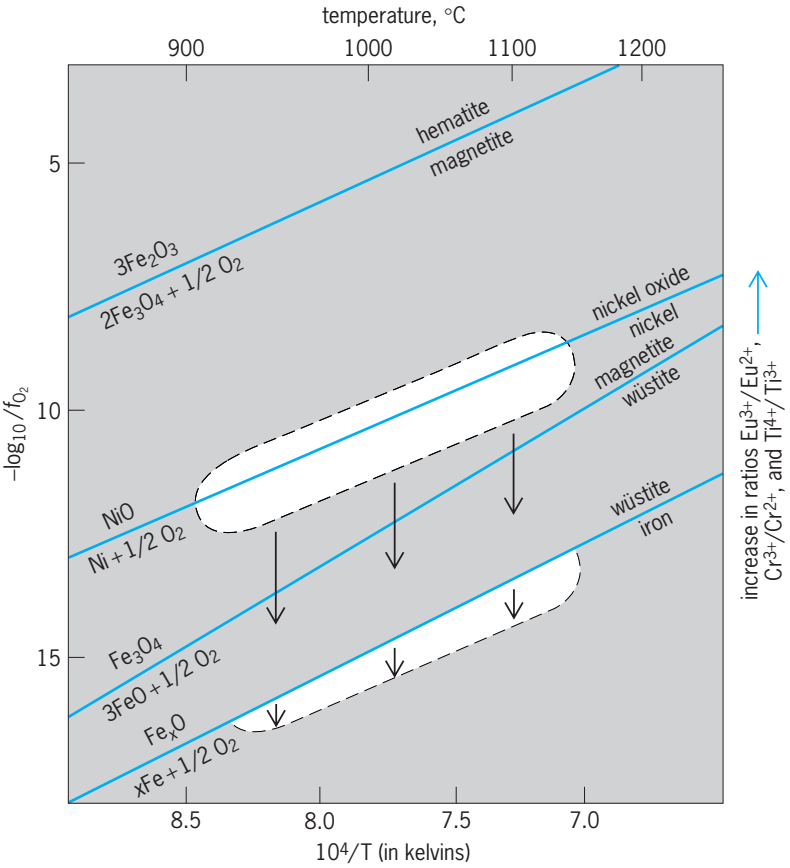


Fig. 4. Range of redox states present within most terrestrial igneous and metamorphic rocks represented in terms of oxygen fugacity (f_{O_2}) as a function of temperature. Most rocks have formed between the oxygen buffers represented by the iron-wustite and magnetite-hematite redox equilibria. In fact, the great majority of erupted rocks lies close to the nickel–nickel oxide redox equilibrium. The arrows indicate a gradation between the areas enclosed in broken lines from (top) most erupted igneous rocks to (bottom) some high-pressure terrestrial, lunar, and meteoritic systems. °F = (°C × 1.8) + 32°.

erupted at the Earth's surface lie within the f_{O_2} range defined by the iron-wustite and magnetite-hematite equilibria (**Fig. 4**). At the more reduced end of this redox range, a readily measurable fraction of the Eu, Ti, and Cr in igneous and metamorphic rocks occurs in the lower of the valence states listed earlier.

In the early years of quantitative experimental studies of silicate systems, the control of oxygen fugacity, iron redox state, and the loss of iron from the systems under investigation through alloying with sample containers was difficult. With a few exceptions, the problems were mostly avoided by studies of iron-free, synthetic analogs. However, it was particularly through the efforts of E. F. Osborn and colleagues that methods of atmospheric control in studies of silicate systems were developed during the 1950s and onward. Initially, these studies were pursued with support from the steel-making industry because of the relevance of optimizing the performance of iron-bearing silicate slags. In addition, it became apparent that many of the processes involved in the heterogeneous phase equilibria of iron-bearing silicates, oxides, melts, and vapor were relevant to natural magmatic systems.

For example, it has long been recognized that within the subalkaline spectrum of igneous rock compositions, two major lineages are present that are characterized by systematic early increases of the FeO/MgO ratio during increases in concentration of SiO_2 (tholeiitic trend), and a relatively constant FeO/MgO value during equivalent increases in SiO_2 (calcalkalic trend), respectively. Although it is recognized that a number of alternative processes may be involved to account for the development of these lineages at different times and places, the timing of saturation and fractional crystallization of a spinel rich in the magnetite (Fe_3O_4) component can be critical. The stability of spinel in a given magma is a function of pressure, temperature, bulk composition, and oxygen fugacity.

Much experimental effort has been expended since 1970 in efforts to control the intensive parameters and retention of iron in synthetic and natural systems during experimental studies. Rewards for this effort include a thorough understanding of the partitioning of FeO and MgO between olivine and silicate melt, and the acquisition of detailed knowledge of the phase behavior of subsolidus and supersolidus relations in the upper mantle.

The nature of solid-melt-fluid speciation in the system C-O-H-S is also well understood under conditions relevant to the deep crust and upper mantle. The predominance of CH_4-H_2O in fluids close to iron-wustite buffer conditions and CO_2-H_2O in more oxidized settings has been determined, and this is clearly relevant to the nature of gas species that could have formed a primitive terrestrial atmosphere during the early evolutionary stages of the Earth.

Finally, the possibility of secular changes in the redox state of the Earth's upper mantle has been examined. It is clear that the return of relatively oxidized oceanic lithosphere to the interior of the planet is a prime agent of local oxidation of the mantle, and indeed analysis of magmas and fragments of the mantle associated with subduction zones indicates that these regions are relatively oxidized compared with other zones of magma generation, such as at mid-oceanic ridges.

Research is focused on the mechanisms of redox control within the Earth's mantle and possible causes for spatial and temporal variations. Comparative studies of the redox evolution of the Earth, Moon, and differentiated meteorites are also of interest with respect to core separation events, metal-silicate equilibria, volatile loss, and recycling. *See* FUGACITY; OXIDATION-REDUCTION. Richard Arculus

Volatile Equilibria at High Temperatures and Pressures

Volatiles are compounds that tend to form gases or fluids in high-temperature and relatively low-pressure geologic environments. They occur as a gas phase, as essential constituents of some minerals, and as dissolved constituents in magmas (molten rock) and fluids. Thermodynamic modeling and experimental research on volatiles determine the conditions of pressure, temperature, and composition at which volatile-bearing rocks form. By definition, volatile equilibria are equilibrium relationships established between volatile-bearing minerals, magmas, and fluids. Volatile equilibria control many high-temperature and high-pressure geologic processes in igneous and metamorphic mineral assemblages, including volcanic eruptions and ore deposition from fluids.

Volatile species. Analyses of volcanic gases, fluid inclusions (small gas or fluid-filled bubbles within rocks), and volatiles in rocks and minerals indicate that the dominant volatiles include water (H_2O) and compounds of carbon (C), sulfur (S), nitrogen (N), fluorine (F), and chlorine (Cl)—that is, carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4), sulfur dioxide (SO_2), hydrogen sulfide (H_2S), ammonia (NH_4), hydrogen fluoride (HF), and hydrogen chloride (HCl). The volatile species that occur in a given igneous or metamorphic environment vary as a function of pressure, temperature, and system composition. The stability of some of the important species is defined by the volatile equilibria shown in reactions (9)–(13). The dominant volatile species released during eruptions of basaltic volcanoes include H_2O, CO_2, and SO_2, and the dominant species generated during thermal metamorphism include H_2O and CO_2 with or without CO, CH_4, and H_2S. The extent to which these volatiles alter the stability of minerals, magmas, and fluids is largely proportional to

$$C + 2H_2 \rightleftharpoons CH_4 \quad (9)$$

$$H_2 + {}^1/_2O_2 \rightleftharpoons H_2O \quad (10)$$

$$C + O_2 \rightleftharpoons CO_2 \quad (11)$$

$$O_2 + {}^1/_2S_2 \rightleftharpoons SO_2 \quad (12)$$

$$H_2 + {}^1/_2S_2 \rightleftharpoons H_2S \quad (13)$$

Maximum concentrations of volatiles in silicate minerals and magmas*

	Components, wt %			
	Water	Carbon dioxide	Chlorine	Fluorine
Igneous minerals	7	4	2	21
Granitic magmas†	10	0.5	1	5
Metamorphic minerals	14	4	1.5	9

*Concentrations in minerals are for mica, amphibole, chlorite, serpentine, topaz, humite, and scapolite.
†Concentrations are for pressures of ≤150,000 lb/in.^{2} (1030 megapascals); the solubilities of H_2O and CO_2 are strongly pressure-dependent.

their concentration in the system (see **table**). *See* VOLCANO.

A fluid phase is not always stable in volatile-bearing systems. Each component in an assemblage exerts a pressure on the system, and the highest relative pressures are exerted by the more volatile components. A fluid phase is unstable if the sum of the individual partial pressures of the volatiles is less than the confining rock pressure. The relationship between partial pressure and total gas pressure is given by Eq. (14), where P_i is the partial pressure of

$$P_i = P_{\text{total}} X_i \tag{14}$$

species i, P_{total} the total fluid pressure in the system, and X_i the concentration of i expressed as the mole fraction. Alternatively, at pressures or temperatures lower than that of a liquid's critical point, a single fluid phase may be replaced by a gas coexisting with a liquid.

Influences. Water is the most significant volatile to influence geologic processes in many environments, primarily because of its greater abundance relative to most other volatiles. Water fluxes the melting of rocks. For example, the presence of only 5 wt % H_2O in granitic and basaltic magmas reduces the liquidus temperatures by 200–300°C (392–572°F). Furthermore, during melting of mica- and amphibole-bearing rocks, the water in these hydrous minerals increases the relative stability of silicate melt by reducing magmatic solidus temperatures [reaction (15)]. An increasing abundance of H_2O in silicate

$$SiO_2 + KMg_3(AlSi_3O_{10})(OH)_2 \rightleftharpoons KAlSi_3O_8 + MgSiO_3 + \text{hydrous melt} \tag{15}$$

magmas also profoundly disrupts the internal structure of silicate magma, reduces magmatic viscosities, enhances rates of diffusion of species though magma, increases or decreases crystallization rates, and increases the explosive nature of volcanic eruptions.

Increasing the extent of thermal metamorphism decreases the stability of hydrous mineral assemblages and increases the relative stability of H_2O-bearing fluids, such as in reaction (16). As the

$$KAl_2(AlSi_3O_{10})(OH)_2 + SiO_2 \rightleftharpoons KAlSi_3O_8 + Al_2SiO_5 + H_2O\ \text{fluid} \tag{16}$$

hydrous assemblage, muscovite plus quartz, is subjected to increasing thermal metamorphism, it is increasingly unstable and gradually dehydrates to an anhydrous assemblage of potassium feldspar and, depending on temperature and pressure, one of the aluminosilicate minerals (kyanite/sillimanite/andalusite) and a water-rich fluid.

An increasing abundance of carbon dioxide in a fluid decreases the stability of hydrous mineral assemblages and variably alters the melting temperatures of hydrous and anhydrous magmas. Because of its relatively low solubility in most magmas, carbon dioxide has a fairly small effect on magmatic solidus and liquidus temperatures, in contrast to water. The solubility of carbon dioxide in silicate magmas has been found experimentally to vary strongly with changes in the composition of the magma. Devolatilization resulting from increasing thermal metamorphism generates H_2O- and CO_2-bearing fluids via dehydration and decarbonation reactions, respectively. For example, fluids rich in CO_2 are generated by equilibria such as that in reaction (17).

$$\underset{\text{carbonate}}{\underset{\text{Calcium}}{CaCO_3}} + SiO_2 \rightleftharpoons \underset{\text{silicate}}{\underset{\text{Calcium}}{CaSiO_3}} + CO_2\ \text{fluid} \tag{17}$$

The release of sulfur dioxide (SO_2) from magma to the atmosphere during explosive volcanic eruptions can produce short-term climate variations. Sulfur dioxide reacts with water vapor in air to form small droplets of sulfuric acid (H_2SO_4), and these aerosol droplets have been shown to reflect sunlight away from Earth and hence to reduce the extent of solar heating of the Earth's surface for short periods.

The volatile elements fluorine, chlorine, and sulfur and their compounds are important in the transport of ore and non-ore minerals in CO_2- and H_2O-bearing fluids. Under conditions of high temperature and pressure, they enhance the solubilities of minerals in such fluids. Fluorine is similar to water in that it shows relatively large solubility in silicate magmas and, as a result, high concentrations of fluorine in silicate magmas can significantly reduce solidus and liquidus temperatures of magmas, enhance diffusivities, and disrupt magma structures. Sulfur compounds are also important for the deposition of many ore metals in the form of sulfide minerals. *See* ORE AND MINERAL DEPOSITS.

Experimental studies. The relative stability of a given silicate assemblage can be predicted through thermodynamically based computations, but specific values of pressure, temperature, and composition at which an assemblage is stable are best fixed by experimental research. In experimental simulations, volatiles and natural minerals or synthetic analogs of

minerals are heated and pressurized in sealed capsules of precious metals. Depending upon the kinetics of the reactions, the establishment of equilibrium requires durations of experimental runs ranging from minutes to months. Examination and chemical analysis of the materials at the end of an experiment determine the identity and composition of the minerals and the melt or fluid that were stable during the experiment.

In mineral-, melt-, and fluid-bearing systems at equilibrium, the most stable assemblage is the one that exhibits the lowest total free energy (G_{total}). In reaction (16), for example, muscovite and quartz are the stable phases at those temperatures where G_{total} of muscovite and quartz is lower than that of the fluid and anhydrous mineral assemblage.

If natural mineral or melt comparisons are not homogeneous or if a mineral assemblage contains minerals that are not mutually compatible, then chemical equilibrium has not been established within the assemblage and it is metastable. *See* HIGH-PRESSURE MINERAL SYNTHESIS; IGNEOUS ROCKS; METAMORPHIC ROCKS; SILICATE PHASE EQUILIBRIA.

James D. Webster

Bibliography. M. Carroll and J. R. Holloway, *Volatiles in Magmas*, Mineralogical Society of America, Reviews in Mineralogy Series, vol. 30, 1994; J. Crank, *The Mathematics of Diffusion*, 1975; J. M. Ferry and D. M. Burt, *Characterization of Metamorphism through Mineral Equilibria*, Mineral Society of America, Reviews in Mineralogy Series, vol. 10, pp. 207–262, 1982; P. Henderson, *Inorganic Geochemistry*, 1982; J. R. Holloway, *Advances in Physical Chemistry 1*, 1981; R. J. Kirkpatrick, Crystal growth from the melt: A review, *Amer. Mineralog.*, 60:798–814, 1975; R. J. Kirkpatrick, Theory of nucleation in silicate melts, *Amer. Mineralog.*, 68:66–77, 1983; G. E. Lofgren, An experimental study of plagioclase crystal morphology: Isothermal crystallization, *Amer. J. Sci.*, 274:243–273, 1974; D. K. Nordstrom and J. L. Munoz, *Geochemical Thermodynamics*, 1987; R. C. Reid, J. M. Prausnitz, and B. E. Roling, *The Properties of Gases and Liquids*, 4th ed., 1987; H. Sigurdsson, *Encyclopedia of Volcanoes*, pp. 149–170, 1999; G. C. Ulmer, *Research Techniques for High Pressure and High Temperature*, 1971.

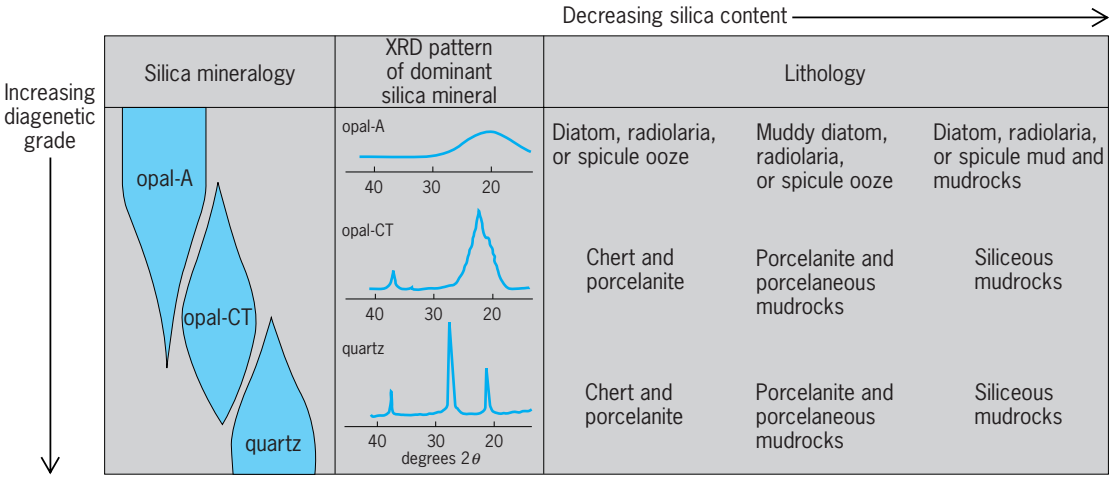
Siliceous sediment

Fine-grained sediment and sedimentary rock dominantly composed of the microscopic remains of the unicellular, silica-secreting plankton diatoms and radiolarians. Minor constituents include extremely small shards of sponge spicules and other microorganisms such as silicoflagellates. Siliceous sedimentary rock sequences are often highly porous and can form excellent petroleum source and reservoir rocks. *See* SEDIMENTARY ROCKS.

In the modern oceans, the distribution of siliceous sediments closely resembles the distribution of nutrients in the uppermost oceanwaters, because diatoms and radiolarians are most abundant in oceanic regions of high productivity. Some of the greatest concentrations of siliceous sediment are adjacent to Antarctica; in the equatorial Pacific Ocean; and along the continental shelves of Peru, the western United States, and west Africa—all of which are oceanographic regions with elevated biologic productivity. The continued productivity results in siliceous sediment being deposited on the sea floor in these regions. Sometimes large-scale die-offs of siliceous planktonic communities can result in the deposition of mats of intertwined diatom assemblages on the sea floor. *See* SILICA MINERALS.

Siliceous sediment undergoes a complicated chemical and physical process during burial on the sea floor (see **illus.**). Diatoms and radiolarians secrete a form of silica that is an amorphous mineral called opal-A. Because seawater is undersaturated with respect to this silica mineral, most of the opal-A microskeletons dissolve on or within the uppermost few centimeters of the sea floor. With continued burial to a few hundreds of meters, the small proportion of opal-A that survives initial burial continues to dissolve, and eventually (with continued sedimentation burying the mud to higher temperatures and pressures) a new siliceous mineral is precipitated, called opal-CT. This opaline mineral consists of disordered cristobalite and tridymite, structures which are also made of silica. With still more burial to many hundreds of meters below the sea floor, the opal-CT in turn becomes chemically unstable, and a new mineral called biogenic quartz is formed. Biogenic quartz is the dominant mineral in the siliceous rock chert (also known as flint or, more rarely, jasper). This mineral is chemically stable, and although it is chemically similar to the quartz found on sandy beaches, its process of formation is entirely different. Biogenic quartz owes its existence to the diatoms and radiolarians that initially thrived in the water column above the sea floor, whereas the quartz found on typical beaches results from the erosion of igneous and metamorphic rocks. Chert and flint are hard, dense rocks, and their fine grain size results in maintenance of sharp edges, a characteristic capitalized upon by early humans who used chert for arrowheads, other weapons, and jewelry. *See* OPAL; QUARTZ.

Given their biologic composition, siliceous sediments provide some of the best geologic records of the ancient oceans. Diatoms did not evolve until the late Mesozoic; thus the majority of siliceous rocks older than approximately 150 million years are formed by radiolarians. Geologists map the distribution of ancient siliceous sediments now pushed up onto land by plate tectonic processes, and can thus determine which portions of the ancient seas were biologically productive; this knowledge in turn can give great insight into regions of the Earth's crust that may be economically productive (for example oil-containing regions). The vast oil reserves of coastal California are predominantly found in the Monterey Formation, a highly porous diatomaceous



Generalized relationship between silica mineralogy and rock types. (*After K. A. Pisciotto, Distribution, thermal histories, isotopic compositions, and reflection characteristics of siliceous rocks recovered by the Deep Sea Drilling Project, in The Deep Sea Drilling Project: A Decade of Progress, vol. 32, ed. by J. E. Warme, R. G. Douglas, and E. L. Winterer, Society of Economic Paleontologists and Mineralogists, Tulsa, 1981*)

siliceous sedimentary sequence distributed along the western seaboard of the United States. These sediments and rocks are particularly well exposed along the beaches of Santa Barbara, where some of the outcrops are so petroliferous that they can be lit with a common household match and can burn slowly for several hours. *See* CHALK; CHERT; LIMESTONE.

R. W. Murray

Bibliography. G. Bohrmann et al., Pure siliceous ooze a diagenetic environment for early chert formation, *Geology*, 22:207–210, 1994; J. R. Hein, *Siliceous Sedimentary Rock-Hosted Ores and Petroleum*, Van Nostrand Reinhold, New York, 1987; A. E. S. Kemp and J. G. Baldauf, Vast Neogene laminated diatom mat deposits from the eastern equatorial Pacific Ocean, *Nature*, 362:141–143, 1993; M. Leinen et al., Distribution of biogenic silica and quartz in recent deep-sea sediments, *Geology*, 14:199–203, 1986; R. W. Murray, Chemical criteria to identify the depositional environment of chert: General principles and applications, *Sed. Geol.*, 90:213–232, 1994; K. A. Pisciotto, Distribution, thermal histories, isotopic compositions, and reflection characteristics of siliceous rocks recovered by the Deep Sea Drilling Project, in *The Deep Sea Drilling Project: A Decade of Progress*, vol. 32, 1981.

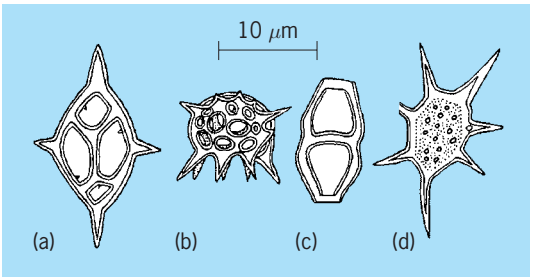
Siliceous sinter

A porous silica deposit formed around hot springs. It is white to light gray and sometimes friable. Geyserite is a variety of siliceous sinter formed around geysers. The siliceous sinters are deposited as the hot subterranean waters cool after issuing at the surface and become supersaturated with silica that was picked up at depth. The sinters are frequently deposited on algae that live in the pools around the hot springs. *See* GEYSER; SPRING (HYDROLOGY).

Raymond Siever

Silicoflagellata

A class of marine planktonic chromists. Their skeletons are composed of hollow, siliceous rods and resemble those of the Radiolaria, with which they have been grouped. They are usually subpyramidal or hemispherical in shape, delicately filigreed, and range in size from 10 to 150 micrometers (see **illus.**).



Examples of fossil and modern Silicoflagellata. (*a*) *Dictyocha*, Cretaceous to Recent; (*b*) *Cannopilus*, Miocene; (*c*) *Naviculopsis*, Eocene to Miocene; and (*d*) *Vallacerta*, Upper Cretaceous.

Two families and 11 genera of silicoflagellates have been described from siliceous sedimentary rocks ranging in age from the early Cretaceous to Recent, in association with abundant diatoms and siliceous sponge spicules; however, this may change pending systematic revision. One genus, *Dictyocha*, lives in the ocean today. *See* CHRYSOPHYCEAE; MICROPALEONTOLOGY; PHYTOPLANKTON.

Daniel J. Jones; Sally Walker

Silicoflagellida

An order of the phylum Protozoa, subphylum Sarcomastigophora, superclass Mastigophora, class Phytamastigophorea. These organisms are marine flagellates which have an internal, siliceous, tubular skeleton; numerous small, discoid, yellow

chromatophores; and a single flagellum. At times the organisms also put forth, from the ends of their skeletal tubes, long, rather threadlike pseudopodia. The skeleton forms a basket within which the moiety of protoplasm lies, but skeletal elements always have at least a thin covering. Nuclei are not conspicuous. Reserve material consists of leucosin.

There is now a single genus, with *Dictyocha fibula* being the most abundant species, although three other species have been described. The organism is not uncommon in Atlantic, Pacific, and Gulf of Mexico waters. However, it is normally not abundant enough to play an important role in ecology.

In past ages a large number of species were abundant, judging from fossil remains. The skeletons vary from quadrate through pentagonal, hexagonal, and octagonal types. *See* PHYTAMASTIGOPHOREA; PROTOZOA.

James B. Lackey

Silicon

A chemical element, Si, atomic number 14, and atomic weight 28.086. Silicon is the most abundant electropositive element in the Earth's crust. The element is a metalloid with a decided metallic luster; it is quite brittle. It has a specific gravity of 2.42 at 20°C (68°F), melts at 1420°C (2588°F), and boils at 3280°C (5936°F). The element is usually tetravalent in its compounds, although sometimes divalent, and is decidedly electropositive in its chemical behavior. In addition, pentacoordinate and hexacoordinate compounds of silicon are known. *See* METALLOID; PERIODIC TABLE.

Crude elementary silicon and its intermetallic compounds are used in alloying constituents to strengthen aluminum, magnesium, copper, and other metals. Metallurgical silicon of 98–99% purity is used as the starting material for manufacturing organosilicon compounds and silicone resins, elastomers, and oils. Silicon chips are used in integrated circuits. Photovoltaic cells for direct conversion of solar energy to electricity use wafers sliced from single crystals of electronic-grade silicon. Silicon dioxide is used as the raw material for making elementary silicon and for silicon carbide. Sizable crystals of it are used for piezoelectric crystals. Fused quartz sand becomes silica glass, used in chemical laboratories and plants as well as an electrical insulator. A colloidal dispersion of silica in water is used as a coating agent and as an ingredient in certain polishes.

Naturally occurring silicon contains 92.2% of the isotope of mass number 28, 4.7% of silicon-29, and 3.1% of silicon-30. In addition to these stable, natural isotopes, several artificially radioactive isotopes are known. Elementary silicon has the physical properties of a metalloid, resembling germanium below it in group 14 of the periodic table. In very pure form silicon is an intrinsic semiconductor, although the extent of its semiconduction is greatly increased by the introduction of minute amounts of impurities. Silicon resembles the metals in its chemical behavior. It is about as electropositive as tin, and decidedly more positive than germanium or lead. In keeping with this rather metallic character, silicon forms tetrapositive ions and a variety of covalent compounds; it appears as a negative ion in only a few silicides and as a positive constituent of oxy acid or complex anions.

Several series of hydrides are formed, a variety of halides (some of which contain silicon-to-silicon bonds), and also many series of oxygen-containing compounds which may be either ironic or covalent in their properties.

Silicon occurs in many forms of the dioxide and as almost numberless variations of the natural silicates. For a discussion of the structures and compositions of the representative classes, *see* SILICATE MINERALS.

In abundance, silicon exceeds by far every other element except oxygen. It constitutes 27.72% of the solid crust of the Earth, whereas oxygen constitutes 46.6%, and the next element after silicon, aluminum, accounts for 8.13%.

Silicon is reported to form compounds with 64 of the 96 stable elements, and it probably forms silicides with 18 other elements. Besides the metal silicides, used in large quantities in metallurgy, silicon forms useful and important compounds with hydrogen, carbon, the halogen elements, nitrogen, oxygen, and sulfur. In addition, useful organosilicon derivatives and silicon-containing polymers have been prepared.

Edwin P. Plueddemann

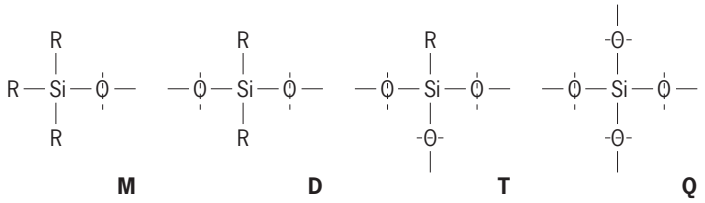
Bibliography. M. A. Brook, *Silicon in Organic, Organometallic and Polymer Chemistry*, Wiley-Interscience, 1999; F. A. Cotton et al., *Advanced Inorganic Chemistry*, 6th ed., Wiley-Interscience, 1999; P. Jutzi and U. Schubert (eds.), *Silicon Chemistry: From the Atom to Extended Systems*, John Wiley & Sons, 2003; U. Schubert (ed.), *Silicon Chemistry*, Springer, 1999.

1																	18
1 H	2											13	14	15	16	17	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

Silicone resins

Soluble or meltable materials based on highly branched polymers composed of alternating silicon and oxygen atoms and having organic (R) substituents attached to the silicon atoms. On curing, these polymers are converted into highly cross-linked, insoluble materials. They may contain four types of units, usually symbolized by the letters M, D, T, Q, as follows:



Each of the oxygen atoms is shared by neighboring units or by the unit and a functional OR group (R = H or alkyl). Therefore, the formulas of units are M—$R_3SiO_{1/2}$, D—R_2SiO, T—$RSiO_{3/2}$, Q—SiO_2.

Silicone resins are combinations of these units, the proportions of which are varied over a broad range. They are characterized by the R/Si ratio, which usually is in the range 1–1.7. Silicone resins constitute one of the three main classes of silicone products. The other two are silicone rubbers (R/Si = 2) and silicone fluids (R/Si > 2). The structure and properties of silicone resins are strongly dependent on the R/Si value. The lower the R/Si value, the less miscible the resin with organic polymers, the easier it undergoes gelation, and the harder it is after curing. Conversely, resins with higher R/Si values (containing more organic groups) are more flexible and compatible with organic resins. The basic component of silicone resins is usually the T unit. Polymers composed exclusively of T units, having the $(RSiO_{3/2})_n$ structure, are called polysilsesquioxanes and may contain fragments of a ladder or a cage structure. T resins having R/Si = 1 are known, but more common are DT resins. The D units give the resin softness and flexibility, which is particularly important in paints and varnishes. M and Q units also are often used. Resins composed exclusively of M and Q units, called MQ resins, are very important and have recently found numerous applications, such as pressure-sensitive adhesives and flexible silicone coatings, as well as high-performance defoamers, release agents, and additives for improving the mechanical properties of silicone rubber. *See* INORGANIC AND ORGANOMETALLIC POLYMERS; ORGANOSILICON COMPOUND; POLYMER; RUBBER.

The organic group is most often methyl or phenyl. Commonly used phenylmethyl silicone resins contain both phenyl and methyl in various proportions. Phenyl groups give the material superior thermal resistance. Cured resins containing a large proportion of phenyl groups are less brittle and more readily miscible with organic polymers. Other organic groups are introduced to give silicone resins special properties. For example, enhanced water repellency is attained by introducing fluoroalkyl groups or long-chain hydrocarbon groups, such as dodecyl ($C_{12}H_{25}$). Silicone resins contain a considerable number of hydroxy (HO—) or alkoxy (for example, CH_3O—) groups which, in the curing process, cross-link the polymer by condensation. These reactive groups may also form hybrids with organic resins. Vinyl groups or SiH groups are sometimes introduced in silicone resins to give them the ability to cure or form hybrids by the addition to olefinic bonds in organic polymers.

The presence of reactive silanol (SiOH) groups in the resins may make them unstable, particularly if the R/Si ratio is low and the molecular mass is relatively high. Their stability may be increased by the partial replacement of hydroxy groups by alkoxy groups. The stability of T resins (R/Si = 1) used for masonry is attained by conversion of the $\equiv$SiOH to $\equiv$SiONa or $\equiv$SiOK groups. [Here $\equiv$ means a triply bonded Si atom; that is, in $\equiv$SiOH only SiOH group reacts and the other three (undefined) do not.]

Silicone resins are usually sold as solutions of the highly branched polymer with a molecular mass of 1000–60,000 daltons in an inert solvent such as an aromatic hydrocarbon, chlorinated hydrocarbon, or alkyl ester. Commercial solvent-free resins are either viscous liquids or powders which melt at 50–80°C (120–180°F). Some silicone resins are available as emulsions or dispersions in water.

Preparation. Of the basic monomers used for the synthesis of silicone resins, CH_3SiCl_3 and $C_6H_5SiCl_3$ are precursors of T units. For the generation of D units, $(CH_3)_2SiCl_2$, $C_6H_5MeSiCl_2$, and $(C_6H_5)_2SiCl_2$ are used. The precursor of the M unit is usually $(CH_3)_3SiCl$. The precursors of Q units are esters of orthosilicic acid $(C_2H_5O)_4Si$ (TEOS) or $(CH_3O)_4Si$ (TMOS), their polymers (for example, ethyl polysilicate 40 which contains 40% SiO_2), inorganic silicates, and $SiCl_4$. Vinyl groups may be introduced using $H_2C{=}CHSiCl_3$ or $H_2C{=}CHSi(CH_3)Cl_2$, while $(CH_3)_2SiHCl$ or CH_3SiHCl_2 is used to introduce SiH groups.

Methyl-substituted monomers CH_3SiCl_3, $(CH_3)_2SiCl_2$, $(CH_3)_3SiCl$ are synthesized using exclusively the Rochow-Müller direct synthesis from elemental silicon and methyl chloride (CH_3Cl) at 250–350°C (480–660°F), catalyzed with copper. $(CH_3)_2SiHCl$ and CH_3SiHCl_2 may be obtained by the same process by supplying HCl to the reactor. Phenyl-substituted monomers are manufactured by direct synthesis from Si and C_6H_5Cl at 500°C (930°F), catalyzed with Ag, or by the Grignard reaction from C_6H_5Cl and $SiCl_4$. *See* GRIGNARD REACTION.

Formerly, TEOS and TMOS were obtained by alcoholysis of $SiCl_4$ produced by the chlorination of silicon. However, direct synthesis from silicon and alcohol has become important. Inorganic silicates are products of reaction of sodium or potassium hydroxide with silica.

Synthesis of silicone resins is based on the hydrolytic polycondensation of a mixture of the organochlorosilane monomers. Reactions (1)–(5)

$$\equiv SiCl + H_2O \rightleftharpoons \equiv SiOH + HCl \quad (1)$$

$$\equiv SiCl + ROH \rightleftharpoons \equiv SiOR + HCl \quad (2)$$

$$\equiv SiCl + HOSi\equiv \rightleftharpoons \equiv SiOSi\equiv + HCl \quad (3)$$

$$\equiv SiOH + HOSi\equiv \rightleftharpoons \equiv SiOSi\equiv + H_2O \quad (4)$$

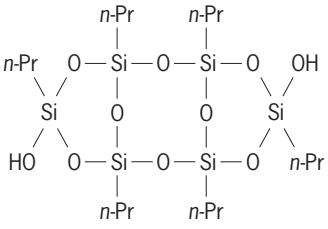
$$\equiv SiOR + HOSi\equiv \rightleftharpoons \equiv SiOSi\equiv + ROH \quad (5)$$

occur in the polycondensation system; however, part of the ≡SiOH and ≡SiOR groups must remain unreacted.

This process may be performed in various ways. Direct hydrolysis-condensation by the addition of the mixture of monomers to an excess of water is difficult to control as considerable differences in hydrolysis rates and in condensation rates of various components lead to a nonuniform product. Better control is gained when water is added slowly to the mixture of monomers and HCl is removed. A mixture of water with alcohol moderates the rate of the process and introduces alkoxy groups to the resin. The presence of alkoxy in place of ≡SiOH makes the resin more stable, and the cured material with remaining alkoxy is more flexible. Using a solvent for the monomers and resins that is immiscible with water slows down the process, diminishes the possibility of gelation, and decreases the polymer viscosity. The process must be stopped well before the gelation point when the desired viscosity is achieved. The process is terminated by cooling, washing, neutralization, and in some cases the addition of solvent.

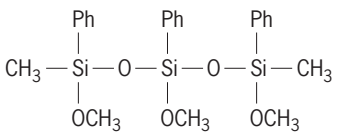
MQ resins are obtained by two routes. The older process uses sodium silicate and trimethylchlorosilane as precursors. The more recent, easier to control route exploits the hydrolytic condensation reaction between hexamethyldisiloxane [$(CH_3)_3SiOSi(CH_3)_3$] and TEOS, TMOS, or ethyl silicate 40.

Some properties of silicone resins may be improved and the price of material may be lower if a hybrid of a silicone resin and an organic resin, such as an alkyd, polyester, epoxy, or acrylic resin, is used. These hybrids are usually obtained by condensation of silanol or alkoxy groups on a silicone resin with hydroxyl groups on an organic resin. If poor resin compatibility is a problem, silicone intermediates (reactive additives), such as (**1**) and (**2**), can be used to increase compatibility.



Dow Corning Z-6018

(1)



Dow Corning Z-6188

(2)

Curing. Curing of silicone resins is the process in which the three-dimensional network is formed. It is most often accomplished by the condensation reaction of the silanol and alkoxysilane groups remaining in the resin, according to Eqs. (4) and (5). The process is run at high temperature using catalysts such as acid and bases, inorganic salts, titanium, and organotin reagents. Curing of silicone coatings by air oxidation involves the cleavage of methyl groups with siloxane-bridge (≡SiOSi≡) formation. This process takes place at temperatures above 150°C (300°F). Resins containing vinyl groups may be cured by free-radical coupling, initiated by peroxides. Curing of resins containing both the SiH and $SiCH{=}CH_2$ groups occurs under mild conditions but needs to be catalyzed by metal complexes. The advantage of this method is that no volatile by-products are formed. *See* FREE RADICAL; ORGANIC SYNTHESIS; REACTIVE INTERMEDIATES.

Properties and applications. Silicone resins have outstanding thermal stability and weathering resistance. They change little on exposure to humidity, oxygen, heat, and sunlight. They are also physiologically inert, retain their gloss for a long time, and have good electrical resistance. Their mechanical and electrical properties are not very dependent on temperature. Thus, although some organic resins may have a higher hardness and strength at room temperature, silicone resins are superior to other resins at extreme temperature conditions. Silicone resins have very interesting surface properties, including very low surface energy. They are water-repellent and foreign bodies do not stick to them. They also show high permeability for gases.

Silicone resins are particularly useful in applications where the material is exposed to extreme temperatures and adverse weather conditions, with many types of silicone resins produced for specific applications.

A variety of silicone resins are used for paints. Heat-resistant paints and varnishes are commonly used on parts for furnaces and ovens, chimneys, pipelines for transporting heating media, exhaust pipes, engine and turbine parts, as well as machinery and apparatus operating at high temperatures. These paints contain a considerable amount of phenyl groups and are often used together with aluminum pigments. They withstand temperatures up to 450°C (840°F) and may work for a long time at 250°C (480°F).

Silicone paints based on methyl-substituted polymers are used on the exterior walls of buildings. They resist weathering, preserve gloss, and repel water, dust, and dirt. These water-repellent silicone-resin coatings make the surfaces hydrophobic, protecting them against humidity. They are most often used for protecting masonry. There are two general types of these materials: water-soluble sodium or potassium polymethylsiliconate resin, and solutions of methyl-substituted silicone resin in an organic solvent, most often a hydrocarbon. These clear, colorless coatings preserve the original appearance of the masonry and leave the masonry pores open, which permits the wall to "breathe."

Among the silicone impregnating varnishes and resins, particularly important are those used for the protecting electrical machinery and devices. These silicone resins provide high thermal stability and excellent electrical resistance and are used for

coating motor stator coils, transformer windings, and generator coils.

Silicone resins also are used as release coatings to prevent other bodies from sticking to surfaces. The most important are glazes for baking pans, trays, and molds. The surface properties of the resins are most critical in these applications, and high-modulus (stiff) methylsilicone resins of a low R/Si ratio are used. In contrast, paper coatings require a resin of lower modulus (more flexible). Silicone resins for pressure-sensitive adhesives are based on MQ resins with an average molecular mass of 1000–10,000 daltons. Silicone DT laminating resins for composites that are produced for the electrical industry contain a large number of silanol groups. Silicones for molding compounds are often solvent-free DT resins.

The range of use of silicone resins is expanding, with many new applications such as antifouling coatings, sensor coatings, resins for microelectronics, materials for photopattering, adhesives for porous ceramic materials, abrasion-resistant coatings, materials for optical communication, and biomedical materials. Particularly important are various silicone-organic hybrid resin materials.

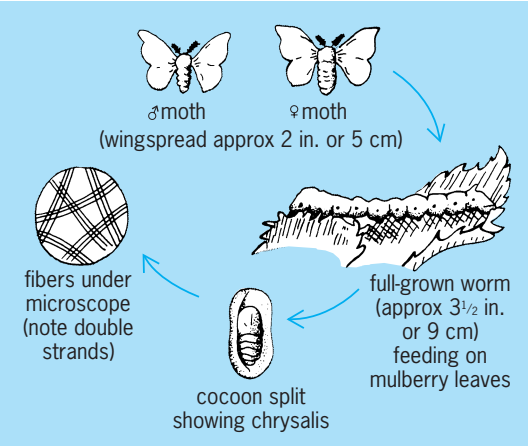
Julian Chojnowski

Bibliography. B. Arkles, *Organosilicon Chemistry IV*, ed. by N. Auner and J. Weis, pp. 606–612, Wiley, Weinheim, 2000; M. A. Brook, *Silicon in Organic, Organometallic and Polymer Chemistry*, Wiley, 2000; B. Hardman and A. Torkelson, *Encyclopedia of Polymer Science and Engineering*, ed. by H. F. Mark et al., Wiley-Interscience, vol. 15, pp. 265-270, 1989; W. Huang, Y. Huang, and Y. Yu, Synthesis of MQ silicone resins through hydrolytic condensation of ethyl polysilicate and hexamethyldisiloxane, *J. Appl. Polym. Sci.*, 70(9):1753–1757, 1998; S. Paul, *Surface Coatings Science and Technology*, pp. 200–216, Wiley, Chichester, 1985; A. Tomanek, *Silicones and Industry: A Compendium for Practical Use, Instruction and Reference*, pp. 71–75, Wacker Chemie, Munich, 1991.

Silk

The lustrous fiber produced by the larvae of silkworms; also the thread or cloth made from such fiber. The United States is the greatest importer and consumer of silk. Silk will probably always be prized by the consumer even though certain manufactured fabrics now have some qualities that were formerly possessed only by silk.

Japan, the first country to use scientific methods in cultivating the silkworm, has always ranked highest in the production of fine silk, although satisfactory types are made in other silk-producing countries. The domestication and cultivation of the silkworm, which began in Japan about 3000 B.C., requires extreme care and close supervision, and the reeling of the filament from the cocoons can be done successfully only by skilled operators whose training is the result of generations of experience.



Life cycle of the silkworm.

Life cycle of the silkworm. Since the discovery that the filament composing the cocoon of the silkworm can be unwound and constructed into a beautiful and durable fabric, silkworms have been bred for the sole purpose of producing raw silk. The production of cocoons for their filament is called sericulture. The cocoon of the silkworm is the second stage of development of the life cycle of *Bombyx mori*, a species which spins a thread of high quality. In sericulture, all four stages of the life cycle of this moth (see **illus.**) are important, because some of the better cocoons must be set aside to permit full development, thus supplying eggs for another hatching. By scientific breeding, silkworms may be hatched three times a year; under natural conditions, breeding occurs only once a year. The life cycle includes (1) the egg, which develops into the larva or caterpillar, the silkworm; (2) the silkworm, which spins its cocoon for protection and to permit development into the pupa or chrysalis; (3) the chrysalis, which emerges from the cocoon as the moth; and (4) the moth, the female of which lays eggs thus continuing the life cycle.

Within 3 days after emerging from the cocoons, the moths mate and the female lays 350–400 eggs. The adults seldom fly, do not eat, and live only a few days. Each healthy egg hatches into a grub, or larva, about $^1/_8$ in. (3 mm) long. The larva requires careful nurturing for approximately 20–32 days. During this period, the tiny worm has a voracious appetite, requiring five daily feedings of chopped mulberry leaves. After four changes of skin, or moltings, the worm reaches full growth in the form of a smooth, grayish-white caterpillar about $3^1/_2$ in. (9 cm) long. After about 6 weeks its interest in food ceases. It shrinks somewhat in size and acquires a pinkish hue, becoming nearly transparent. A constant restless rearing movement of the head indicates that the worm is ready to spin its cocoon. Clusters of twigs or straw are provided for this purpose.

The silkworm begins to secrete a proteinlike substance through its spinneret, a small opening under its jaws. With a bending motion a filament is spun around the worm in the form of the figure eight.

The silkworm is hidden from view within 24 h; in 3 days the cocoon is completed. It is about the size and shape of a peanut shell. The filament is in the form of a double strand or fibroin, which is held together by a gummy substance called sericin, or silk gum. Chemically, the silk fibroin and sericin are composed of approximately 95% protein and 5% wax, fats, salts, and ash. The liquid substance hardens immediately on exposure to the air. If left undisturbed, the chrysalis inside the cocoon develops into a moth within 2 weeks. To emerge, the moth must break through the top of the cocoon by excreting an alkaline liquid that dissolves the filament. As this cutting through damages the cocoon so that the filament cannot be unwound in one long thread, the growers terminate the life cycle at this point by a process known as stoving or stifling. The cocoons are heated to suffocate the chrysalis, but the delicate silk filament is not harmed. *See* LEPIDOPTERA.

Filature operations. The cocoons are delivered to a factory, called a filature, where the silk is unwound from the cocoons and the strands are collected into skeins. Some of the cocoons are produced scientifically in such factories. They are sorted according to color, size, shape, and texture, as all these affect the final quality of the silk.

After the cocoons have been sorted, they are put through a series of hot and cold immersions, because the sericin must be softened to permit the unwinding of the filament in one continuous thread. Raw silk consists of about 80% fibroin and 20% sericin. In this step only about 1% of the sericin is removed because this silk gum is a needed protection during the further handling of the delicate filament.

The process of unwinding the filament from the cocoon is called reeling. The care and skill used in the reeling operation prevent defects in the raw silk. As the filament of a single cocoon is too fine for commercial use, 3–10 strands are usually reeled at a time to produce the desired diameter of raw silk thread. The cocoons float in water, bobbing up and down as the filaments are drawn upward through porcelain eyelets and rapidly wound on wheels or drums while the operator watches to detect flaws. As the reeling of the filament from each cocoon nears completion, the operator attaches a new filament to the moving thread. Skilled operators have an uncanny ability to blend the filaments, always retaining the same diameter of the rapidly moving silk strand. The sericin acts as an adhesive, aiding in holding the several filaments together while they are combined to form the single thread.

The usable length of the reeled filament is from 1000 to 2000 ft (300 to 600 m). The remaining part of the filament is valuable raw material for the manufacture of spun silk.

The term reeled silk is applied to the raw silk strand that is formed by combining several filaments from separate cocoons. It is reeled into skeins, which are packed in small bundles called books, weighing 5–10 lb (2.3–4.6 kg). These are put into bales, ranging in weight from 135 to 145 lb (61 to 66 kg). In this form the raw silk is shipped to all parts of the world.

From the filature, the books of reeled silk go to the throwster where they are transformed into silk yarn, also called silk thread, by a process known as throwing. Persons engaged in this work are called throwsters. Silk throwing is analogous to the spinning process that changes cotton, linen, or wool fibers into yarn. Unlike the manufacture of these fibers, however, the manufacture of silk yarn does not include carding, combing, and drawing out, the usual processes for producing a continuous yarn. The raw silk skeins are sorted according to size, color, and length or quantity. Then they are soaked in warm water with soap or oil. This softening of the sericin aids in handling the thread. After being mechanically dried, the skeins are placed on light reels, from which the silk is wound on bobbins.

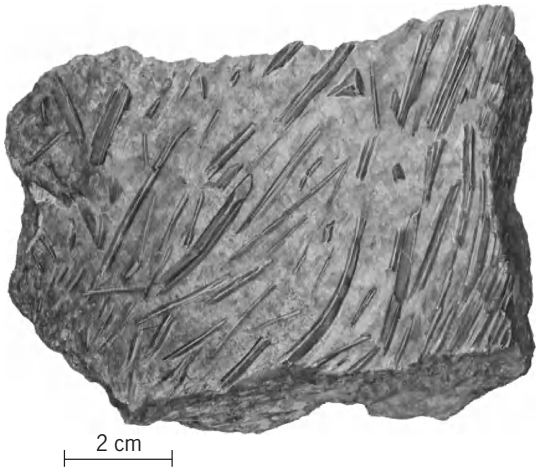
During this winding operation single strands may be given any desired amount of twist. If two or more yarns are to be doubled, they are twisted again in the same or in a reverse direction, depending on the kind of thread to be made. To equalize the diameter, the thread is run through rollers. It is then inspected and packaged ready for shipment to manufacturers for construction into fabric.

Wild silk. Wild or tussah silk may be distinguished from cultivated silk by its coarse, thick form, which appears flattened. Cultivated silk is a narrow fiber with no markings. Wild silk is a broader fiber with fine, wavy longitudinal lines running along its surface, giving it a dark hue under a microscope. *See* NATURAL FIBER.

M. David Potter

Sillimanite

A nesosilicate mineral of composition $Al_2O[SiO_4]$, crystallizing in the orthorhombic system, space group Pbnm, $a = 0.744$, $b = 0.760$, and $c = 0.575$ nanometers, with four formula units in the cell. Sillimanite commonly occurs in slender crystals or parallel groups, and is frequently fibrous, hence the synonym fibrolite (see **illus.**). There is one perfect



Sillimanite crystals from Chester, Connecticut. (*Specimen from Department of Geology, Bryn Mawr College*)

cleavage, luster is vitreous, color is brown, pale green, or white, hardness is 6–7 on Mohs scale, and the specific gravity is 3.23.

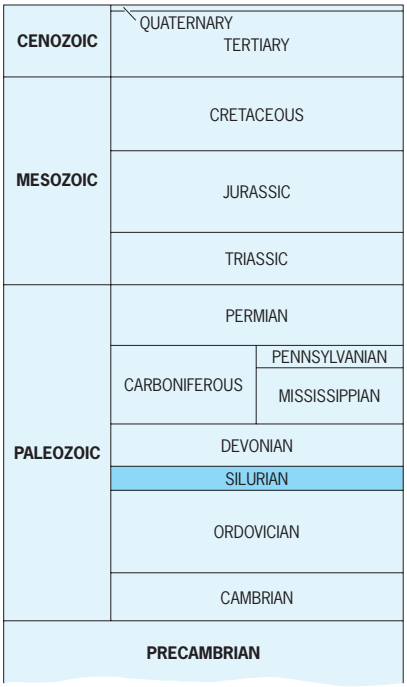
Sillimanite, andalusite, and kyanite are polymorphs of $Al_2O[SiO_4]$. In sillimanite one Al atom is coordinated by four and the other by six oxygens; in andalusite the Al atoms are five- and six-coordinate; and in kyanite all Al atoms are six-coordinate. The triple point where all three polymorphs coexist has been a source of much debate, but is probably close to 3.76 kilobars (376 megapascals) and 501°C (934°F), with kyanite the high-pressure phase. The three $Al_2O[SiO_4]$ polymorphs are important in assessing the metamorphic grade of the rocks in which they crystallized. Sillimanite occurs in high-temperature metamorphic argillaceous rocks. Common associated minerals are cordierite, biotite (a mica), and almandine (a garnet). In rocks which are poor in silica, sillimanite can coexist with corundum. *See* ANDALUSITE; KYANITE; SILICATE MINERALS.

Paul B. Moore

Bibliography. D. M. Kerrick, *Reviews in Mineralogy*, vol. 22: *The Al_2SiO_5 Polymorphs*, 1990.

Silurian

The third oldest period of the Paleozoic Era, spanning an interval from about 412 to 438 million years before the present. The Silurian system includes all sedimentary rocks deposited and all igneous and metamorphic rocks formed in the Silurian Period. Both the base and top of the Silurian have been designated by international agreement at the first appearances of certain graptolite species in rock sequences at easily examined and well-studied outcrops. *See* GEOLOGIC TIME SCALE.

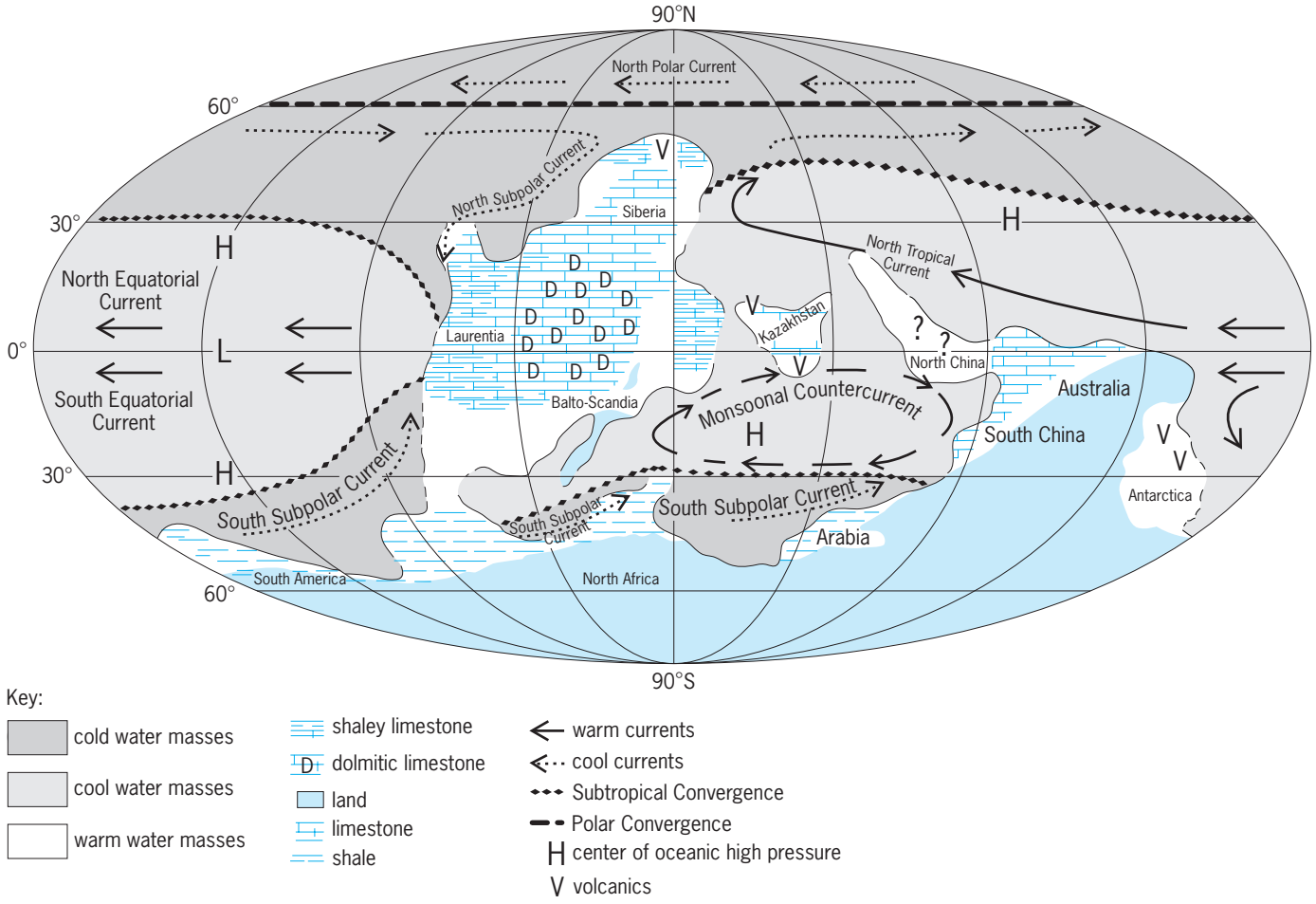


Divisions. The type area of the Silurian System is in adjoining parts of Wales and England. Based on studies of rocks and their contained faunas in that area, R. I. Murchison divided the Silurian System into three series. Each series is distinguished by a characteristic faunal association. Epochs are time intervals during which rocks of the corresponding series accumulated. From oldest to youngest, the three widely recognized divisions of the Silurian are Llandovery, Wenlock, and Ludlow. Analyses of faunas obtained from Silurian rocks led to recognition of a fourth and youngest series and epoch within the Silurian, the Pridoli. The Pridoli type area is in Bohemia, an area close to Prague, where marine faunas younger than those of the Ludlow and older than those of the Devonian have been obtained from outcrops. The stratotype or type section for the base of the Devonian is in the same area as rocks bearing typical Pridoli faunas.

Intervals with durations shorter than those of the epochs have been recognized in Britain. Three divisions of the Llandovery, two of the Wenlock, and two of the Ludlow are recognized on the basis of shelly (primarily brachiopod) and graptolite faunas. The divisions are termed stages. Ages are the time intervals during which rocks of the corresponding stages accumulated. Graptolites, the fossil remains of colonial, marine, planktic (floating) organisms, have been used to divide the Silurian in many parts of the world into intervals even shorter than those of the stages. These intervals are zones. Graptolite zonal successions have been developed in dark, graptolite-bearing shales of Silurian age in many parts of the world. Twenty-eight generally recognized graptolite zones have been proposed as divisions of the Silurian that had shorter durations than stages. Most of these zones may have had durations of about a million years, making separation of evolutionary and biogeographic developments within Silurian graptolite-bearing strata remarkably precise. Fifteen conodont zones are recognized as zonal divisions of the Silurian. They are most useful in carbonate successions. Organic-walled microfossil (chitinozoans and acritarchs) zonations have been developed that are especially useful in the high-latitude siliclastic successions. *See* GRAPTOLITHINA.

Paleogeography and lithofacies. Plate positions and plate motions during the Silurian significantly influenced the depositional environments, climates, and life of the period. Integration of data from remanent magnetism, distributions of reefs and other prominent successions of carbonate rocks, positions of shorelines, and positions of glacial deposits may be used to indicate certain features of Silurian paleogeographies and changes in them through the Silurian (see **illus.**). *See* DEPOSITIONAL SYSTEMS AND ENVIRONMENTS; FACIES (GEOLOGY); PALEOGEOGRAPHY; PALEOMAGNETISM; PLATE TECTONICS.

The most prominent feature of Silurian paleogeography was the immense Gondwana plate. It included much of present-day South America, Africa, the Middle East, Antarctica, Australia, and the Indian subcontinent. Numerous small plates lay near its margins.



Lithofacies of the Upper Silurian (Ludlow), modified from A. J. Boucot et al. (1968). Paleogeographical reconstructions after Scotese (1986). (*After Spec. Pap. Paleontol. 44, 1991*)

Some of the small plates close to northern or equatorial Gondwana included certain areas of North China, Tibet, Southeast Asia, Asiatic China, New Guinea, and New Zealand. Small plates that were in mid to low latitudes included the plates that today make up southern Europe and Florida, as well as small plates now found as fragments in the Alps, South China, and Tarim (in Asiatic China). The modern South American and African portions of Gondwana lay in high latitudes about the South Pole. That pole probably was positioned approximately in eastern South America. A large salient (landform projection) that was composed of areas of the Middle East, India-Pakistan, Antarctica, and Australia extended north across the Equator from modern eastern Africa into the Northern Hemisphere. The small plates that constitute modern Southeast Asia-Malaysia were within the Northern Hemisphere tropics.

During the Silurian, many plates continued the relative northward motion that had commenced during the mid-Ordovician. As a consequence of these motions, the Avalon plate (South Wales, southern England, and nearby continental Europe; and the Avalon Peninsula of Newfoundland and parts of maritime Canada) collided with the eastern margin of Laurentia (much of modern North America, Greenland, Scotland, and portions of northwestern Ireland and western Norway) in the latest Ordovician. As that collision developed, a large mountainous area formed. It was the source of siliclastics that comprise the Queenston deltaic deposits in Appalachian rock sequences. It was also the source of siliclastic sediments that accumulated in Avalon plate sites. Baltica (the Baltic states, Scandinavia, and eastern Europe east to the Urals) moved north across the Equator and collided with the Avalon and Laurentian plates during the early part of the Silurian. Late Silurian deposits of east-central Laurentia include significant thicknesses of salt and other evaporites. This stratigraphic record suggests that by the latter part of the Silurian, the eastern side of Laurentia was close to 30° south latitude. The plate fragments that today constitute much of southern Europe moved northward to approach, and perhaps begin to collide with, each other. Most of these plates moved into the tropics during the latter part of the Silurian. The core of one such plate, Perunica (modern Czech Republic), was the site of a large volcano for much of the Silurian. Potentially, that plate resembled a modern atoll because carbonates flank the volcanics. The Kazakh plate (modern Kazakhstan) may have been a set of atolls in the Silurian. Volcanic rocks

of different types are common there, as are carbonates. The plate seems to have included one or more tropical atolls. The Siberian plate (much of modern Asiatic Russia, Mongolia, and areas in northwestern China) moved northward through the Northern Hemisphere tropics. Portions of it may have entered the relatively temperate conditions north of 30° north latiude by the end of the Silurian. The northern part of modern Africa also moved northward during the Silurian, coming close to or even entering the tropics by the end of the period. *See* CONTINENTS, EVOLUTION OF.

Plate positions and plate motions as well as topographic features of the plates controlled depositional environments and lithofacies. These, in turn, significantly influenced organismal development and distributions.

Much of high-latitude Gondwana was the scene of continental glaciation in the latest Ordovician. Initially, the ice melted rapidly, leading to a rapid sea-level rise that continued from the latest Ordovician into the early Silurian. Central areas of Gondwana rose after the ice melted, presumably as a result of isostatic rebound. That rebound was followed by tectonic doming. The consequence of a rising central Gondwana is reflected in the steady northward and westward spread of continental and nearshore marine deposits away from land areas and across continental shelves during the Silurian. The Gondwanan land was the source for much of the highly organic-rich mudstone that characterizes North African and Middle Eastern Silurian strata. Isotopic ratios of strontium-87 to strontium-86 may be obtained from brachiopod shells and conodonts. In such analyses, a relatively higher proportion of strontium-87 to strontium-86 suggests that an old granitic source land is being eroded. Studies of the ratio of the strontium-87 to strontium-86 for Silurian brachiopods and conodonts are consistent with continued erosion of a relatively old granitic source during the Silurian. As the southern European and modern Alpine plates moved northward to enter the tropics, they became sites of, initially, cool-water carbonate deposition. That was followed by reef formation and extensive bahamian-type carbonate deposition. *See* CHEMOSTRATIGRAPHY; GLACIAL EPOCH; ISOTOPE; STRONTIUM.

Three brief glacial episodes have been recognized in the Llandovery stratigraphic record of the Amazon Basin. Each of these episodes took place essentially at the time of each of the boundaries between the stage divisions of the Llandovery Series. The South American stratigraphic record also suggests that a brief glaciation took place in the late Wenlock, at about the height of impact of the Baltic plate against Avalon-Laurentia. The South American Silurian record of glaciation suggests that the modern eastern part of South America lay close to or across the South Pole at that time.

Plates that were within the tropics were sites of extensive carbonate deposition. Reefs developed along and near the margins of most plates that were within the tropics.

Silurian Northern Hemisphere plates, other than a portion of Siberia, are not known north of the Northern Hemisphere tropics. Presumably, nearly all of the Northern Hemisphere north of the tropics was ocean throughout the Silurian.

Ocean circulation. Absence of plates bearing continental or shallow shelf marine environments north of about 45° north latitude indicates that ocean circulation in most of the Silurian Northern Hemisphere was zonal. Ocean currents were relatively strong and flowed from east to west north of 60° north. Major ocean currents between 30 and 60° north flowed from west to east. Surface currents would have been deflected south along the west side of Siberia, and upwelling conditions could have formed in that area.

Ocean surface currents in the tropics would have been influenced strongly by the prevailing westerlies. Primary surface circulation from the Equator to 30° north and south of it would have been from east to west. The large peninsulalike salient of Gondwana would have deflected currents northerly along it in the Northern Hemisphere and southerly along it in the Southern Hemisphere. North China's eastern shores would have been sites of upwelling. The presence of richly fossiliferous coralline limestones and volcanics there is consistent with surface-water turbulence and upwelling conditions along the margins of volcanic islands. Much the same conditions appear to have prevailed along the margins of the Kazakh plate.

The large size of the Gondwana plate and the presence of land over much of it would have led to development of seasonal monsoon conditions. Monsoonal conditions would have led to seasonal reversal of major surface circulation adjacent to the land, as is seen in modern India. In the austral summer–Northern Hemisphere winter, winds would have blown offshore from the warm land and resulted in an east to west current flow near 30° south and a west to east surface current near the Equator. In the Northern Hemisphere summer and austral winter, surface water flow west of the lands on the Gondwana plate would have reversed. These seasonal reversals in surface current directions could have created long-term, year-long upwelling in a pattern similar to that observed in modern oceans off the Somali coast. Consequently, intense upwelling across northern Africa could have been generated during the Silurian in a pattern similar to that in the modern Arabian Ocean. That upwelling generated vast quantities of organic matter which became organic-rich shales that characterize the North African Silurian. In general, surface circulation west of tropical Gondwana would have been east to west. However, surface circulation between the several plates that drifted into the tropics as well as those that moved closer to each other than they had been previously would have created many small east-to-west flowing gyres in the tropics. Modest-to-strong upwelling along many plate margins was likely as a consequence. The occurrence of numerous long-studied reefs, such as those on the island of Gotland off the Swedish coast, close to or along

tropical plate margins is consistent with plate margin upwelling.

Surface circulation south of 30° south would have hit the western side of Laurentia and flowed generally northward along it to about 30°, at which latitude the surface currents probably turned to flow east to west between 30° south and the Equator. *See* CORIOLIS ACCELERATION; PALEOCEANOGRAPHY; UPWELLING.

Climate and hydrology. Collision of the Avalonian and Laurentian plates in the latest Ordovician coincides with development of the Southern Hemisphere continental glaciation. Erosion of the land area formed at the Avalon-Laurentian plate collision generated a large volume of coarse to fine-grained siliclastic materials. Such extensive erosion could have reduced the atmospheric concentration of carbon dioxide, creating climate conditions cool enough to allow glaciation to develop. Although atmospheric carbon dioxide concentration probably was significantly greater during the Silurian than it is today, solar heat coming to the Earth was about 4–5% less than today. Thus, depression of the atmospheric carbon dioxide content could have been enough to allow glaciation. Reduction in the extent and height of the land being eroded could have resulted in increased atmospheric carbon dioxide concentration, ending the conditions suited to generating glaciers by the latter part of the Wenlock.

That part of South America (modern eastern South America) near the South Pole for the early part of the Silurian was not just cold, but also the site of as many as four brief glacial episodes. These episodes took place during the time that the Avalon and Baltica plates collided with each other and with the Laurentian plate.

Rainfall would have been significant on the mountainous lands formed along the Avalon-Laurentian plate collision boundary because that land lay within the prevailing winds. *See* PALEOCLIMATOLOGY.

Atmosphere. Analyses of major rock suites have led to the suggestion that the Silurian global air temperature was about 9°F (5°C) warmer than today. The postulated global air temperature is consistent with an estimated carbon dioxide content of three to four times that of the modern atmosphere. Oxygen content of the Silurian atmosphere has been estimated to have been about two-thirds to three-fourths the present content. A lesser oxygen concentration in the atmosphere would have resulted in less oxygen available to be stirred into the oceans. A lesser oxygen concentration in the surface oceans coupled with a warmer global temperature meant that less oxygen was available to marine organisms in the Silurian seas than is available to modern oceanic organisms. Silurian marine organisms, especially those living at depths significantly below the surface, must have survived on significantly less oxygen than do modern organisms. Furthermore, the ocean oxygen minimum zone would have attained shallower depths than it does in modern oceans. Oxygen content in ocean waters may have declined to near zero by 330 ft (100 m) in tropical waters. Atmospheric conditions would have had a significant influence on the distributions of marine benthic dwellers. *See* ATMOSPHERE, EVOLUTION OF.

Land life. Both nonvascular and vascular plants continued to develop in land environments following their originations in the early mid-Ordovician. Many of these Silurian plants were mosslike and bryophytelike. Plants with vascular tissues had developed in the mid-Ordovician. These plants continued their spread in terrestrial environments during the Silurian. Psilophytes assigned to the genus *Cooksonia* were relatively widespread in Late Silurian terrestrial environments. The probable lycopod (club moss) *Baragwanathia* apparently lived in nearshore settings in modern Australia during the latter part of the Silurian.

Silurian land life also included probable arthopods and annelid worms. Fecal pellets of wormlike activity have been found as well as remains of centipede-, millepede-, and spiderlike arthropods. *See* PALEOBOTANY; PALEOECOLOGY.

Marine life. Shallow marine environments in the tropics were scenes of rich growths of algae, mat-forming cyanobacteria, spongelike organisms, sponges, brachiopods, bryozoans, corals, crinoids, and ostracodes. Nearshore marine siliclastic strata bear ostracodes, small clams, and snails and trilobites. Certain nearshore strata bear the remains of horseshoe-crab-like eurypterids. Some of them may have been significant predators. *See* ALGAE; BRACHIOPODA; BRYOZOA; CRINOIDEA; CYANOBACTERIA; OSTRACODA; TRILOBITA.

Fish are prominent in a number of Silurian nearshore and some offshore marine environments. Jawless armored fish of several kinds occur in Silurian strata. These fish include many species of thelodonts that had bodies covered with minute bony scales, heterostracans, and galeaspids that had relatively heavily armored head shields, and anaspids that possessed body armor consisting of scales and small plates. Jawed fish were relatively rare in the Silurian. They were primarily spiny sharks or acanthodians. As well, there are remains of true sharklike fish and fish with interior bony skeletons (osteichthyes) in Late Silurian rocks. *See* ANASPIDA; HETEROSTRACI; OSTEICHTHYES; THELODONTIDA.

The acanthodians appear to have been relatively common in the latter part of the Silurian. The oldest known placoderms were found in Silurian strata in South China. Fish began to diversify in the latter part of the Silurian in many shelf sea environments. The planktic colonial marine graptolites are the prominent organism found in rocks that formed under anoxic or near-anoxic waters. Their remains are most plentiful in rocks that accumulated on the outer parts of shelves and in basins of the Silurian. *See* ANOXIC ZONES.

The extinct microfossil group, the conodonts, were relatively common in many carbonates deposited in shelf seas. Small, slender shells of squidlike cephalopods occur in many shelf-sea rock suites, including certain of the black, organic-rich graptolite-bearing sequences. These cephalopods appear to

have been nektic in habit. *See* CEPHALOPODA; CONODONT.

Biogeography. Both land plants and marine animals were distributed in patterns reflective of latitudinal temperatures during the Silurian. Those organisms living along the margins of the Gondwana plate in the general position of 30 to 45° south latitude constituted one floral and faunal realm, the Malvinokaffric, which persisted throughout the Silurian. Organisms that typify it were adapted to cool climates.

Land plants and marine organisms living during the early part of the Silurian (Llandovery into Wenlock) in warm temperate to tropical conditions outside of the cool temperate to polar Malvinokaffric Realm were essentially cosmopolitan. This distribution reflected sparsity of tropical marine life in the tropics after the major Late Ordovician into earliest Silurian extinctions among marine organisms. Both land plants and marine organisms were increasingly provincial during the latter part of the Silurian. Marine bottom-dwelling invertebrates living on shelves of plates in the tropics were noticeably provincial by the close of the Silurian. The northern part of the Siberian plate was characterized during the latter part of the Silurian by a unique association of shelled marine invertebrates of which the brachiopod *Tuvaella* is characteristic. Laurentia-Baltoscanian shallow marine environments were inhabited by marine bottom-dwelling faunas, of which the brachiopod *Salopina* has been selected to give its name. The Kazakh plate shelf seas had a unique, endemic marine invertebrate fauna.

Nearshore marine fish had distributions similar to those of the marine bottom-dwelling invertebrates. Silurian fish from South China are an exception. They constituted a distinct faunal province.

Among Late Silurian land plants, the truly tracheophytic *Baragwanathia* known from eastern Australia was distinct from coeval land plants known in other parts of the tropics. Although sparse, tropical land plants appear to have had distributions similar to those of the marine organisms. This evidence suggests that plate positions and the ocean and wind circulation patterns they influenced were major factors in the distribution of life during the latter part of the Silurian. *See* BIOGEOGRAPHY.

Economic resources. Silurian dark, organic-rich rocks form one of the six prominent suites of petroleum source rocks known. The six suites have provided more than 90% of the world's known oil and gas reserves. Silurian source rocks have generated approximately 9% of the world's reserves. Most of the Silurian source beds for petroleum and gas are those that lay under the upwelling waters along the northern and northwesterly margins of the Gondwana plate. Other known prominent Silurian source rocks are those that accumulated under upwelling waters on the eastern side of the Baltoscanian-Avalonian plates and those that accumulated under upwelling waters along the westerly side of the Laurentian plate.

Sedimentary iron ores accumulated in tidal wetlands on the Avalonian, especially the West Avalon, plates during the Llandovery. At that time, the Avalonian plates lay south of the tropics.

Silurian carbonate rocks have been quarried in many parts of the world for building stone and for the raw material for cement. Quarries in the area of the type Wenlock, at Wenlock Edge in Britain, have yielded richly fossiliferous limestones used in construction. Late Silurian orthoceroid cephalopod-bearing limestones interbedded with the black, graptolite-bearing shales in southern Europe have been quarried for a spectrum of ornamental uses, including tables and decorative paneling. Silurian salt deposits have been mined extensively in the eastern United States. These salts were a major economic resource in the 1800s. Silurian carbonate rocks in Nevada bear large quantities of gold, which occurs as tiny flakes. Today the amount of gold recovered from the Nevada sites has made this area one of the two or three largest gold producers in the world. Silurian rocks have played significant roles in a number of local and regional economies. *See* FOSSIL; PALEOZOIC; PETROLEUM GEOLOGY.

William B. N. Berry

Bibliography. M. G. Bassett, P. D. Lane, and D. Edwards (eds.), *The Murchison Symposium: Proceedings of an International Conference on the Silurian System*, Spec. Pap. Palaeontol. 44, Palaeontological Association, London, 1991; W. B. N. Berry and A. J. Boucot, *Correlation of the African Silurian Rocks*, GSA Spec. Publ. 147, 1973; W. B. N. Berry and A. J. Boucot, *Correlation of the North American Silurian Rocks*, GSA Spec. Publ. 102, 1970; J. Gray and W. Shear, Early life on land, *Amer. Sci.*, 80:444–456, 1992; J. C. Gutierrez-Marco and Rabano (eds.), *Proceedings of the 6th International Graptolite Conference of the GWPG (IPA) and the 1998 Field Meeting of the International Subcommission on Silurian Stratigraphy (ICS-IUGS)*, Instituto Tecnologico Geominero de Espana, Temas Geologico-Mineros, 23, 1998; C. H. Holland and M. G. Bassett (eds.), *A Global Standard for the Silurian System*, Nat. Mus. Wales Geol. Ser. 10, 1989; E. Landing and M. Johnson (eds.), *Silurian Cycles: Linkages of Dynamic Stratigraphy with Atmospheric, Oceanic, and Tectonic Changes*, New York State Mus. Bull. 491, 1998; G. T. Moore et al., A paleoclimate simulation of the Wenlockian (Late Early Silurian) world using a general circulation model with implications for early land plant paleoecology, *Palaeogeog. Palaeoclimatol. Palaeoecol.*, 110:115–144, 1994.

Siluriformes

The catfishes, a large and highly diverse order of the Ostariophysi. Another ordinal name commonly used is Nematogathi, from the Greek *nematos*, thread, and *gnathos*, jaw, a descriptive name referring to the long barbels around the mouth of most catfishes. Catfishes differ from other members of the Ostariophysi in having a more complex Weberian apparatus and in lacking intermuscular bones, as well as lacking subopercle, symplectic, or intercalar bones in the head. They are further characterized by either

lacking scales (in which case they are covered with tough skin) or being covered with overlapping bony plates; having vomer, pterygoid, and palatine bones that are usually toothed; having maxillae that are rudimentary and toothless; lacking a parietal bone; usually having one to four pairs of barbels (some members of the family Loricariidae lack visible barbels); having (usually) an adipose fin and having pectoral and dorsal fins that are each preceded by a strong spine in most species.

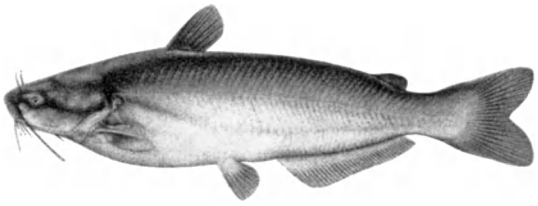
The order includes about 2800 species in 35 families and is represented on all continents, albeit only by fossils in Antarctica. Catfishes are chiefly freshwater species; Ariidae (tropical waters of both hemispheres) and Plotosidae (Indo-Pacific) are considered marine families, but they too include freshwater species. Aspredinidae (tropical South America), considered a freshwater family, has some brackish species. Twelve families of catfishes are known only in Central and South America; three families are restricted to Africa; five are endemic to eastern Asia; three are shared by Africa and Asia; two are Indonesian, one Eurasian; one ranges from India to southeast Asia;, and one, the Ictaluridae, is endemic to North America.

The following families were selected to represent the world of catfishes and give the reader some insight into their diversity.

Ictaluridae. Ictaluridae is the largest family of fishes indigenous to North America, where their native range is eastward from the Continental Divide to the Atlantic and Gulf drainages and from southern Canada southward to Guatemala. Characters that distinguish the ictalurids are four pairs of barbels, one pair nasal, one pair of maxillary, and two on the chin; naked skin; usually six soft dorsal fin rays; dorsal and pectoral fins with spines (dorsal spine absent in *Prietella*); and an adipose fin free or adnate.

Three of its seven genera (*Prietella, Satan*, and *Trogloglanis*) are represented by blind, subterranean species. The bullheads (*Ameiurus*) with seven species, the catfish (*Ictalurus*) [see **illustration**] with nine species, and the flathead catfish (*Pylodictis*) are large-sized members of the family. Pond-reared channel catfish (*I. punctatus*) support a large food industry in the southeastern United States. The 25 species in the genus *Noturus* are pygmies among the catfishes, the largest hardly exceeding 150 mm (6 in.) and most being far smaller than that. Madtoms spend rather secretive lives in small streams, where they hide by day and become active at night. They have a poison gland at the base of the pectoral fin spine enabling them to deliver a painful sting.

Clariidae. These are the so-called air-breathing catfishes or labyrinth catfishes of Africa, India, southeastern Asia, western Indonesia, and the Philippines. They are distinguished by having a very long dorsal fin not preceded by a spine; having four pairs of barbels; and lacking pectoral and pelvic fins in some species. Most interesting are the labyrinthic organs in the upper branchial cavities, which greatly increase the surface area for exchange of respiratory gases, thus allowing the fish to breathe atmospheric air. Not all clariids are so equipped, but those that are can move short distances over land, thus the "walking catfish." The ability to move over land enhances dispersal, as demonstrated by *Clarias batrachus*, a species accidentally introduced in Florida.



Blue catfish (*Ictalurus furcatus*). (*Drawing © Joseph R. Tomelleri*)

Ariidae. Ariid catfishes are in general appearance similar to the genus *Ictalurus* (Ictaluridae), but easily distinguished from them by the absence of nasal barbels. They are represented in all tropical and subtropical seas of the world, and some species enter freshwater on a regular basis, while others occupy freshwater throughout their lives. Most species of the family practice oral incubation (eggs are carried in the mouth until hatching). The male carries the eggs throughout their development, and the young are not released until they can fend for themselves. The mature eggs are the largest known among teleost fishes, measuring 14–18 mm (0.5–0.7 in.) in diameter. During the incubation period the oral cavity is enlarged by depressing the hyoid bones, making room for several dozen 40-mm (1.6 in.) young.

Doradidae. Doradids are called the thorny catfishes. They are freshwater species of South America, primarily in Guiana, Brazil, and Peru. The body, with few exceptions, has a lateral row of bony plates, each usually armed with a spine; there are three pairs of barbels, nasal barbels being absent; the mandibular barbel may be branched; pectoral and dorsal fin spines are stout and serrated; the skull extends as a bony plate to the base of the dorsal fin. These are the "talking catfishes." By movements of the pectoral spines or by vibrations of the swim bladder, the fishes make a growling sound. The smaller species of the family, along with the similar families Callichthyidae and Loricariidae, also from South America, are popular aquarium fishes.

Trichomycteridae. Members of this family occur in Costa Rica, Panama, and throughout South America. They are identified by the following features: elongate body lacking scales; chin barbels usually absent, nasal barbels usually present; dorsal and anal fins short; pelvic fins absent in some species; usually no adipose fin; and posteriorly directed spines on the opercle. These characteristics adapt the fish for a parasitic life. Indeed, they do enter the gill cavity of larger fish, where they nip the gills and feed on the ensuing blood. The infamous candiru even parasitize humans, by entering the urethra in a rheotaxic response to a stream of urine. Once the candiru enters the urethra, the opercular spines prevent its removal and surgery is required.

Not only are catfishes diverse in morphology,

biology, and life histories, but also they vary tremendously in size. Adults of the little madtoms (*Noturus*) of North America may weigh only a few grams, while the Mekong giant catfish (*Pangasianodon gigas*, family Pangasidae) has been reported by the Thai Inland Fisheries Department to weigh up to 293 kg (646 lb). *See* ACTINOPTERYGII; CYPRINIFORMES; OSTEICHTHYES. Herbert Boschung

Bibliography. T. M. Berra, *An Atlas of Distribution of Freshwater Fish Families of the World*, University of Nebraska Press, Lincoln, 1981; J. G. Lundberg, The comparative anatomy of the toothless blindcat, *Trogloglanis pattersoni* Eigenmann, with a phylogenetic analysis of the ictalurid catfishes, *Misc. Publ. Mus. Zool. Univ. Michigan*, 163:1–85, 1982; J. G. Lundberg, The phylogeny of ictalurid catfish: A synthesis of recent work, pp. 392–420 *in* R. L. Mayden, (ed.), *Systematics, Historical Ecology, and North American Freshwater Fishes*, Stanford University Press, Stanford, California, 1992; J. S. Nelson, *Fishes of the World*, 3d ed. Wiley, New York, 1994; W. R. Taylor, A revision of the catfish genus *Noturus* Rafinesque with an analysis of higher groups in the Ictaluridae, *Bull. U.S. Nat. Mus.*, 282:1–315, 1969.

Silver

A chemical element, Ag, atomic number 47, atomic mass 107.868. It is a gray-white, lustrous metal. Chemically it is one of the heavy metals and one of the noble metals; commercially it is a precious metal. There are 25 isotopes of silver with atomic masses ranging from 102 to 117. Ordinary silver is made up of the isotopes of masses 107 (52% of natural silver) and 109 (48%). *See* PERIODIC TABLE.

1																	18
1 H	2											13	14	15	16	17	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

Although silver is the most active chemically of the noble metals, it is not very active in comparison with most other elements. It does not oxidize as iron does when it rusts, but it reacts with sulfur or hydrogen sulfide to form the familiar silver tarnish. Electroplating silver with rhodium prevents this discoloration. Silver itself does not react with dilute nonoxidizing acids (hydrochloric or sulfuric acids) or strong bases (sodium hydroxide). However, oxidizing acids (nitric or concentrated sulfuric acids) dissolve it by reaction to form the unipositive silver ion, Ag^+. The Ag^+ ion is colorless, but a number of silver compounds are colored because of the influence of their other constituents.

Silver is almost always monovalent in its compounds, but an oxide, fluoride, and sulfide of divalent silver are known. Some coordination compounds of silver, also called silver complexes, contain divalent and trivalent silver. Although silver does not oxidize when heated, it can be oxidized chemically or electrolytically to form silver oxide or peroxide, a strong oxidizing agent. Because of this activity, silver finds considerable use as an oxidation catalyst in the production of certain organic materials.

Soluble silver salts, especially $AgNO_3$, have proved lethal in doses as small as 2 g. Silver compounds may be slowly absorbed by the body tissues, with a resulting bluish or blackish pigmentation of the skin (argyria).

Silver is a rather rare element, ranking 63rd in order of abundance. Sometimes it occurs in nature as the free element (native silver) or alloyed with other metals. For the most part, however, silver is found in ores containing silver compounds. The principal silver ores are argentite, Ag_2S, cerargyrite or horn silver, AgCl, and several minerals in which silver sulfide is combined with sulfides of other metals; stephanite, $5Ag_2S \cdot Sb_2S_5$; polybasite, $9(Cu_2S, Ag_2S) \cdot (Sb_2S_3, As_2S_3)$; proustite, $3Ag_2S \cdot As_2S_3$; and pyragyrite, $3Ag_2S \cdot Sb_2S_3$. About three-fourths of the silver produced is a by-product of the extraction of other metals, copper and lead in particular. *See* SILVER METALLURGY.

Pure silver is a white, moderately soft metal (2.5–3 on Mohs hardness scale), somewhat harder than gold. When polished, it has a brilliant luster and reflects 95% of the light falling on it. Silver is second to gold in malleability and ductility. Its density is 10.5 times that of water. The quality of silver, its fineness, is expressed as parts of pure silver per 1000 parts of total metal. Commercial silver is usually 999 fine. Silver is available commercially as sterling silver (7.5% copper) and in ingots, plate, moss, sheets, wire, castings, tubes, and powder.

Silver, with the highest thermal and electrical conductivities of all the metals, is used for electrical and electronic contact points and sometimes for special wiring. Silver has well-known uses in jewelry and silverware. Silver compounds are used in many photographic materials. In most of its uses, silver is alloyed with one or more other metals. Alloys in which silver is an ingredient include dental amalgam and metals for engine pistons and bearings. *See* PHOTOGRAPHIC MATERIALS; SILVER ALLOYS. William E. Cooley

Bibliography. F. A. Cotton et al., *Advanced Inorganic Chemistry*, 6th ed., Wiley-Interscience, 1999; J. R. Davis (ed.), *Metals Handbook: Desk Edition*, 2d ed., ASM International, 1998; D. R. Lide, *CRC Handbook of Chemistry and Physics*, 85th ed., CRC Press, 2004.

Silver alloys

Combinations of silver with one or more other metals. Pure silver is very soft and ductile but can be hardened by alloying. Copper is the favorite hardener and normally is employed in the production of sterling silver, which must contain a minimum of 92.5% silver, and also in the production of coin silver.

In the United States coin silver contained 90% silver, the balance being copper. To reduce the consumption of silver in the coinage, the 10-cent and 25-cent pieces were replaced by a composite material with cupro-nickel layers outside a core of copper. The density and electrical conductivity of the composite are such that the coins are accepted by coin-operated mechanisms. The 50-cent piece was made of outer layers of a silver-copper alloy with a core of copper silver. This peculiar combination must be ascribed to political metallurgy.

Pure silver melts at 961.93°C (1763.5°F) on the International Practical Temperature Scale of 1968, but this is lowered by the addition of copper to a minimum of 779°C (1434°F) at 28% copper. This silver-copper eutectic and modifications containing other elements such as zinc, tin, cadmium, phosphorus, or lithium are widely used for brazing purposes, where strong joints having relatively good corrosion resistance are required. Where higher strengths at elevated temperature are required, silver-copper-palladium alloys and other silver-palladium alloys are suitable. The addition of a small amount of silver to copper raises the recrystallizing temperature without adverse effect upon the electrical conductivity. For this reason, copper containing 10–20 oz (280–560 g) of silver per ton is desirable for commuter bars and armature windings in certain dc machinery.

An unusual use for silver is the 25% silver—75% palladium alloy used as a diffusion septum transmitting only hydrogen. The high electrical and thermal conductivities of pure silver, slightly exceeding those of copper, along with its resistance to oxidation, have led to the large use of silver for electrical contacts. Because of the tendency of silver to form sulfide films, it is desirable to employ voltages in excess of about 12 V and reasonable pressures in using such contacts. Frequently, the silver for contact is alloyed with 10% copper or with a small amount of cadmium, or, better still, with cadmium oxide, which improves the behavior of the contact material under many conditions. Cadmium alloyed with silver raises electrical resistivity. This effect, undesirable for electrical contacts, is reduced if the cadmium is present as the oxide. Oxide content materials may be made by powder metallurgy methods or by heating a silver cadmium alloy in air for some hours during which oxygen diffuses into the alloy and converts the cadmium to the oxide.

Noble-metal alloys. Silver may be alloyed with gold or palladium in any ratio, producing soft and ductile alloys; certain of these intermediate alloys are useful for electrical contacts, where resistance to sulfide formation must be achieved. This requires about 40–50% palladium or slightly more gold. Sterling and coin silver cost slightly less. In an early effort to reduce the cost of silver articles, the technique of bonding silver to copper and rolling the composite material into sheets was developed in England, particularly in the vicinity of Sheffield; the product was therefore known as Sheffield plate. However, with the development of the method of electrodepositing silver from a double cyanide solution by G. R. and H. Elkington in 1844, the production of Sheffield plate almost disappeared and was replaced by electroplated silver. This was the first commercial use of electroplating, and basically the same solution is still employed. The Elkingtons found that the most satisfactory base metal upon which silver could be plated was the ancient Chinese alloy known as paktong, which was found to contain nickel, copper, and zinc. It is still made under the general name of German silver or nickel-silver and still constitutes the most satisfactory base material. *See* ELECTROPLATING OF METALS.

Technical alloys. Silver is a useful component for high-duty bearings in aircraft engines, where it may be overlaid with a thin layer of lead and finally with a minute coating of indium. *See* METAL COATINGS.

The fact that silver is readily oxidized and reduced is a limitation to its use for certain electrical purposes, particularly where silver conductors are in contact with vulcanized fiber or similar materials and are subjected to a reasonable voltage gradient. Under these circumstances, the silver tends to migrate and develop conducting paths within the insulation which are difficult to detect.

Specially developed alloys of silver with tin, plus small percentages of copper and zinc in the form of moderately fine powder, can be mixed with mercury to yield a mass which is plastic for a time and then hardens, developing relatively high strength despite the fact that it contains about 50% mercury. This material was developed specifically for dental use and is generally known as amalgam, although the term amalgam actually includes all the alloys of mercury with other metals. Dental amalgam is widely used as a tooth-filling material and is a useful dental material. It possesses some limitations in strength and also discolors in the mouth, but silver has some germicidal effect which probably is helpful in this instance in preventing further decay, which is particularly important in children's teeth. *See* AMALGAM.

Silver is a component in many of the colored gold alloys used not only for jewelry but also for dentistry; when it is added to the gold-copper alloys or the palladium-copper alloys, a large improvement results in the strengths obtainable by heat treatment. The addition of platinum or palladium to some of these alloys further augments their strength or hardenability and also renders them more suitable for rubbing electrical contacts. The softness of silver, its insolubility in iron, and its freedom from oxidation at high temperatures make silver powder useful in preventing the threads in high-temperature bolting from sticking at temperatures up to about 650°C (1200°F). In the absence of silver, such bolts may bond together after a short time so that they cannot be unscrewed.

Where silver is to be used at high temperatures, the pure metal should be employed; it has been so used for electrical windings in motors operating as high as 500°C (930°F). At moderate temperatures, silver coatings are frequently useful in electrical devices to ensure good electrical contact in bolted joints and to provide a low-resistivity coating in equipment which will be operated at high frequency where the skin conduction is of primary importance. Thin silver or gold electroplate also may be applied to copper and copper parts to facilitate soldering, particularly if the parts are to be stored for some time and may become oxidized and difficult to solder.

Because silver does not oxidize on heating, it is used in applying electrically conductive coatings to ceramics. A paste containing finely divided silver plus certain additives is applied to the ceramic. Upon heating to redness, the silver coating becomes firmly bonded to the ceramic. With care, the product can be soldered; the method is used in making connections to ceramic capacitors and similar devices. Similar coatings are used for decorating glassware. These may be coated with additional silver by electrodeposition and finally with rhodium to prevent tarnishing.

Because of its resistance to acetic acid and many other organic materials, as well as to alkalies, silver-lined equipment finds use in the chemical industry—in autoclaves, piping, and similar pieces of equipment, some of it very large. In high-temperature applications, the high solubility of oxygen in silver and the rapid diffusion of oxygen through it must be recognized. Silver, alloyed with a base metal, will suffer internal oxidation when heated in air, and a silver coating applied to a base metal, such as iron, affords no useful protection at high temperatures because the oxygen will go through the silver, oxidizing the iron at the interface. Furthermore, silver will be weakened along grain boundaries if heated in an oxidizing atmosphere and then in a reducing atmosphere, particularly if the latter contains hydrogen. *See* SILVER; SILVER METALLURGY. Edmund M. Wise

Bibliography. A. Butts (ed.), *Silver, Economics, Metallurgy and Use*, 1967, reprint 1975; R. P. Elliot (ed.), *Constitution of Binary Alloys: First Supplement*, 1965, reprint 1989; M. Hansen, *Constitution of Binary Alloys*, 2d ed., 1988; W. Hume-Rothery and R. E. Smallman, *The Structure of Metals and Alloys*, 1988; F. A. Shunk (ed.), *Constitution of Binary Alloys: Second Supplement*, 1969, reprint 1986.

Silver metallurgy

The art and science of extracting silver metal economically from various ores, and the reclamation of silver from the myriad types of industrial processes or scrap produced therefrom. It includes all processes of separating silver from its ores, alloys, and solutions, as well as the smelting, refining, and working of the metal and its alloys and compounds. It deals with the technical application of the chemical and physical properties of silver, to its concentration, extraction, purification, alloying, working, and compounding to meet the requirements of ever-expanding technical needs.

Silver is widely distributed in nature and occurs in both native and combined forms. Although one of the most important ores of silver, tetrahedrite ($Cu_8Sb_2S_7$ with silver sulfide in chemical combination), is hypogene in origin, native silver, argentite (Ag_2S), and cerargyrite (AgCl) are distinctly supergene; and pyrargyrite (Ag_3SbS), stephanite (Ag_5SbS_4), polybasite (Ag_9SbS_6), and proustite (Ag_3AsS) are probably hypogene, having developed as a result of supergene sulfide enrichment. In a few cases, ores are treated for silver only, or for combined gold and silver values. However, at least three-fourths of the world silver production is obtained as a by-product of base-metal ores, mainly those of copper, lead, and zinc. *See* ACANTHITE; CERARGYRITE; PYRARGYRITE; TETRAHEDRITE.

Extractive processes. Silver extraction has been effected by a number of processes. A few of the important historical processes, as well as modern-day practice, are discussed below.

Amalgamation. This is a process of separating metal from ore by alloying the metal with mercury. The Patio process was introduced in Pachuca, Mexico, in 1557. The silver ore was finely ground, mixed with water to form a thin mud, and then treated with salt, copper sulfate, and mercury. The entire mass was then thoroughly mulled by walking horses or mules through the mixture. The resulting silver amalgam was retorted to separate the silver and mercury. Large amounts of silver were produced by this method from the time of its inception until the end of the nineteenth century, when it was superseded by other methods. Other amalgamation processes of importance are the Tintin, Cazo, Kröhnke, Washoe, Combination, Boss, Barrel or European, Pan or Reese-River, and Franke-Tina, all of which were derived from the basic Patio process.

Chloridization. Chloridization is a method of producing silver by precipitating it from aqueous solutions. In chloridization methods the silver contained in ores or metallurgical products is first converted by means of a chloridizing roast into a compound which is soluble in water or in certain aqueous solutions. The silver is then precipitated as an insoluble compound by suitable reagents and the precipitate worked for the metal. The soluble silver compounds are in the form of chloride, sodium thiosulfate solution, sulfate, or the double cyanide of silver and potassium. Principal chloridization processes are the Augustin process, which uses brine as a solvent; the Patera process, in which sodium thiosulfate is used; the Kiss process, in which calcium thiosulfate is used; the Russell process, in which silver metal or sulfide is brought into the solution with sodium-copper thiosulfate; and the Ziervogel process, in which the silver is converted to the sulfate and dissolved in hot water.

Cyanidation. This process differs only in minor details, such as strength of the solution and time of treatment, from the cyanide process used in the recovery of gold. Further, it differs from the chloridization process described above in that no preliminary metallurgical treatment other than fine grinding is required, as the cyanide is capable of dissolving

metallic silver and, to a greater or lesser extent, all the silver compounds occurring in the various ores. *See* GOLD METALLURGY.

Pattinson lead process. Although the Pattinson lead process has been superseded by the Parkes process described below, it is still occasionally used. The Pattinson process is based on the fact that, if molten argentiferous lead is slowly cooled, crystals of silver-poor lead separate from the still molten portion which is correspondingly enriched in silver. If the richer liquid portion is separated, it can again be separated into a poorer solid portion and a still richer liquid alloy. This operation can be repeated until the silver is concentrated to approximately 750 oz/ton (24,000 g/metric ton), at which point no further separation is possible. Similarly, the portions poor in silver may be successfully treated as above until, by repetition, lead very low in silver is produced. For economic reasons this process is no longer in extensive use.

Parkes process. This process is based on the greater affinity of silver for zinc than for lead. Slab zinc is added to an argentiferous lead bath, the temperature of which has been raised higher than the melting point of the zinc. When the zinc has been melted and thoroughly mixed into the lead bath, the bath temperature is lowered and a silver-zinc alloy separates and floats on the top of the kettle. This zinc crust, as it is now called, is pressed off to remove the excess lead. The crust is broken up, cooled, and later retorted to recover the major part of the zinc, which is then reused. The resultant high silver-lead retort bullion is then cupelled to recover the silver and gold as doré and the lead as litharge. *See* LEAD METALLURGY.

Cupellation. This is the oldest and most widely known method of separating and recovering gold and silver from lead. It is used extensively to recover doré metal from the silver-lead end product of the Parkes process. The silver-lead bullion resulting from retorting of the zinc crust of the Parkes process is charged to a reverberatory-type cupellation furnace. After the charge has melted, air is blown across the top of the molten bath, causing oxidation of the lead and other impurities which separate as an impure litharge slag, leaving behind the silver and gold as a doré alloy. The doré is cast into anodes and sent to the refinery, where it is electrolytically separated into pure gold and silver.

Slimes smelting. By far the most important method of silver production results from the smelting and subsequent treatment of the silver-rich slimes resulting from the electrolytic refining of copper. Before smelting, the raw slimes are subjected to treatment with a sulfuric acid and air leach to remove excess copper and to combine the silver with the selenium and tellurium that are present. The resultant leached slimes are filtered through plate and frame presses to remove the excess solution and charged to a reverberatory-type furnace with appropriate fluxes. The melt slag resulting from this initial melting contains most of the base-metal impurities. Upon completion of this step, the material remaining in the furnace consists of gold, silver, selenium, tellurium, and residual base metals, and is called matte. At this point, treatment of the matte begins first with sodium carbonate to purge residual sulfur, if any, and then with niter (potassium nitrate) and air, forming a water-soluble slag containing the selenium and tellurium. In some cases caustic soda is used as a collector in conjunction with the niter and air treatment. Upon completion of the selenium and tellurium removal, the white metal or impure bullion is subjected to continued niter treatments, sometimes with the addition of lead, to remove the residual base metals. The resulting doré is then cast into anodes for subsequent electrolytic refining.

Refining. Refined metal from the doré furnaces (or from any previously mentioned production method) is treated electrolytically in either the Möbius or Balbach-Thum cell or variants thereof. The Möbius cell is more conventional, since because of the vertical suspension of the anodes and the cathodes, it is similar to the arrangement of a copper refinery. In the Balbach-Thum cell the cathode is placed on the cell bottom and the anodes are suspended in a basket parallel to the cathode. The major difference in the two methods is the fact that the Möbius cell requires less space than the Balbach-Thum cell in relation to the quantity of silver produced but produces 15–25% scrap which must be reprocessed, whereas the Balbach-Thum cell produces no scrap. The silver crystals produced by both methods are collected, washed, and cast into fine silver bars which are then ready for industrial or artistic use.

Recovery from secondary sources. Of importance to the world silver supply is the recovery of silver from scrap generated by the arts and industry. The more important sources of secondary silver are sterling and silver plate production; electrical components such as batteries, relays, and contact points; solders and brazing and other alloys; chemicals; mirrors, lamps, and jewelry; catalysts; and photographic and related processes. Because of the complexity of reclamation and the value of the above items, the sources of secondary silver may enter the smelting and refining process at the initial blast or reverberatory furnace smelting stage, the copper or lead refining stage, or, if high enough in silver purity, the doré production stage. In all cases, however, the final product is refined silver. *See* SILVER; SILVER ALLOYS.

R. D. Mushlitz

Bibliography. A. Butts (ed.), *Silver: Economics, Metallurgy, and Use*, 1967, reprint 1975; L. Coudurier, D. W. Hopkins, and I. Wilkomirsky, *Fundamentals of Metallurgical Processes*, 1985; M. A. Meyers and K. K. Chanla, *Mechanical Metallurgy*: *Principles and Applications*, 1984; H. W. Pollack, *Materials Science and Metallurgy*, 4th ed., 1988.

Silviculture

The theory and practice of controlling the establishment, composition, and growth of stands of trees for any of the goods (including timber, pulp, energy, fruits, and fodder) and benefits (water, wildlife habitat, microclimate amelioration, and carbon

sequestration) that they may be called upon to produce. In practicing silviculture, the forester draws upon knowledge of all natural factors that affect trees growing upon a particular site, and guides the development of the vegetation, which is either essentially natural or only slightly domesticated, to best meet the demands of society in general and ownership in particular. Based on the principles of forest ecology and ecosystem management, silviculture is more the imitation of natural processes of forest growth and development than a substitution for them.

Origins. Silviculture is the oldest deliberate application of ecology, recognized as a practice long before the term "ecology" was coined. Some of the most complex silvicultural techniques have been practiced for centuries in the tropics within tree garden systems, but on very small scales. However, in most parts of the world the dominant historical influence has been the clearing of forests for farmland either permanently or through more traditional cultivation of swidden systems (slash-and-burn agriculture). In Europe and the Middle East, much of this clearance occurred in pre-Roman times. Up until the late Middle Ages, the main use of woodland was in the provision of small and medium-scale materials for fuelwood, building poles, fencing, and wattle. These products were supplied mostly by a variety of short rotation coppice systems (thickset of small trees). Demand for timber was satisfied by standard trees left to continue to grow among the coppice or to form isolated trees in wood pastures where animals grazed the undergrowth or grass beneath. In the early 1600s, wood scarcity and the decline in land productivity became an important resource issue across much of Europe. Modern silviculture was therefore born out of necessity to develop restoration techniques for sustaining the productivity of woodlands and forests; even in the twenty-first century, exploitation of forests and depletion of resources usually precede silvicultural restoration. In the eighteenth and nineteenth centuries, silviculture began to break away from its empirical basis to find a foundation in the underlying sciences that were developing.

Treatments. The techniques proceed on the assumption that the natural vegetation of any site normally tends to extend itself to occupy all available growing space, making the fullest possible use of the available growth factors, such as light and moisture. The forester attempts, usually by cutting while harvesting useful wood products, to create vacancies in the forest vegetation that will provide environments favorable either to the establishment of new, desirable trees (reproduction treatments) or to the enhanced growth of remaining trees (intermediate operations).

Intermediate operations. These are made during the life of a particular crop of trees to alter species composition, to provide income, to increase the amount or value of the goods and services produced, and to alter forest structure in order to improve wildlife habitat. Under intensive management, intermediate operations may include various kinds of release cuttings to free desirable trees not past the sapling stage

Fig. 1. Release treatment for timber tree saplings in a mixed dipterocarp rainforest in Malaysia. (*Yale University School of Forestry and Environmental Studies*)

from the competition of taller or faster-growing trees that either are undesirable species or are defective specimens of otherwise acceptable species (**Fig. 1**). If the trees to be removed in such cuttings are not to be harvested, they may be killed either by girdling or by applying herbicides. *See* HERBICIDE.

If trees large enough to have rough bark are to be killed in place, the girdling is usually done by cutting away or injecting herbicides into a band of bark and cambium. This halts the downward movement of carbohydrates from the leaves to the roots; the starvation of roots ultimately kills the whole tree. Smaller trees or clumps of sprouts with thin bark also can be killed by basal spraying in which oil solutions of herbicides are applied entirely around the bases of the stems to girdle them and also to kill any dormant buds beneath the bark. It is possible to kill whole stands of undesirable woody plants by spraying foliage from the ground or air; this technique depends on essentially normal movement of materials through the phloem to carry the herbicides from the leaves to the roots at times when the roots will be killed. Under proper conditions, foliage spraying can be used to kill broad-leaved trees without significant harm to narrow-leaved conifers beneath them. *See* PLANT ANATOMY; PLANT TRANSPORT OF SOLUTES.

Thinnings are made in immature stands to stimulate the growth of the trees that remain and to increase the total production (**Fig. 2**). Thinning has its basis in that new stands start with many hundreds or thousands of trees per acre and ultimately decline to several hundreds or merely dozens as a result of severe competition. The crowns and roots

of individual trees expand, and the individuals that grow fastest in height overtop and ultimately kill the laggards.

Artificial thinning represents a judicious acceleration and guidance of this process. As long as the trees of a stand are sufficiently numerous to be capable of expanding to occupy the available growing space completely, the gross volume of raw wood produced by a stand of given composition and age on a particular site remains nearly independent of the number of trees per acre. Because a given volume of wood has greater value and utility if produced in a small number of large trees rather than a large number of small trees, it is generally desirable to reduce the stocking periodically to the smallest number of good trees that will fully occupy the site. The process can enhance the growth of the remaining trees in diameter but not in height. A little production is lost because some space is temporarily unoccupied by foliage.

Thinning also increases the actual yield of a stand through the harvest of surplus marketable trees that might otherwise be killed by competition and lost to decay. The practice often imitates the natural process by eliminating the smaller trees in favor of the larger. However, if the smaller trees happen to be more desirable than the larger trees, and if they are of sufficient vigor, it is possible to alter the pattern of removals to rescue them. Thinning enables the forester to guide an existing stand along lines of optimum development and to secure current income from an otherwise immature crop.

Thinning can be used to reallocate growing space for the purpose of altering wildlife habitat or the visual esthetics in a stand. For example, removing some of the larger trees will increase the vertical structural heterogeneity of the forest by releasing smaller trees that will grow at different rates among each other. Removing all the smaller trees in a stand can increase the depth and viewshed for a person walking along a trail.

Pruning is the removal of branches from the stems of living trees, ordinarily for the purpose of increasing the volume of clear (knot-free) wood. The practice is best confined to high-quality trees destined to be favored by thinning as part of the final crop.

Reproduction treatments. These are made when the prospective growth of an old stand is deemed less valuable or desirable than that of a new stand. Reliance is often on natural seeding or upon the release of "advance regeneration" already present beneath the old stand. Some species, mostly angiosperms, can renew themselves simply by sending up new sprouts from stumps or roots after their tops are killed by cutting or fire. Under intensive practice, it is common to depend on artificial (or direct) seeding or, more commonly, the planting of nursery-grown seedlings. These forms of artificial regeneration are used when natural regeneration is regarded as too slow, where there is no source of natural seed, when selected superior strains are to be introduced, or simply when there is reason to use logging methods that are incompatible with natural regeneration. *See* HERBICIDE.

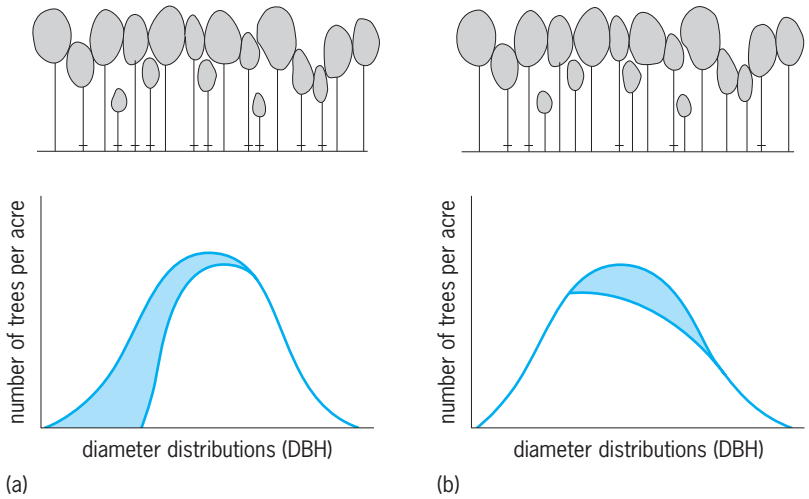


Fig. 2. Types of thinning. (*a*) Previously untreated even-aged stand marked (by horizontal lines at the base of the stems) for a low thinning whereby trees removed are the smallest trees in the stand. (*b*) Crown thinning whereby trees removed are the direct competitors of the remaining dominant canopy. In the diameter distributions (DBH), the shading depicts that part of the original stand that is removed in the thinning.

Successful regeneration from seed, whether naturally or artificially applied, demands basic knowledge of the ecological requirements for germination and survival of the desired species. An abundant source of seed is essential. The condition of the seed bed and the arrangement of the remaining trees also must be adjusted to provide an adequate number of spots where the environment is more favorable to establishment of the desired species than to any others. The native flora of any forest region includes species adapted to colonize almost any kind of vacancy that might be created in the forest by natural disturbances, great or small. The forester attempts to simulate the kind of disturbance that will lead to establishment of the particular species desired. *See* DENDROLOGY.

Stand patterns and processes. The spatial patterns in which old trees are removed and the species that replace them determine the structure and developmental processes of the new stands. If all the trees are replaced at once with a single species, the result is so-called pure stand or monoculture in which all of the trees form a single canopy of foliage that is lifted ever higher as the stand develops. If several species start together from seeds, from small trees already present, or from sprouts, as a single cohort, the different species usually tend to grow at different rates in height. Some are adapted to develop in the shade of their sun-loving neighbors. The result is a stratified mixture (**Fig. 3**). Such stands grow best on soils or in climates, such as in tropical moist forests, where water is not a limiting factor and the vegetation collectively uses most of the photosynthetically active light. If trees are replaced in patches or strips, the result is an uneven-aged stand which may be of one species or many.

These different spatial and temporal patterns of stand structure are created by different methods of reproduction. The simplest is the clear-cutting

Fig. 3. Stand patterns. (*a*) Stratified mixture of red oak and eastern hemlock in a temperate moist forest in northeastern Connecticut. (*b*) Stratified mixture of mahogany with several subcanopy species of palm and an understory shrub layer in a moist subtropical rainforest in Puerto Rico. (*Yale University School of Forestry and Environmental Studies*)

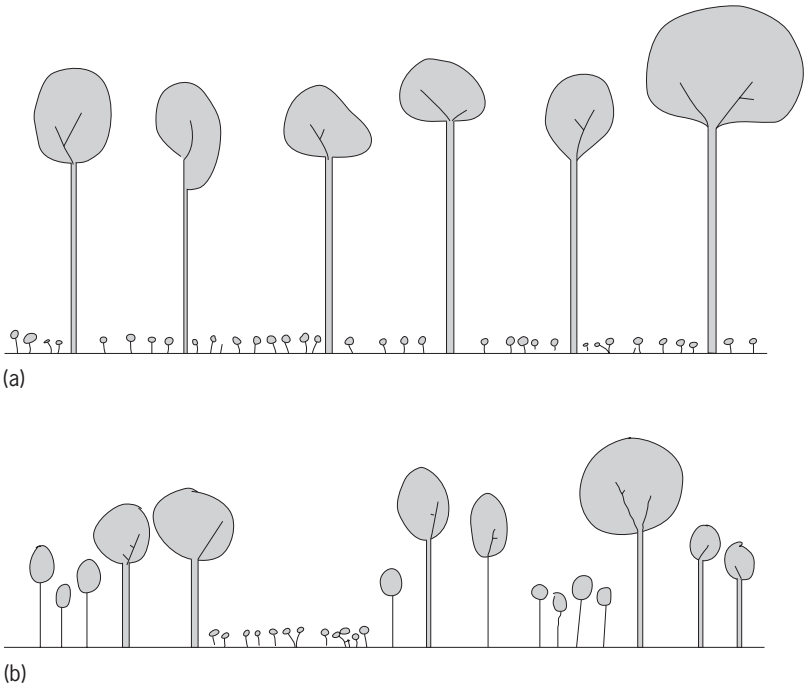


Fig. 4. Reproduction treatments. (*a*) Even-aged stand after uniform shelterwood cuttings to establish advance regeneration. The uniform shelterwood method involves the removal of the entire stand in a series of partial cuttings that extend over a fraction of the rotation. (*b*) Uneven-aged stand after single tree selection treatments. The selection method involves the removal of scattered individuals or strips or groups of trees at relatively short intervals; such cuttings are repeated indefinitely.

method, in which virtually all of the vegetation is removed. Although it is sometimes possible to rely on adjacent uncut stands as sources of seed, it is usually necessary to reestablish the new stand by artificial seeding or planting after clear-cutting. The seed tree method differs only in that a limited number of trees are temporarily left on the area to provide seed.

In the shelterwood method, enough trees are left on the cutting area to reduce the degree of exposure significantly and to provide a substantial source of seed (**Fig. 4***a* and **Fig. 5***a, b*). In this method the growth of a major portion of the preexisting crop continues, and the old trees are not entirely removed until the new stand is well established. The three methods just described lead to the creation of essentially even-aged stands.

The various modifications of the selection method, however, lead to the creation of uneven-aged stands characterized by the presence of three or more distinct age classes (Fig. 4*b* and **Fig. 6**). Reproduction treatments under the selection method occur in the same stand at more or less equal intervals of time and are not concentrated near the end of the life of the crop as in the even-aged stands. Both the shelterwood and selection methods are subject to very large amounts of variation and modification; both are applied in one form or another to the management of almost all species and forest types, except the few that actually need exposure to become established. In reality, selection and shelterwood methods have a high degree of overlap. This overlap can be

Fig. 5. Shelterwood methods. (*a*) Uniform shelterwood cutting in an 100-year-old stand consisting of oak and other hardwood trees in Connecticut. (*b*) Group shelterwood where regeneration has been established in distinct canopy openings throughout a beech stand in the Czech Republic. (*Yale University School of Forestry and Environmental Studies*)

perceived as dependent upon the size of the openings created within a single entry for a selection treatment (for example, the amount of trees removed) as compared to the arrangement, age-class distribution, and amount of reserve trees actually left standing until the next rotation for a shelterwood.

The so-called coppice methods involve establishment of new stands from sprouts arising from the stumps or roots of cut trees. They are used mainly for the production of quick crops of such species as the poplars for fuel and pulp. However, some species, such as oaks and coast redwoods, can grow to very large sizes from sprouts.

Clear-cutting is not necessarily crude silviculture; it may even represent the final harvesting operation following a series of frequent intermediate cuttings. Conversely, light periodic partial cuttings do not invariably represent highly efficient practice, but may involve "high-grading," in which the few good trees of an otherwise untended stand are cut to leave it dominated by undesirable trees.

The application of partial cutting often involves the concept of financial maturity as a guide in determining which trees to cut and which to reserve for additional growth. According to this concept, the market value of each standing tree is regarded as a capital investment, and an attempt is made to predict the increase in value of the tree between the cutting immediately in prospect and the next cutting. If this increase in value does not represent an acceptable rate of compound interest on the investment represented by the present value of the tree, the tree is marked for immediate cutting. Under this concept, fast-growing trees of good quality are reserved, while poor unthrifty trees are likely to be cut.

Ecological factors. The choice of methods of regeneration cutting depends on the ecological status of the species and stands desired. Species that characterize the early stages of succession, the so-called pioneers, will endure, and usually require the kind of exposure to sunlight resulting from heavy cutting or, in nature, severe fire, catastrophic windstorms, floods, and landslides. These pioneer species usually grow rapidly in youth but are short-lived and seldom attain large size. They are exemplified by slash, lodgepole, and jack pines as well as by aspens, paper birch, and red alder. The species that attain greatest age and largest size are ordinarily those which are intermediate in successional position and tolerance of shade. They are moderately resistant to exposure when young; while they can usually endure the shade of light-foliaged pioneer trees, shrubs, or herbs, they seldom persist beneath a closed canopy of their own species. Some of them will become established after the severe exposure of clear-cutting, but they often start best with light initial shade such as that created by shelterwood cutting. This group is exemplified by Douglas-fir, loblolly pine, the five-needled pines (white pines), sweet gum, tulip poplar, and the ashes. Their longevity and large size result from the fact that they are naturally adapted to reproduce after disturbances occurring at relatively long intervals. *See* ECOLOGICAL SUCCESSION; FOREST ECOSYSTEM.

The shade-tolerant species representing late or climax stages in the succession are adapted to reestablish themselves in their own shade. They usually require such protection when young; they can survive as stunted advance regeneration for many years and suddenly start rapid growth when released. These species represent natural adaptations to the kinds of fatal disturbance caused by insects, disease, and atmospheric agencies rather than to the more complete disturbance caused by fire. Since they can reestablish themselves almost continuously, longevity is not necessary for survival of the species. If these species are already established as advanced regeneration, new stands can be created simply by removing the overstory in any pattern that is otherwise desirable, provided that the small trees are

Fig. 6. Single-tree selection method of reproduction in a Norway spruce–silver fir stand in Bavaria, Germany. (*Yale University School of Forestry and Environmental Studies*)

not destroyed in the process. If advanced growth is not established when the time comes to replace the stand, comparatively light partial cutting (shelterwood or selection) is necessary to accelerate the regeneration process. Among the species of this group are the true firs, spruces, hemlocks, maples, beeches, and oaks. Some of the limitations that ecological factors place on the germination and early survival of seedlings can be evaded by planting ones that have survived the crucial early stages in the protected environment of a nursery.

Mixed and pure stands. Much silvicultural practice is aimed at the creation and maintenance of pure, even-aged stands of single species. This approach is analogous to that of agriculture and simplifies administration, harvesting, and other operations. The analogy is often carried to the extent of clear-cutting and planting of large tracts with intensive site preparation, especially with species representative of the early or intermediate stages of succession.

Mixed stands are more difficult to handle from the operational standpoint but are more resistant to injury from insects, fungi, and other damaging agencies which usually tend to attack single species. They are also more attractive than pure stands and make more complete use of the light, water, and nutrients available on the site. They usually do not develop unless these site factors are comparatively favorable; where soil moisture or some other factor is limiting, it may be possible for only a single well-adapted species to grow. If the site is highly favorable, as in tropical rainforest, river floodplains, or moist ravines, it is so difficult to maintain pure stands that mixed stands are inevitable.

In mixed stands, the different species have a tendency to occupy different strata in the crown canopy, depending on their rate of height growth or position in succession. If all start in the same year or within a few years of one another, any pioneer species that are present quickly occupy the upper stratum even though they may succumb after a few decades. Thereafter the ascendancy passes to species that are intermediate in succession, and in nature they may dominate for the remainder of the next century. When they begin to dwindle in number from natural attrition, the long-suppressed species of the late successional stages begin to take over more of the upper canopy levels. In the silvicultural treatment of such stands, it is often appropriate to anticipate, imitate, accelerate, and guide the sequence of events. The modifications usually take the form of cuttings in which individual trees or strata are removed when they are mature. It is possible to proceed in such ways that either the even- or uneven-aged structure is created or maintained; in other words, mixed stands can be either even- or uneven-aged.

The various strata representing different sets of species and successional stages can also appear or be caused to develop at successive intervals separated by many decades. If otherwise desirable, the harvesting of the successive overstories can proceed in a manner that even more nearly duplicates the events of natural succession than where all species become established nearly simultaneously.

Accessory practices. The application of silviculture involves a number of accessory practices other than cutting. Fire, like the ax, is an agent that can be used wisely to build the forest or unwisely in its witless destruction. In localities of high fire risks, it may be desirable to burn the slash (logging debris) after cutting. Not only does this reduce the potential fuel, but it may also help the establishment of seedlings by baring the mineral soil or reducing the physical barrier represented by the slash. Slash disposal is most often necessary where the cutting has been very heavy or where the climate is so cold or dry that decay is slow.

Deliberate prescribed burning of the litter beneath existing stands of fire-resistant species is sometimes carried out even in the absence of cutting to reduce the fuel for wild fires, to kill undesirable understory species (including those representing stages in plant succession later than those desired), to enhance the production of forage for wild and domestic animals, and to improve seedbed conditions. Prudence demands that such burning be prescribed only by experts under carefully chosen weather conditions and with ample provision for controlling the fires. The practice has been used mainly in stands of fire-resistant two- and three-needled pines, primarily in the southern United States, where it closely imitates the effect of natural fires. *See* FOREST FIRE.

Heavy disk plows, rolling brush cutters, modified bulldozers, and similar equipment also find use in the elimination of undesirable vegetation and preparation of the soil for the establishment of new crops by planting or natural seeding. The growth of a stand is sometimes improved by chemical fertilization or by drainage of swamps and bogs.

Silvicultural practices may be conducted not only for the growth of wood products but also for development of forest stands of form and composition consistent with wise management of water, wildlife, scenery, and other forest resources. Almost all forests provide sources of water. Earthworms and other soil organisms that feed on fallen leaf litter generally maintain soils of high porosity. Water that sinks into porous soils does not erode or move sediments into streams but can emerge from clear-water springs after it has been filtered through the soil. Almost any type of forest vegetation will maintain suitable soil porosity. The cutting of trees does not cause erosion, floods, or reduced rainfall, but may actually reduce transpiration losses of water enough to cause small increases of stream runoff, usually during dry periods. However, turbid water can flow off roads and trails, so it is important that water is diverted from them as soon as possible and, preferably, into areas where the sediment will filter out before they reach streams. In regions where winter snows are melted mainly by the sun rather than rain, water can be stored in snow drifts; these can be induced to form by cutting narrow openings in the forests.

Silviculture that is aimed at wildlife management (such as that for timber management) works by controlling the structure and species composition of stands. However, it also involves the creation of more kinds of stands. Since animals are mobile, it is much more difficult to learn their life histories than is the case with trees, and to determine the kinds of stands to create. The objective is to build the kinds of habitats that will call forth the desired animal population.

If maximum diversity of plant and animal species is desired, forests are required to have a wide range of age classes, stand structures, and species. Some wildlife species require conditions (such as hollow trees suitable for nests) that go with stands or trees older than those that would be grown just for timber production. Others, including many game animals, may require very young stands or even the elimination of trees in favor of shrubs, grasses, or herbaceous plants. If grazing of cattle is to be done beneath the trees, the establishment of grasses is promoted by heavy thinning, elimination of understory vegetation, or prescribed burning.

Managing cutting areas for esthetic purposes may include a variety of measures. All cutting areas look bad, so it is best to screen them from public view by leaving trees and removing debris along routes of travel. Log loading areas should also be kept out of sight. Where cutting areas will be visible for long distances, it is best to avoid straight boundaries and to effect irregular shapes that simulate natural openings. *See* CONSERVATION OF RESOURCES.

Ecosystem management. Integrated schedules of treatment for stands are called silvicultural systems. They cover both intermediate and reproduction treatments but are classified and named in terms of the general method of reproduction cutting contemplated. Such programs are evolved for particular situations and kinds of stands with due regard for all the significant biological and economic considerations involved. These considerations include the desired uses of the land, kinds of products and services sought, prospective costs and returns of the enterprise presented by management of the stand, funds available for long-term investment in stand treatments, harvesting techniques and equipment employed, reduction of losses from damaging agencies, and the natural requirements that must be met in reproducing the stand and fostering its growth.

It is also necessary to coordinate silvicultural systems for the treatment of individual stands with plans for aggregation of stands that are called forests in an administrative sense. It is generally desirable to plan the length of rotations and the arrangement of different age classes of trees or stands (**Fig. 7**) so that the forest as a whole will be the source of continuous, sustained yield of products as well as other benefits that forests can provide indefinitely (water, recreation). For a discussion of the planning, organization, and the administration of whole forests for these objectives *see* FOREST MANAGEMENT.

Identifying the natural pattern within a landscape is the best starting point for defining appropriate scales of stands for forest management. Physiographical attributes of soil nutrition, texture, and soil moisture availability to plants, along with coarser descriptors such as aspect, slope, and topographic position, when integrated together, are the best determinants

Fig. 7. Two age classes of ponderosa pine developed from natural seeding. The stand in the background is about 60 years old and that in the foreground is 35 years old. (*Yale University School of Forestry and Environmental Studies*)

of site productivity. The ability to identify how site productivity changes across different temporal and spatial scales provides the base guideline for defining thresholds of management practice from one site to another. Conceptually this seems logical, but in reality threshold limits in site productivity usually change across landscapes as continua rather than within any clearly identifiable abrupt transition. Such an example might be a ridge subject to drought, with a thin to bedrock soil that obviously will have a lower site productivity with a narrower set of management options and ecological constraints, as compared to a fertile, moist, and deep soil in the adjacent valley. Economic values demand that the two sites be managed differently, which means that somewhere between the two sites the management on the more productive site will be delineated from that on the less productive. This will likely accentuate abruptness between ridge and valley ecosystems.

Therefore, though ecological processes are without boundaries, ecosystem management is only feasible when silvicultural operations can be considered at certain discrete scales. The three most important scales for accommodating ecosystems processes include forest stands, forest watersheds, and physiographic regions. The different scales are appropriate for different kinds of forest management and should be integrated as a whole. Stand scales are best for managing forests where direct manipulation for the purposes of yielding a product is necessary (including timber, fruits, and fuelwood). Other service-related values must be considered at watershed (landscape) scale (such as water management, recreation, and wildlife habitat). The whole physiographic region is a useful scale for conceiving management operations for sequestering carbon or preserving biodiversity. David M. Smith; Mark S. Ashton

Bibliography. J. W. Barrett (ed.), *Regional Silviculture of the United States*, 3d ed., 1994; T. W. Daniel, J. A. Helms, and F. S. Baker, *Principles of Silviculture*, 2d ed., 1979; R. D. Nyland, *Silviculture: Concepts and Applications*, 1996; C. D. Oliver and B. C. Larson, *Stand Dynamics*, 1996; D. M. Smith et al., *Practice of Silviculture: Applied Forest Ecology*, 9th ed., 1997; U.S. Forest Service, *Silvics of North America*, 1990; U.S. Forest Service, Silvicultural Systems for the Major Forest Types of the United States, *Agr. Handb.*, no. 445, 1983.

Simple machine

Any of several elementary machines, one or more of which is found in practically every machine. The group of simple machines usually includes only the lever, wheel and axle, pulley (or block and tackle), inclined plane, wedge, and screw. However, the gear drive and hydraulic press may also be considered as simple machines. The principles of operation and typical applications of simple machines depend on several closely related concepts. *See* BLOCK AND TACKLE; HYDRAULIC PRESS; INCLINED PLANE; LEVER; SCREW; WEDGE; WHEEL AND AXLE.

Two conditions for static equilibrium are used in analyzing the action of a simple machine. The first condition is that the sum of forces in any direction through their common point of action is zero. The second condition is that the summation of torques about a common axis of rotation is zero. Corresponding to these two conditions are two ways of measuring work. In machines with translation, work is the product of force and distance. In machines with rotation, work is the product of torque and angle of rotation.

Work is the product of a force and the distance through which it moves. For example, the work done in raising a 10-lb object 15 ft is 150 ft-lb. In this example the work done on the weight goes into increasing the potential energy of the object. Work and energy, both potential and kinetic, have the same units, and in general the purpose of a machine is to convert energy into work. *See* EFFICIENCY; MECHANICAL ADVANTAGE; WORK.

For rotating machines, it is more convenient to consider torque and angular displacements than force and distance. Work is then expressed as the product of the torque and the angle (in radians) through which the object rotates while acted on by the torque. Torque, in turn, is the force exerted at a given radius from an axis of rotation. Thus, a 10-lb force at the end of a 15-ft crank exerts a torque of 150 lb-ft. *See* TORQUE.

Power is the rate of doing work. For example, one horsepower is arbitrarily defined as 550 ft−lb per second, or 33,000 ft-lb per minute. *See* MACHINE; POWER.

Richard M. Phelan

Simulation

A broad collection of methods used to study and analyze the behavior and performance of actual or theoretical systems. Simulation studies are performed, not on the real-world system, but on a (usually computer-based) model of the system created for the purpose of studying certain system dynamics and characteristics. Simulation is both an experimental science and an art or craft in the development of the precise, "valid" model that captures the necessary elements of the system, and the use of the model to analyze system behavior under various scenarios.

Simulation is used in many disciplines to model different types of systems. For example, simulation can be used in task or situational training areas in order to allow humans to anticipate certain situations and be able to react properly; decision-making environments to test and select alternatives based on some criteria; scientific research contexts to analyze and interpret data; and understanding and behavior prediction of natural systems, such as in studies of stellar evolution or atmospheric conditions.

The focus of this article is the use of simulation in decision making. Simulation helps provide information to aid the decision maker in selecting the proper alternatives. Simulation then becomes a decision-aiding tool, and with simulation a decision maker can try out new designs, layouts, software programs, and systems before committing resources to their acquisition or implementation; test why certain phenomena occur in the operations of the system under consideration; compress and expand time; gain insight about which variables are most important to performance and how these variables interact; identify bottlenecks in material, information, and product flow; better understand how the system really operates (as opposed to how everyone thinks it operates); and compare alternatives and reduce the risks of decisions.

Advantages. Simulation is a powerful and flexible tool because a simulation study can be tailored to provide answers to many different questions. The goal of each simulation study varies; however, there are some fundamental reasons why people use simulation to analyze real-world systems:

1. Simulation can be used to predict outcomes. If a bank wants to reconfigure its operations, simulation can be used to test various layouts and scenarios of operation.

2. Most real-world systems are stochastic (changes in the system are determined by chance) and dynamic (the system changes over time). In this case, the development of an analytical solution for studying the systems is either not feasible or impractical. *See* STOCHASTIC PROCESS.

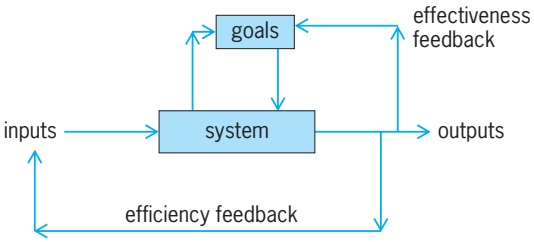
3. Simulation can account for the effects of variance. Multiple runs of the model can be made under different conditions, and not only the average outcomes but also the variability of outcomes can be examined.

4. Simulation analysis promotes what is known as total solutions. A simulation model is designed and built to take into account all aspects that directly impact the system; indirect impacts exist in the system environment. Since the total system is reflected in the model, any outcomes suggested by the analysis of the model represent total solutions to the system behavior.

5. Simulation can bring together expertise, knowledge, and information about a system and its operations. The process of model development should incorporate any outside human expertise and knowledge about the system. In addition, information on the current status of inputs and processes is incorporated directly into the model and its analysis. Finally, simulation can be a very cost effective tool, especially in the design of new systems. Many different scenarios can be examined and analyzed. Thus, even new and creative ideas for system modifications can be examined without harm to the current operations.

Simulation is essentially a four-step process to formulate (build) a valid model of a real world system, develop operational scenarios of interest (from the real world) that exercise the model, analyze the output of these scenarios, and translate and project the model outputs back to the real-world system.

Systems. The word "system" refers to a set of elements (objects) so interconnected (interdependent, interrelated) as to aid in driving toward a desired goal. This definition has two connotations: First, a



System feedback.

system is made of parts (elements) that have relationships between them (or processes that link them together). These relationships or processes can range from relatively simple to extremely complex. One of the necessary requirements for creating a "valid" model of a system is to capture, in as much detail as possible, the nature of these interrelationships. Second, a system constantly seeks to be improved. Feedback (output) from the system (*see* **illus.**) must be used to measure the performance of the system against its desired goal. Both of these elements are important in simulation. *See* SYSTEMS ENGINEERING.

System environment and system state. A system is often affected by changes that occur outside the system. These changes are said to occur in the system environment. In the modeling of systems, it is important to decide on the boundary between the system and its environment. Changes in the system can be reflected (or modeled) by changes in the system state. The state of the system is defined to be the collection of variables necessary to describe the system at any time.

For example, in the study of a bank operation, the elements in the system include the tellers, the waiting area (how many people are allowed to wait at one time), and the services (processes) provided by the tellers to the customers. The environment of this system includes the town in which the bank is located. If it is a payday for a large company in town, the bank may see more customers on that particular day. The state of the bank system could be defined as the number of tellers who are working, the number of tellers who are currently serving customers, and the numbers of customers waiting. Each time one of these variables changes, the state of the bank system changes.

Classification of systems. Systems can be classified in three major ways. The classification of systems (*see* **table**) revolves around the types of elements that exist in the system, the nature of time and how the system state changes in relation to time, and whether or not the system changes over time at all. This categorization affects the type of modeling that is done and the types of simulation tools that are used.

Classification of systems and models

Classifiers	Systems
Elements	Deterministic versus stochastic
Time	Discrete-event versus continuous
Change	Static versus dynamic

A system that has elements or interrelationships betwen the elements known with complete certainty is called a deterministic system. In deterministic systems, every time a particular input is placed into the system a known output always results. Deterministic systems are easier to model than stochastic systems. Stochastic systems have at least one element or interrelationship that is determined by chance and therefore must be modeled with probabilities. The bank is an example of a system containing stochastic elements because the arrival of customers and the length of time to service a customer are both stochastic.

Time is an important element is many real-world systems. The nature of time in a system and how time affects the system state is an important classifier. A system that changes in a continuous manner is a continuous-time system. Differential calculus, specifically differential equations, are often used to describe the nature of the system state through a model. In the bank example, events, such as the arrival of a customer or the completion of service of a customer, happen at discrete instants in time; these systems are known as discrete-event systems.

The nature of change in the system also acts as a classifier. Systems that do not change over time are called static, whereas systems that do change over time are called dynamic. Most real-world systems are both dynamic and stochastic, and simulations models are usually developed for such systems. Different tools exist to model systems that are continuous-time or discrete-event.

Modeling. The purpose of any model is to enable its users to draw conclusions about the real system by studying and analyzing the model. The major reasons for developing a model, as opposed to analyzing the real system, include economics, unavailability of a "real" system, and the goal of achieving a deeper understanding of the relationships between the elements of the system.

The model should include only the elements of the system that are within the boundary of the system relative to the objectives of the study. While the system environment does have an effect on the model outcome, the environment does not have to be modeled explicitly.

Models of actual or theoretical physical systems can take many forms. Models, like the systems they represent, can be static or dynamic, discrete or continuous, and deterministic or stochastic. Simulation models are composed of mathematical and logical relations that are analyzed by numerical methods rather than analytical methods. Numerical methods employ computational procedures to run the model and generate an artificial history of the system. Observations from the model runs are collected, analyzed, and used to estimate the true system performance measures. *See* MODEL THEORY.

Simulation studies. There is no single prescribed methodology in which simulation studies are conducted. However, most simulation studies cover each

of the steps below to some level of detail depending on the study objectives. Most simulation studies proceed around four major areas: the problem, the system and model, the experiment, and the analysis.

Problem formulation. In the problem area, the major activity is in the formulation of the problem. In formulating the problem, a key step in any problem-solving endeavor, the analyst must strive to understand and define the problem in a precise and systematic form. The analyst must be able to state a clear understanding for the entities in the system and the interrelationships between the various entities. In this stage a clear and measurable set of objectives of the simulation analysis must be stated, and with those objectives a well-defined performance metric (or index) must be determined. In the problem definition, all assumptions as to the performance of the model must be clearly and explicitly stated. The scenarios that drive the simulation must be developed so that key inputs to the system (and therefore the model) are also stated. Finally, any restrictions on the system solutions (physical, political, economic, and social) must be clearly identified.

The output of this step is a detailed system description. In this description a baseline system is developed to show the movement of entities through the system; this can be done using a flow diagram or schematic depiction of the baseline system. The diagram should include entry points and exits to and from the system, any processing stations, and decision points where an entity might take a different path throughout the remainder of the system.

Once a baseline system is detailed, the analyst can start to design a new system diagram that would show possible changes to the baseline system that will be studied to improve overall system performance. The new system diagram can entail small changes to the baseline or significant deviations from the processing of entities from the baseline. This new system diagram is used to generate a list of potential questions that are to be answered with the simulation efforts. These questions must be answerable using the existing performance criteria (metrics that were defined in the earlier part of the problem formulation step).

The activities that are undertaken in this step help the simulation analyst to prepare for the next step in the problem analysis phase of the simulation study: defining and developing the model.

Model development. The development of a system model requires both science and art. A great detail of experience in developing models is required to obtain a valid and meaningful model of the system under study. The analyst must determine to what level of detail the system needs to be modeled and what are the essential features of the model, which are critical to the study objectives. In addition, the analyst must make an effort to involve the users of the model in the building process. If the users of the model are not satisfied that the model solves their problems, the modeling effort will be wasted.

The building of a system model is a top-down and often incremental process. The analyst will begin by developing a high-level model that captures the flow of entities through the system and leaves many of the finer details for future development. The analyst continues to refine the model in an effort to ensure that the model is a true representation of the system.

There is a constant interplay between the development of the model and the collection of needed input data. Before building the model (and as the model changes), the analyst must collect data from the real-world system and represent that data within the processes of the model. In a stochastic system, most of the real-world data that are collected are transformed and represented by probability distributions. Each of the processes within the model must attempt to mimic the real-world processes as closely as possible. The transformation of real-world data into a probability distribution is a well-documented statistical procedure, and the analyst must be able to use judgment and experiences to choose the right distribution for the situation and then follow the statistical procedure to ensure that the parameters of the distribution are accurate. *See* DISTRIBUTION (PROBABILITY); STATISTICS.

Model verification is used to check the correctness of the model functionality. In order to check the correctness, multiple runs of the simulation model are performed to ascertain that the correct outputs are generated for the particular inputs. Model validation is a much more difficult endeavor. It requires a check of the degree to which the model is representative of the real system; this usually requires an independent credibility check. Users and outside experts may be called on to review the structure of the model, the level of detail of the model, and the output of the model in comparison to the real-world system.

Another important activity in understanding the system and developing a valid model is the use of sensitivity analysis. Sensitivity analysis is a group of statistical techniques that help the analyst (and users) to understand the effect of input variables on the output. Sensitivity analysis is undertaken to examine the input variables and ascertain which variables have a high degree of "sensitivity." It is important in the understanding of the system to assess how small changes in the input of the system can have large effects on the output. Sensitivity analysis is extremely important in the design of the simulation experiments that are undertaken with the model. It can also lead to further refinement (more detail) in the data collection process or the study of a particular input parameter. *See* SENSITIVITY (ENGINEERING).

Running the model. Simulation is an experimental science, so care must be taken to design experiments with the simulation model to obtain results that address the objectives of the study. Simulation is not designed to find the "best" or optimal answer to the problem at hand, but the nature of simulation makes it possible to compare many different alternatives. Solutions found using simulation are only as good as the alternatives being investigated. In other words, simulation is not a discovery science that allows a

creative and innovative solution to be uncovered unless that particular solution is being tested directly.

In designing simulation experiments, it is largely the job of the analyst to determine the proper inputs that need to be changed in order to meet the study's objectives. By controlling the change in these inputs over multiple runs of the model, the analyst can minimize the effect of variability that is introduced to the output of the mode. This in turn makes the results of the study more meaningful.

Many systems that simulation deals with are both dynamic and stochastic. Since a simulation model can be viewed as an input-output transformation, and since many of the inputs are random variables, the outputs of a simulation model are also random. Since these outputs are random, a single run of the model is insufficient in making a judgment of the result of the model. Multiple runs (or replications) of a stochastic model must be made, and statistical inference techniques on these replications must be used to make credible recommendations for system changes.

Analyzing the model output. Output analysis is the analysis of data generated by a simulation. Output analysis is used to predict system performance or compare the performance of two or more system designs. In the problem-formulation stage of the study, the analyst developed a problem description that included the specific metrics that will be used to measure the performance of the system. These metrics will differ from study to study depending on the objectives of the study. However, regardless of the metric to be used, each run of the simulation will yield a single value for the metric and this value will differ from run to run. The job of the analyst is to determine the long-term behavior of the system; this is accomplished by taking multiple runs of the simulation model and averaging these runs.

When analyzing simulation output data, a distinction is made with regard to the type of system being modeled. A terminating (or transient) system is one that starts at time 0, under a prespecified set of initial conditions, and proceeds until some specified time, at which time the system "closes." The objectives and the metrics for a terminating system may differ from that of a nonterminating or steady-state simulation. In analyzing data from a steady-state simulation, the goal is to make a valid statistical inference about the long-term average behavior of the system based on the sample average from a specified number of replicate simulation runs. The tool most often used to make this determination is confidence intervals. A confidence interval is an interval estimate for a parameter that specifies a level of confidence. A confidence interval provides a way of quantifying imprecision. The goal of a confidence interval procedure is to form an interval, with end points determined by the sample, that will contain, or "cover," the target parameter with a prespecified (high) probability called the confidence level.

Statistical inference methods also allow the comparison of various competing system designs or alternatives. For example, estimation and hypothesis testing make it possible to discuss the outputs of the simulation and compare the system metrics.

Applications. Simulation is a versatile tool that can be applied to many different systems operating (or proposed) in society. Many of the applications of simulation are in the area of manufacturing and material handling systems. Manufacturing systems are the focus of many simulation textbooks, and simulation is taught in many engineering and business curricula with the focus of the applications also being on manufacturing systems. The characteristics of these systems, such as physical layout, labor and resource utilization, equipment usage, products, and supplies, are extremely amenable to the simulation modeling methods described above. *See* COMPUTER-INTEGRATED MANUFACTURING; FLEXIBLE MANUFACTURING SYSTEM.

There are many other areas where simulation is prominent. Transportation systems, service systems (such as hospitals and banks), communication and network systems, environmental systems, business process reengineering, and social and economic systems are all candidate fields for the use of simulation.

Simulation began in the late 1950s and 1960s as a very expensive and specialized tool that was generally used by large corporations, primarily steel and aerospace companies. With the advent of personal computers and the relative richness of current simulation software packages, simulation has become a mature tool that many smaller firms can use with little capital investment. *See* COMPUTER-AIDED ENGINEERING; MATHEMATICAL SOFTWARE; MICROCOMPUTER; SOFTWARE.

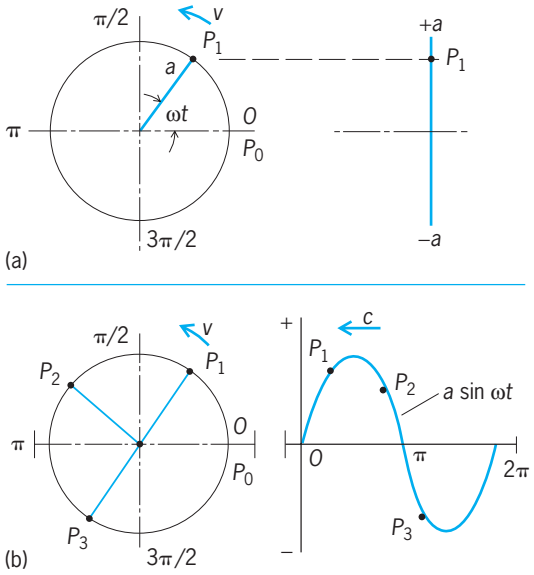
Current applications of simulation are concerned primarily with the design or redesign of complex systems. In the future, simulation will concentrate its efforts on the actual control of the system. This requires that models be kept current, but also allows for easy testing of the new systems over time. *See* COMPUTER-AIDED DESIGN AND MANUFACTURING; CONTROL SYSTEMS; ENGINEERING DESIGN.

James Pomykalski

Bibliography. J. Banks et al., *Discrete-Event System Simulation*, 3d ed., Prentice Hall, Upper Saddle River, NJ, 2000; T. J. Gogg and J. R. Mott, *Improve Quality and Productivity with Simulation*, 3d ed., Consulting Group, Palos Verdes, CA, JMI, 1996; B. Hannon and R. Matthias, *Dynamic Modeling*, 2d ed., Springer-Verlag, New York, 2000; A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3d ed., McGraw-Hill, New York, 2000; W. D. Kelton, R. P. Sadowski, and D. A. Sadowski, *Simulation with Arena*, McGraw-Hill, Boston, 1998; A. B. Pritsker, *Introduction to Simulation and SLAM II*, 4th ed., Halsted Press, New York, 1995.

Sine wave

A wave having a form which, if plotted, would be the same as that of a trigonometric sine or cosine function. It generally results from the solution of a problem having a one-dimensional space coordinate, such as the transverse vibrations of a string, longitudinal vibrations of a bar, or the propagation of plane waves of electromagnetic radiation or sound. *See*



Simple harmonic motion and sine wave. (*a*) Point *P* moves around the circle at constant tangential speed *v*; its projection moves up and down between limits $+a$ and $-a$. (*b*) Same conditions as in preceding, except that the plane on which the projection is plotted is traveling to the left at speed *c*, thereby generating a sine wave. Angular speed is ω rad/s; the total angle traced out from zero position equals ωt rad.

ELECTROMAGNETIC RADIATION; SOUND; VIBRATION; WAVE EQUATION; WAVE MOTION; WAVE MOTION IN FLUIDS.

The sine wave may be thought of as the projection on a plane of the path of a point moving around a circle at uniform speed. For example, in **illus.** *a*, assume that the moving point travels around the circle at constant speed v. The projection onto a plane would trace back and forth on the line indicated as the point went around the circle. If the plane is now moved to the left at constant speed c, as in illus. *b*, the resulting trace has the form of the graph of the sine function.

The sine wave trace of illus. *b* describes a body moving in simple harmonic motion. This motion changes in magnitude and time so that it repeats itself exactly, as long as uniform speed is maintained. It is characteristic of one-dimensional vibrations and one-dimensional waves having no dissipation. *See* HARMONIC MOTION.

The sine wave is the basic function employed in harmonic analysis. It can be shown that any complex motion in a one-dimensional system can be described as the superposition of sine waves having certain amplitude and phase relationships. The technique for determining these relationships is known as Fourier analysis. *See* FOURIER SERIES AND TRANSFORMS; WAVEFORM.

William J. Galloway

Single crystal

In crystalline solids the atoms or molecules are stacked in a regular manner, forming a three-dimensional pattern which may be obtained by a three-dimensional repetition of a certain pattern unit called a unit cell. When the periodicity of the pattern extends throughout a certain piece of material, one speaks of a single crystal. A single crystal is formed by the growth of a crystal nucleus without secondary nucleation or impingement on other crystals. *See* CRYSTAL STRUCTURE; CRYSTALLOGRAPHY.

Growth techniques. Among the most common methods of growing single crystals are those of P. Bridgman and J. Czochralski. In the Bridgman method the material is melted in a vertical cylindrical vessel which tapers conically to a point at the bottom. The vessel then is lowered slowly into a cold zone. Crystallization begins in the tip and continues usually by growth from the first formed nucleus. In the Czochralski method a small single crystal (seed) is introduced into the surface of the melt and then drawn slowly upward into a cold zone. Single crystals of ultrahigh purity have been grown by zone melting. Single crystals are also often grown by bathing a seed with a supersaturated solution, the supersaturation being kept lower than is necessary for sensible nucleation.

When grown from a melt, single crystals usually take the form of their container. Crystals grown from solution (gas, liquid, or solid) often have a well-defined form which reflects the symmetry of the unit cell. For example, rock salt or ammonium chloride crystals often grow from solutions in the form of cubes with faces parallel to the (100) planes of the crystal, or in the form of octahedrons with faces parallel to the (111) planes. The growth form of crystals is usually dictated by kinetic factors and does not correspond necessarily to the equilibrium form. *See* CRYSTAL GROWTH; CRYSTALLIZATION; ZONE REFINING.

Physical properties. Ideally, single crystals are free from internal boundaries. They give rise to a characteristic x-ray diffraction pattern. For example, the Laue pattern of a single crystal consists of a single characteristic set of sharp intensity maxima. *See* X-RAY DIFFRACTION.

Many types of single crystal exhibit anisotropy, that is, a variation of some of their physical properties according to the direction along which they are measured. For example, the electrical resistivity of a randomly oriented aggregate of graphite crystallites is the same in all directions. The resistivity of a graphite single crystal is different, however, when measured along different crystal axes. This anisotropy exists both for structure-sensitive properties, which are strongly affected by crystal imperfections (such as cleavage and crystal growth rate), and for structure-insensitive properties, which are not affected by imperfections (such as elastic coefficients).

Anisotropy of a structure-insensitive property is described by a characteristic set of coefficients which can be combined to give the macroscopic property along any particular direction in the crystal. The number of necessary coefficients can often be reduced substantially by consideration of the crystal symmetry; whether anisotropy, with respect to a given property, exists depends on crystal symmetry.

The structure-sensitive properties of crystals (for example, strength and diffusion coefficients) seem

governed by internal defects, often on an atomic scale. *See* CRYSTAL DEFECTS. David Turnbull

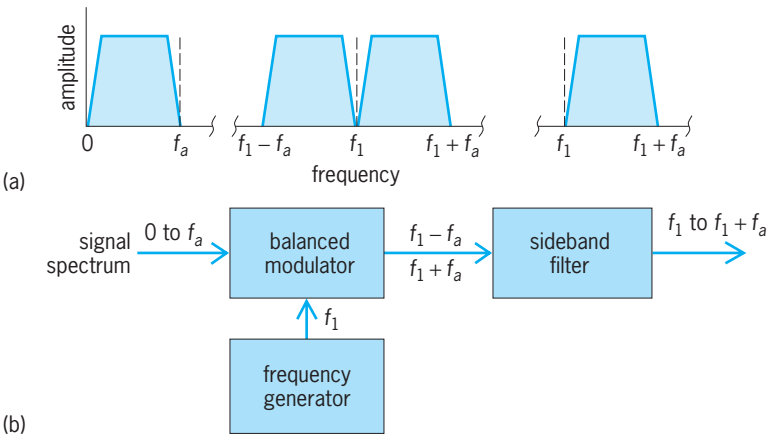
Bibliography. I. V. Markov, *Crystal Growth for Beginners*, 1995; B. R. Pamplin, *Crystal Growth*, 2d ed., 1980; J. P. Van der Eerden, *Fundamentals of Crystal Growth*, 1993.

Single sideband

An electronic signal-processing technique in which a spectrum of intelligence is translated from a zero reference frequency to a higher frequency without a change of frequency relationships within the translated spectrum. Single-sideband (SSB) signals have no appreciable carrier. After translation, only the single-sideband energy remains. This form of intelligence transmission requires distortion-free amplification of the SSB signal prior to transmission and occupies only the spectrum bandwidth of that intelligence. These advantages allow SSB to be selected for communication of voice and digital data wherever there is a premium on having a minimum of transmitted power and available frequency spectrum.

Amplitude-modulated (AM) signals have identical upper and lower sidebands symmetrically located on each side of the translation frequency, which is often called the carrier. The SSB spectrum differs from the AM spectrum in having little or no carrier and only one sideband. (A double-sideband signal is one in which only the carrier is suppressed, leaving both sidebands. This technique is not commonly used.) *See* AMPLITUDE MODULATION.

Generation. In the SSB signal-processing action, the intelligence spectrum to be translated is applied to the signal input port of a balanced modulator. A higher-frequency sinusoidal signal, often called a carrier, is applied to the other input port of this circuit. Its function is to translate the zero reference spectrum to the carrier frequency and to produce the upper and lower sidebands, which are symmetrically located on each side of the carrier. The carrier frequency power is suppressed to a negligible value by the balanced operation of the modulator and does not appear at the output. Generally, the balanced modulator operates at an intermediate frequency which is lower than the frequency of transmission. Following the balanced modulator is a sideband filter which is designed to remove the unwanted sideband signal power and to allow only the desired intelligence spectrum to pass. A block diagram and drawings of the signal characteristics at each major circuit interface are shown in the **illustration**. *See* AMPLITUDE MODULATOR; MODULATOR.



Generation of a single-sideband signal. (*a*) Spectrum. (*b*) Block diagram.

The outputs of several sideband filters having adjacent frequency passbands may be added to produce a succession of intelligence channel groups, each with differing information content. These may be translated to higher frequency and amplified for transmission; amplified and transmitted directly; or used as an input to a frequency, phase, or pulse code modulation system. *See* FREQUENCY MODULATION; MULTIPLEXING AND MULTIPLE ACCESS; PHASE MODULATION; PULSE MODULATION.

The SSB signal-handling processes are carefully designed to avoid distortion of the frequency relationships within the sideband spectrum. Amplifier circuits and components are chosen to provide a highly linear signal-handling capability. Linearity is measured by determining the level of the third-order tone and the fifth-order tone produced when two tones (f_1, f_2) occupy the intelligence channel. Typically linearity requirements allow a third-order tone level less than 1/1000 (30 decibels below) the desired tone's power level and a fifth tone almost 1/1,000,000 (60 decibels below) the desired tone's power level. Filters are designed for equal amplitude response and uniform time delay of the frequency components in the signal passband.

A technique is under development and in early commercial use which compresses both the frequency range and the amplitude range of the intelligence spectrum before translation. This technique is referred to as amplitude compandered sideband (ACSB). ACSB processing improves the signal-to-noise performance for a given transmitted power and furthermore decreases the required transmission bandwidth.

Reception. Reception of SSB signals is essentially a reverse process from signal generation. The signal energy derived from the antenna is selected for a desired frequency range by radio-frequency filtering, is amplified to a desired level, and is translated down to an intermediate frequency. The signal-processing circuitry required to derive the desired intelligence channel spectrum from the intermediate frequency requires a demodulator which is similar to the modulator of the SSB generator. ACSB reception requires an inverse signal processing of the intelligence spectrum to that used in transmission in which the original frequency and amplitude range are precisely restored.

Frequency stability requirements. Frequency stability of a high order is required of the signal generators that provide the translation frequencies in both SSB generation and reception. In the precise recovery of the intelligence spectrum there is a direct dependence upon total system frequency

accuracy in returning the received intelligence sideband to zero reference frequency. Voice communication intelligibility suffers little with a system frequency error of up to 50 Hz, although digital data may be disrupted severely with only a few hertz error. Many SSB systems use a frequency synthesizer which makes all translation frequencies dependent upon a single piezoelectric frequency generator which can be designed for frequency accuracy on the order of 1 part in 10^9 under severe environmental conditions. Atomic standards for frequency generation can provide reference frequency accuracy on the order of 1 part in 10^{12}.

The Doppler shift, caused by motion of high-velocity vehicles carrying SSB equipment, for example, supersonic aircraft or satellites, necessitates transmitting some reference frequency power (suppressed carrier) to provide at the receiver a source for frequency or phase locking of the frequency standard. In ACSB transmission and reception there is need for a high relative stability of the reference (carrier) frequency. Thus a pilot tone is transmitted, and upon reception it provides a stable frequency reference to which the translator frequency is phase-locked. *See* AUTOMATIC FREQUENCY CONTROL (AFC); DOPPLER EFFECT.

Communications. There are many advantages in the use of SSB techniques for communication systems. The two primary advantages are the reduction of transmission bandwidth and transmission power. The bandwidth required is not greater than the intelligence bandwidth and is one-half that used by amplitude modulation. The output power required to give equal energy in the intelligence bandwidth is one-sixth that of amplitude modulation.

Propagation of radio energy via ionospheric refraction and when the transmitted signal energy is reflected from terrain objects or from buildings (as in land mobile communications) provides the possibility for multiple paths of differing path length which cause a selective cancellation of frequency components at regular frequency spacings. This produces in AM a severe distortion of the intelligence because of the critically dependent carrier-to-sideband amplitude and phase relationships. Frequency-modulated (FM) multipath signals in conjunction with the motion of a vehicular receiver cause a rapid fading to occur so that when the median received signal is near the required noise quieting threshold level the receiver will emit bursts of noise.

SSB is much less affected under these conditions. SSB does not have an abrupt noise quieting threshold as does FM, since with SSB there is a nearly linear increase of noise as the signal level decreases. SSB has proved to be a highly satisfactory communication signaling method where signal-to-noise ratio levels of under 10:1 are present, and SSB can produce intelligible voice communication at less than unity signal-to-noise ratio.

SSB and its close relative ACSB are inherently less affected by interference than are either AM or FM due to the narrow bandwidth occupied by SSB/ACSB, and since the reference frequency is adjusted for clarity of the desired signal, the undesired signal appears noiselike. Thus in SSB the signal-to-noise ratio of the desired signal is reduced by interference. In FM the interfering signal either can capture the demodulator when the interference is stronger than the desired signal or can severely distort the desired signal when the interference is less than the desired signal. In AM a weaker interfering signal will appear as an intelligible signal at a lower level than the desired signal. SSB techniques are finding wide use in wire line, coaxial cable, long-range high-frequency communications, microwave multiplex, ground-to-air–air-to-ground voice communication and land mobile (ASCB) communications, and are being used in experimental satellite-to-Earth channels. High-channel-density microwave systems which formerly employed FM multiplex techniques are now being installed using SSB-modulation techniques to achieve more than twice the number of channels in the same spectrum bandwidth. SSB microwave has equal or improved performance when compared to that of FM multiplexed microwave. *See* RADIO-WAVE PROPAGATION.

History. The discovery of SSB can be credited to John R. Carson in 1915. He found from the mathematical equations that there was a preservation of the premodulation spectrum in the remaining sideband after the removal of the lower sideband and the carrier of an AM signal. He experimentally proved this by using the filtering capability of a frequency-selective antenna. His Patent No. 1,449,382, filed in 1915 and granted in 1923, recognized this. Transoceanic experiments were also in progress by 1923, and were soon followed by commercial radio telephone channels. Wider utility followed developments in stable frequency generators and improved filter technology and amplifier components.

David M. Hodgin

Bibliography. American Radio Relay League, *Radio Amateur's Handbook*, revised periodically; American Radio Relay League, *Single Sideband for the Radio Amateur*, revised periodically; J. M. Gale, *Marine SSB Operation*, 1990; National Association of Broadcasters Staff, *NAB Engineering Handbook*, 1985; W. E. Sabin and E. O. Schoenike, *Single Sideband Systems and Circuits*, 1995; M. Schwartz, *Information Transmission, Modulation, and Noise*, 4th ed., 1990.

Single-site catalysts (polymer science)

Metallocenes and related compounds which have a single catalytic site and are used to produce uniform polymers and copolymers with tailored structures and physical properties. Metallocenes are "sandwich compounds" in which a π-bonded metal atom is situated between two aromatic ring systems (**Fig. 1**). *See* CATALYSIS; METALLOCENES; POLYMER.

Classical catalysts. In 1953, Karl Ziegler succeeded in polymerizing ethylene into high-density polyethylene at standard pressure and room temperature, based on the use of the catalyst titanium

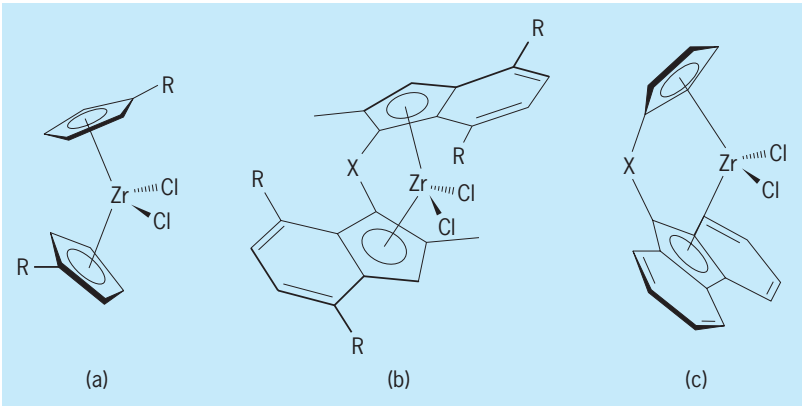
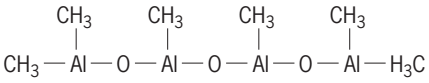


Fig. 1. Structures of metallocenes (in this case, zirconocenes). (*a*) Nonchiral structure produces atactic polypropylene. (*b*) Racemic mixture of this chiral structure produces isotactic polypropylene. (*c*) C_s-symmetric structure produces syndiotactic polypropylene. R = alkyl substitution [H, CH_3, C_2H_5, C_6H_5]; X = bridge [—CH_2—CH_2—, $(CH_3)_2Si$, $(CH_3)_2C$].

tetrachloride combined with diethylaluminum chloride. Some months later, Giulio Natta demonstrated a catalyst system that was capable of polymerizing propylene into semicrystalline polypropylene. For their work, Ziegler and Natta shared the Nobel Prize for Chemistry in 1963. Modern Ziegler-Natta catalysts are mixtures of solid and liquid compounds, often containing $MgCl_2/TiCl_4/Al(C_2H_5)_3$ and, for propylene polymerization, different internal and external donors such as ethylbenzoate, silanes, or ethers to increase stereoregularity (tacticity). An important catalyst for ethylene polymerization is the Phillips catalyst, that is, chromium trioxide on a silica support. This catalyst is used in the gas-phase polymerization of ethylene, but is unable to produce polypropylene with methyl groups always on the same side of the chain (isotactic). As Ziegler-Natta catalysts are heterogeneous and complex systems with different active sites, the resulting polymer structure can be controlled only to a limited degree. *See* HETEROGENEOUS CATALYSIS; POLYMERIZATION; POLYMER STEREOCHEMISTRY AND PROPERTIES; POLYOLEFIN RESINS.

Metallocene catalysts. In contrast to Ziegler systems, metallocene catalysts have only one active site (single-site catalysts) and produce polymers with a narrow molecular weight distribution $M_w/M_n = 2$ (M_w and M_n are the weight average and number average molecular weights). Their catalytic activity is 10–100 times higher than that of the classical Ziegler-Natta systems. In addition, metallocene catalysts are soluble in hydrocarbons or liquid propylene, and their structure can be easily changed. These properties allow one to predict accurately the properties of the resulting polyolefins by knowing the structure of the catalyst used during their manufacture, as well as to control the resulting molecular weight and distribution, comonomer content, and tacticity by careful selection of the appropriate reactor conditions. *See* MOLECULAR WEIGHT.

Metallocenes, in combination with the conventional aluminumalkyl cocatalysts used in Ziegler systems, are capable of polymerizing ethylene, but only at a very low activity. In 1977 in Hamburg, it was discovered that methylalumoxane (MAO) greatly improved catalysis with metallocenes, enhancing their activity, surprisingly, by a factor of 10,000. Methylaluminoxane



is a compound in which aluminum and oxygen atoms are arranged alternately and free valences are saturated by methyl substituents. It is prepared by the careful partial hydrolysis of trimethylaluminum and consists mainly of units of the basic structure $Al_4O_3Me_6$. As the aluminum atoms in this structure are coordinatively unsaturated, the basic units join, forming cluster compounds. These have molecular weights from 1200 to 1600 and are soluble in hydrocarbons. Metallocene catalysts, especially zirconocenes, treated with MAO can polymerize up to 100 tons of ethylene per gram of zirconium (Fig. 1). Such highly active catalysts can remain in the product. The insertion time (for the insertion of one molecule of ethylene into the growing chain) is only 10^{-5} s. A comparison to enzymes is not far-fetched. It is generally assumed that the function of MAO is to first undergo a fast ligand exchange reaction with the metallocene dichloride, thus producing the metallocene methyl and dimethylaluminum compounds. In the next step, either Cl^- or CH_3^- is abstracted from the metallocene compound by an aluminum center in MAO, thus forming a metallocene cation and an MAO anion. The alkylated metallocene cation represents the active center. Meanwhile, other weakly coordinating cocatalysts, such as tetra(perfluorophenyl)borate anions, $[(C_6F_5)_4B]^-$, have been successfully used to activate metallocenes and other single-site catalysts. Polyolefins with different microstructures and characteristics can be custom-made by varying the ligands of the metallocene (Fig. 1). By combining different olefins and cycloolefins with one another, the range of characteristics can be further broadened. Conventional heterogeneous catalysts have not yet produced polyolefins with tailored microstructures or chemically uniform copolymers. Using metallocene catalysts, it was possible for the first time to produce polyethylenes, polypropylenes, and copolymers with narrow molecular-weight distribution, syndiotactic (alternating stereoregular structure) polypropylene, syndiotactic polystyrene, cyclopolymers of 1,5-hexadiene, and cycloolefin copolymers (COCs) with high catalytic activity. *See* COPOLYMER; POLYSTYRENE RESIN.

Other single-site catalysts. "Half sandwich" metallocenes are highly active for the copolymerization of ethylene and 1-octene (**Fig. 2**). These catalysts produce long-chain branched polyolefins by incorporating oligomers with vinyl endgroups, which are formed during polymerization by β-hydrogen transfer. Long-branched copolymers of ethylene with 1-octene show elastic properties as long as the

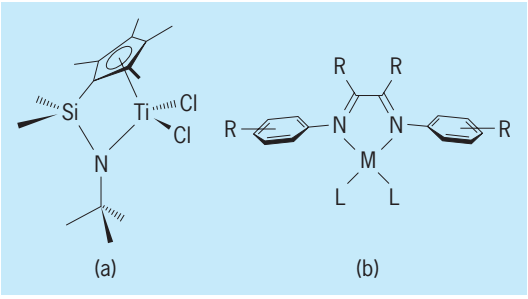


Fig. 2. Structures of (*a*) "half-sandwich" metallocenes, dimethylsilylamido(cyclopentadienyl)titanium compound for the production of linear low-density polyethylene (LLDPE), and (*b*) diimin complexes with M = Ni, Pd; R = H, alkyl; L = halogen.

comonomer content is more than 20%. Other elastomers with different microstructures can be synthesized from dienes.

New palladium, nickel, iron, and cobalt complexes activated by MAO or borate cocatalysts are also active for the polymerization of olefins. Cationic Pd(II) and Ni(II) complexes are able to polymerize ethylene and other α-olefins when sterically bulky α-diimine ligands and noncoordinating counterions are used. The lower oxophilicity and presumed greater functional group tolerance of late-transition metals relative to early-transition metals make them likely targets for the development of catalysts for the copolymerization of ethylene with polar monomers under mild conditions.

Polypropylene. Propylene and long-chain olefins can be assembled in a stereoregular manner, such that the building blocks have a defined and recurring arrangement in the resulting polymer chain. The structural arrangement of polypropylene has a considerable influence on its functional properties. A random arrangement leads to amorphous (atactic) polypropylene which flows at room temperature, whereas stereoregular polypropylene (isotactic and syndiotactic) is crystalline having a melting point of 165°C (330°F). In addition to isotactic and atactic polypropylene, syndiotactic polypropylene as well as isoblock and stereoblock polypropylene have been made for the first time in large quantities and in high purity (**Fig. 3**). The block lengths of equal tacticity in isoblock polypropylene (meaning the position of the methyl groups in one direction) can be varied in a broad range on the nanoscale (4–100 units). This leads to microcrystalline polypropylene which is suitable for making transparent films. The development of applications for elastic stereoblock polypropylene has just begun. Polypropylenes made using metallocenes exhibit distinct differences to conventionally produced polypropylenes, such as narrow molar mass distributions, higher stiffness, and greater tensile strength (see **table**). This is caused not only by the uniform structure but also by the extremely low fractions of low-molecular-weight (oligomeric) products. These fractions amount to less than 0.1%, compared to 2–4% in Ziegler-Natta polypropylene.

Elastomers. One of the biggest impacts of single-site catalysts promises to be in the manufacture of elastomers with elastic properties similar to rubber. The copolymers of ethylene and propylene with a molar ratio of 1:0.5 up to 1:2 are of great industrial interest. These ethylene-propylene copolymers (EPM) show elastic properties. In addition, ethylene-propylene-diene terpolymers (EPDM), made by including 2–5 wt % of a diene as a third monomer, are elastomers. The diene, 5-ethylidene-2-norbornene (ENB) [norbornene = bicyclo[2.2.1]-hept-2-ene], is mainly used. Since there are no double bonds in the backbone of this polymer, it is stable to oxidation and photodegradation reactions. *See* RUBBER.

Until recently, most processes for the production of EPM and EPDM rubber used soluble or highly dispersed vanadium catalysts. The elastomers produced by vanadium catalysts are pale yellow in color because of traces of remaining vanadium catalysts. Similar, less colored elastomers can be obtained with the single-site catalysts because the concentration of the single-site catalysts used for the polymerization is about a factor of 100 lower than the vanadium catalysts.

Cycloolefin copolymers. Single-site catalysts can be used to polymerize and copolymerize strained cyclic olefins such as cyclopentene or norbornene. The melting points for these homopolymers are extremely high, such as 400°C (750°F) for

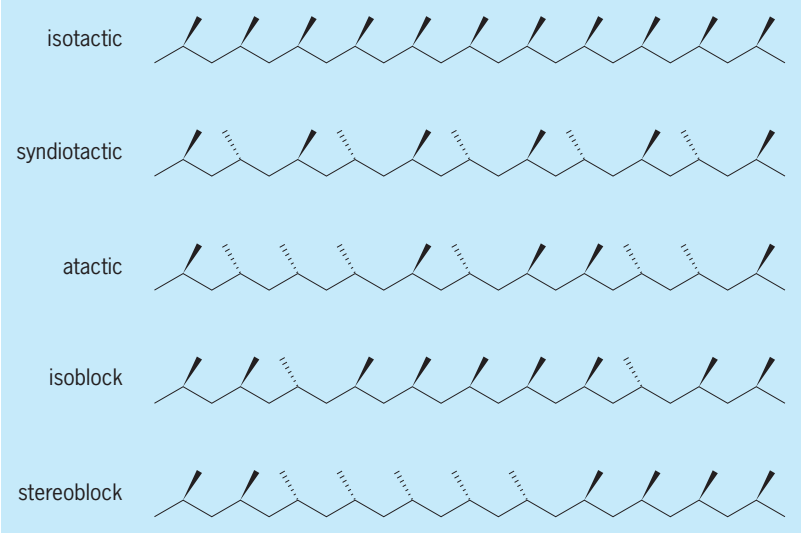


Fig. 3. Microstructures of polypropylenes.

Comparison of some properties of polypropylene made by Ziegler-Natta and metallocene catalysts

	Polypropylene from metallocene catalyst	Polypropylene from Zigler-Natta catalyst
Melting point (°C)	161	162
M_w/M_n	2.5	5.8
Stress tensile (N mm^{-2})	1620	1190
Hardness (N mm^{-2})	86	76
Extractable fraction (%)	0.1	2–4

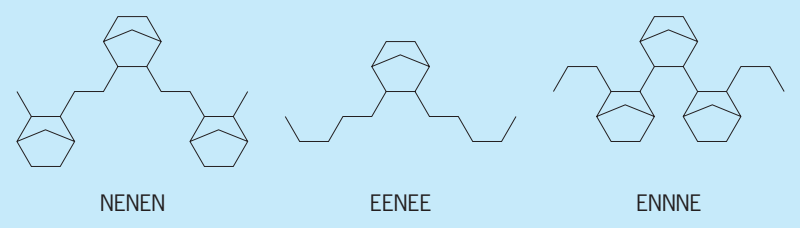


Fig. 4. Structures of norbornene (N)/ethylene (E) copolymers. Alternating blocks can also occur.

polycyclopentene, and as a result they may be used as engineering plastics. Copolymers of cyclic olefins with ethylene (COC) represent a new class of thermoplastic amorphous materials (**Fig. 4**). COCs are characterized by excellent transparency and very long life at high service temperatures. In addition they are chemically resistant and can be melt-processed. Due to their high carbon/hydrogen ratio, these polymers feature a high refractive index, for example, 1.53 for the ethene-norbornene copolymer at 50 mol % norbornene incorporation. Their stability against hydrolysis and chemical degradation, in combination with their stiffness, makes them desirable materials for optical applications such as compact discs, lenses, optical fibers, and films. W. Kaminsky

Bibliography. S. D. Ittel, L. K. Johnson, and M. Brookhart, Late-metal catalysts for ethylene homo- and copolymerization, *Chem. Rev.*, 100:1169–1203, 2000; W. Kaminsky, Highly active metallocene catalysts for olefin polymerization, *J. Chem. Soc., Dalton Trans.*, pp. 1413–1418, 1998; J. Scheirs and W. Kaminsky, *Metallocene Based Polyolefins*, vols. 1 and 2, Wiley, Chichester, 1999; K. B. Sinclair, Future trends in polyolefin materials, *Macromol. Symp.*, 173:237–261, 2001.

Sintering

The welding together and growth of contact area between two or more initially distinct particles at temperatures below the melting point, but above one-half of the melting point in kelvins. Since the sintering rate is greater with smaller than with larger particles, the process is most important with powders, as in powder metallurgy and in firing of ceramic oxides.

Powder. Although sintering does occur in loose powders, it is greatly enhanced by compacting the powder, and most commercial sintering is done on compacts. Compacting is generally done at room temperature, and the resulting compact is subsequently sintered at elevated temperature without application of pressure. For special applications, the powders may be compacted at elevated temperatures and therefore simultaneously pressed and sintered. This is called hot pressing or sintering under pressure.

Sintering is observed as an increase in mechanical properties (strength, ductility, and so on) and in many physical properties (for example, electrical and thermal conductivity). In many, but by no means all, sintering processes, the density of the compact increases, that is, the dimensions of the compacts become smaller (the compacts "shrink"), but the shape of the compact is generally preserved. The final density of the sintered compact depends upon the pressure with which the powder is compacted—the higher the pressure, the greater the density—and upon the shrinkage of the compact during sintering. Compacts from a single component powder, for example, a powder of a pure metal or a pure oxide, must be sintered below the melting point of the component, that is, without a liquid phase. Certain compacts from a mixture of different component powders may be sintered under conditions where a limited amount of liquid, generally less than 25 vol%, is formed at the sintering temperature. This is called liquid-phase sintering, important in certain powder-metallurgy and ceramic applications. *See* CERAMICS.

Mechanism. The driving force in sintering is surface energy, which decreases because the total surface area decreases as sintering proceeds. When two spheres originally in tangential contact sinter, the area of contact between the spheres increases and a neck is formed. In amorphous materials such as glasses, the process causing neck formation is viscous flow (as when two drops of water coalesce to form one). This type of viscous flow is not possible in crystalline solids. Here the most important material transport process is self-diffusion. Because the material in the neck surface has a highly convex curvature, the number of defects in its crystal structure (vacant lattice sites) is considerably higher than on a flat or concave surface. These defects move by self-diffusion from the convex neck surface to the adjacent flat surface, which means material moves in the opposite direction, that is, the neck grows (**Fig. 1***a*). Instead of coming from the adjacent flat surface, the material forming the neck may also move by self-diffusion from the grain boundary between the two spheres to the convex neck surface (Fig. 1*b*). This latter type of movement explains why compacts shrink, because in this case the centers of the spheres approach each other during sintering.

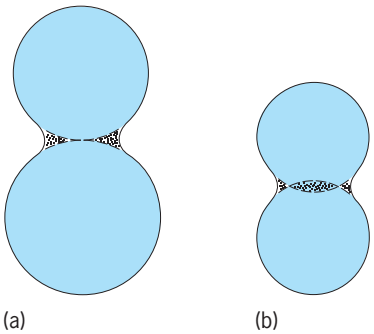


Fig. 1. Neck formation. (*a*) Neck growth through movement of defects from the neck surface to the adjacent flat surface. (*b*) Movement of neck-forming material from the grain boundary to the neck surface. (*After G. H. Gessinger et al., Continuous observation of the sintering of silver particles in the electron microscope, ASM Trans. Quart., 61(3):598–604, 1968*)

Densification (shrinkage of compacts) is inhibited by rapid grain growth during sintering. During grain growth the grain boundaries, which are the sources of the material which is transported into the convex neck surface, that is, into the pores of the compact, are swept out of the compact.

Liquid. The mechanism of liquid-phase sintering is more complicated. It involves viscous flow of the liquid phase, but also solution of the solid phase in the liquid phase and its reprecipitation in such a way as to make the compact more dense. As in solid-phase sintering, the driving force in liquid-phase sintering is a decrease in surface energy. *See* CERMET; POWDER METALLURGY. F. V. Lenel

Boundary migration. Powder compacts for sintering are usually made of mixtures of elemental particles, which form alloys as composition change occurs during the sintering treatment. When solute atoms diffuse into the grain boundaries, either present in the original particles or formed in the early stage of the sintering treatment, the grain boundaries may migrate, forming alloyed regions behind them. Such grain boundary migration has been observed in a number of metallic and ceramic materials, and is usually called diffusion-induced grain boundary migration (DIGM) [**Fig. 2***a*]. Solute atoms are supplied from vapor, melt, or solid. During liquid-phase sintering, thin liquid films form between the grains and may migrate if the grains are not chemically in equilibrium with the liquid (Fig. 2*b*). This phenomenon, called chemically induced liquid-film migration or simply liquid-film migration (LFM), shows characteristics similar to DIGM. Liquid-film migration can be induced either by adding solute atoms to the liquid phase or by changing the temperature. If the amount of liquid in the grain-liquid alloy is large at high temperatures, the solid grains may become corrugated by reprecipitation on certain surface regions and dissolution of adjacent regions. When the composition change by solute addition is large, new grains often form and grow, resulting in diffusion-induced recrystallization (DIR) [Fig. 2*c*].

Mechanism. Both DIGM and LFM occur when there is depletion of solute atoms from an alloy, such as by the evaporation of zinc from the surface of a polycrystalline iron-zinc alloy, and both occur at relatively low temperatures, where lattice diffusion of the solute atoms is not rapid. Often the boundaries move against their curvatures. As for any reaction to equilibrium, the driving force for DIGM (and LFM) arises from a decrease in the chemical free energy associated with the composition change. It has been proposed and verified in some experiments that solute atoms diffuse into the grains in front of the moving boundaries and produce coherency strains, which become the driving force for migration. The migration velocity then depends on the size and concentration change of the solute atoms. If the boundaries do not migrate, lattice diffusion is the dominant mode of composition change. The mechanism of forming new grains in DIR is not well understood, but once the grains form, their boundaries move by the same mechanism as for DIGM and LFM.

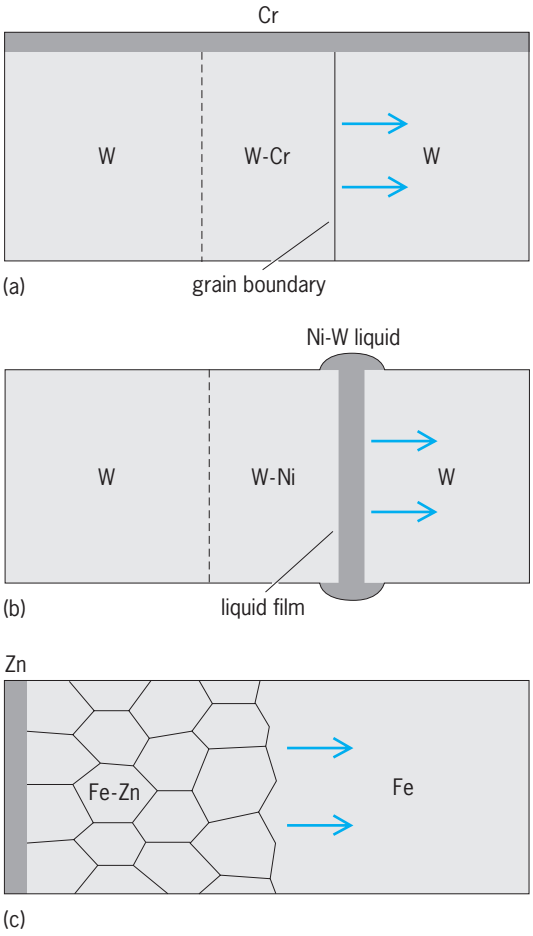


Fig. 2. Boundary migration. (*a*) Diffusion-induced grain boundary migration in tungsten-chromium (W-Cr). (*b*) Liquid-film migration in tungsten-nickel (W-Ni). (*c*) Diffusion-induced recrystallization in iron-zinc (Fe-Zn). The initial positions of the grain boundary and liquid film are shown by broken lines and the migration directions by arrows.

Applications. Both DIGM and LFM and possibly DIR are found in a variety of practical processes, such as heat treatment, welding, oxidation, sputtering, and melting. In the alloys prepared by liquid-phase sintering, rapid microstructural and compositional changes occur by DIGM and LFM during cooling from the sintering temperature. In electronic ceramics, the boundaries migrate during infiltration with a liquid phase. Such microstructural and composition changes produce significant effects on the resulting properties. The overall composition changes by DIGM, LFM, and DIR in polycrystals, including those produced by sintering, are much more rapid than that expected by diffusion only along stationary boundaries and in the lattice. *See* ALLOY; DIFFUSION. W. Gust; D. Y. Yoon

Bibliography. G. Flight, *Introduction to Ceramics*, 1990; W. D. Kingery, *Introduction to Ceramics*, 2d ed., 1976; B. Kubicki, *Sintered Machine Elements*, 1994; G. C. Kuczynski (ed.), *Sintering*, 1980; G. C. Kuczynski (ed.), *Sintering Processes*, 1980; E. Rabkin, C. Y. Ma, and W. Gust, Diffusion-induced grain boundary phenomena in metals and oxide ceramics, in J. Novotny (ed.), *Science of Ceramic*

Interfaces II, pp. 353–369, Elsevier, Amsterdam, 1994; L. H. Yaverbaum (ed.), *Technology of Metal Powders: Recent Developments*, 1980; D. Y. Yoon, Theories and observations of chemically induced interface migration, *Int. Metall. Rev.*, 40:149–179, 1995.

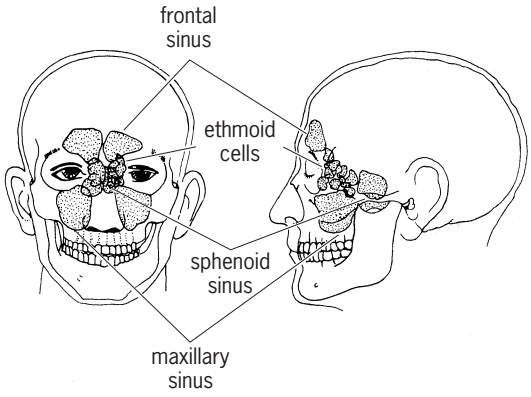
Sinus

Any space in an organ, tissue, or bone, but usually referring to the paranasal sinuses of the face. In humans, four such sinuses (see **illus.**), lined with ciliated, mucus-producing epithelium, communicate with each nasal passage through small apertures. The ethmoid and sphenoid sinuses are located centrally between and behind the eyes. The frontal sinuses lie above the nasal bridge, and the maxillary sinuses are contained in the upper jaw beneath the orbits. The mastoid portion of the temporal bone contains air cells lined with similar epithelium.

These sinuses vary in size and shape with the individual and with age. They are minute or absent in the infant, and usually develop to adult size as the skull bones enlarge after eruption of the permanent teeth.

Fluid or inflammatory products may accumulate in the sinuses, altering the voice resonance and producing headache or pain. Such accumulations are frequently the result of acute or chronic infections, or sinusitis. In this type of inflammation, abundant mucus is produced, which cannot drain into the nose because the narrow ducts are obstructed by the swollen mucous membrane. Pressure rises in the sinus, causing pain in the form of localized headaches. The localization varies according to the sinus involved. Extension of the infection to the neighboring structures becomes dangerous because of the close relationship to the brain. In chronic sinusitis, polyps can develop that protrude into the nose.

The membranes of the sinus of humans are well supplied with blood, lymphatics, and nerves, particularly nerves sensitive to pain and pressure. Erect posture decreases natural drainage of the human paranasal sinuses.



Frontal and side view of face, showing sinuses. (*After N. L. Hoerr and A. Osol, eds., Gould Medical Dictionary, 2d ed., Blakiston–McGraw-Hill, 1956*)

Similar sinuses are found in most mammals and may be very large, as in elephants, in which they serve to lighten the head, but they have no known function in humans. Most submammalians lack them, but crocodilians have comparable cavities emptying into the nasal cavities. Thomas S. Parsons

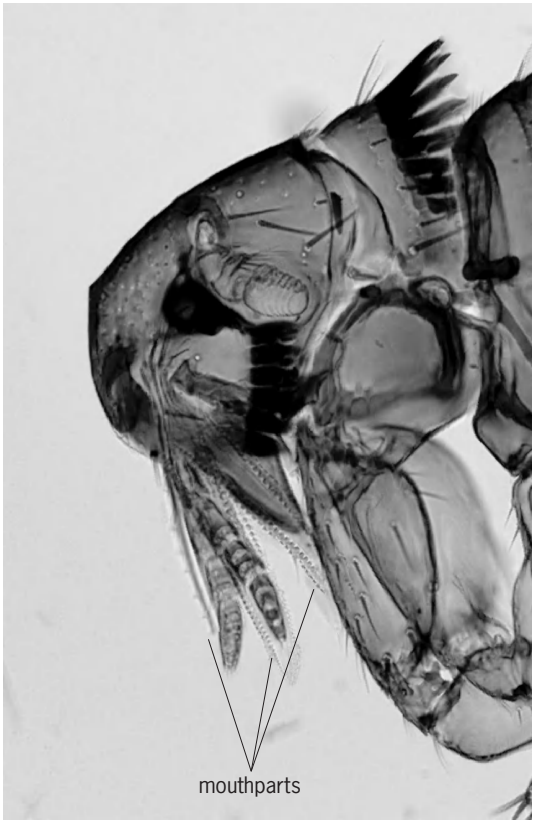
Siphonaptera

An order of insects commonly known as fleas. These insects are bloodsucking ectoparasites of animals and humans and may transmit serious diseases.

Adult fleas are easily recognized. They are small, wingless, about $^1/_8$ in. (3 mm) in length, dark brown or black, and laterally flattened. They have three pairs of legs, with the third pair modified for jumping. The body of a flea is characteristically oval-shaped and armed with combs, setae, and spines, which prevent the flea from being dislodged from the host. The head bears mouthparts adapted for bloodsucking (see **illustration**).

Life history. Fleas undergo complete metamorphosis (holometabolous); that is, their life history involves four distinct phases: the egg, larva, pupa, and adult. The only exception is the neosomic *Tunga monositus* (sand flea), which has only two stages. Female adult fleas may lay large numbers of eggs (400 or more). Flea eggs are white, ovoid, and nonadhesive. They are laid in the bedding or nest of the host, where they eventually hatch into the larva. Fleas have three larval stages (instars), which are characterized by differences in size (1.5 – 10 mm). The first larval stage has an egg tooth, which is used to open the eggshell. Flea larvae are eyeless and legless, and sparsely covered with bristles. They usually bear an anal comb of bristles on the tenth segment. Only the head and mouthparts are sclerotized. Larvae are seldom seen because they usually hide in the nesting material or in the surroundings of the host. They feed on organic material, mainly droplets of dried, digested blood from the adults. When fully developed, the larva spins a silken cocoon with incorporated debris, to form the pupal stage. This is a nonfeeding, resting period, which may last from a few days to several months, often depending on host availability. The adult flea emerges from the pupa. Under favorable conditions, the entire life cycle may be completed in as little as 2 weeks.

Adult fleas feed exclusively on the blood of the host. They have specialized sucking mouthparts with three piercing blades (stylets) see illustration. Adult fleas have three-segmented antennae, and their eyes are rudimentary. Most species possess combs on their head or the first thoracical segment. The thorax bears three pairs of legs. The third pair, in combination with an elastic structure (pleural arch) within the thorax, enables the flea to jump. The flea abdomen is divided into tergites (dorsal plates) and sternites (ventral plates), which are connected through flexible tissue that enables the abdomen to expand during a blood meal.



Head of *Spilopsyllus cuniculi* (rabbit flea), with extended mouthparts.

The last abdominal segment has a sensory plate, called the sensillium, containing a variety of sensory pits, whose function is mainly unknown. Adult fleas spend considerable time off the host, and may be found in or around the bedding or nest of the host. They only visit the host to feed, and may feed multiple times during one blood meal.

Medical importance. During the three decades following World War II, the study of fleas increased, and the world flea fauna became very well known; specialists now recognize 2575 species, belonging to 244 genera (see **table**). The great majority of fleas occur on mammals (95%); the remainder live on birds. The host specificity of fleas, as well as their seasonal and geographical distribution, is largely dependent on specific requirements of the larvae. The least host specificity is shown by those taxa whose immature stages tolerate wide ranges of temperature, humidity, and nutritional requirements (for example, cat fleas). Adult fleas may occasionally attack humans if given the opportunity of contact. Since many fleas are carriers and transmitters (vectors) of disease, humans can contract infections through their bite.

Plague, which is responsible for millions of deaths throughout the history of humankind, is a disease of rats It is transmitted between rats and from rats to humans by fleas, particularly the oriental rat flea (*Xenopsylla cheopis*). This is a cosmopolitan species occurring on rats throughout the world. Plague occurs naturally in the wild (sylvatic plague) and usually affects rodents, such as ground squirrels, chipmunks, and prairie dogs of the western United States (Colorado, New Mexico, Arizona, Wyoming). Although many species of fleas are capable of transmitting sylvatic plague, human cases are rare, perhaps because humans do not often come in contact with the animal and insect reservoirs.

Murine or endemic typhus fever, another disease transmitted from domestic rats to humans by rat fleas, was of great public health significance in the southeastern United States during the period 1944–1950, when thousands of cases occurred. Tularemia, or rabbit fever, a disease of small wild animals, particularly of the western and midwestern United States, may be transmitted to humans by fleas. However, this disease, which is not common, is more often contracted while skinning infected animals. Flea allergy is the term applied by physicians to the severe reactions which sometimes result from flea bites. This disease is regarded as a major dermatological problem in California. Often the urticarial papular eruption resulting from flea bites in allergic persons is mistakenly diagnosed as hives and attributed to some food. The fleas involved are those species that most often occur as household pests, the dog flea (*Ctenocephalides canis*), the cat flea (*C. felis*), and the human flea (*Pulex irritans*). In tropical America a flea known as the chigoe, nigua, or sand flea (*Tunga penetrans*) can penetrate into the skin between the toes of humans and cause festering sores which may lead to tetanus and gangrene. The sticktight, or chicken flea (*Echidnophaga gallinacea*), is similar to the nigua but attacks the heads of poultry, cats, and dogs and sometimes annoys

Important species of Siphonaptera

Scientific name	Common name	Importance
Ctenocephalides canis	Dog flea	Intermediate host of tapeworm, readily attacks humans, cosmopolitan, rare flea
Ctenocephalides felis	Cat flea	Common household pest, intermediate host of tapeworm, readily attacks humans, very abundant on cats and dogs, cosmopolitan
Leptopsylla segnis	Mouse flea	Common on mice and rats, may transmit plague and tularemia among rodents, cosmopolitan
Oropsylla montana		Transmits sylvatic plague and tularemia among rodents, common on prairie dogs
Pulex irritans	Human flea	Probably originated with caviomorph rodents, attacks humans, not as common as cat and dog fleas, transmits plague and murine typhus
Xenopsylla cheopis		Main vector of the plague and endemic typhus, cosmopolitan

humans. Fleas are also intermediate hosts of cat and dog tapeworms, and children may become infected by accidentally swallowing dog or cat fleas. *See* PLAGUE; TULAREMIA.

Control of fleas. Bubonic plague and murine typhus fever are controlled by measures directed against rodents, such as poisoning, trapping, and rodent-proofing buildings, and by the application of carbaryl or bendiocarb dust to rodent runs and harborages to kill rodent fleas. Dog and cat fleas can also be controlled through application of the same pesticides. To kill all stages of dog and cat fleas, cellars, outbuildings, and living quarters should be sprayed with bendiocarb (wettable powder), carbaryl (wettable powder), chloropyrifos (emulsifiable concentrate), diazinone (emulsifiable concentrate), and pyrethrum (aerosols, spray). Simultaneously, pets should be treated by veterinarians to control fleas on their bodies. Flea collars impregnated with carbamate or organophosphate insecticides can be used to repel fleas. In humans, shoes and overlapping light-colored clothing should be worn as prophylactic measures along with application of repellent spray containing diethyltoluamide (DEET) on clothing and exposed skin.

Katharina Dittmar; Irving Fox

Bibliography. R. Lewis, Resume of the Siphonaptera (Insecta) of the world, *J. Med. Entomol.*, 35:377–389, 1998; A. Mallis, S. A. Hedges, and D. Moreland, *Handbook of Pest Control*, 9th ed., 2004; D. K. Mardon, *Pygiopsyllidae: An Illustrated Catalogue of the Rothschild Collection of Fleas (Siphonaptera) in the British Museum (Natural History)*, British Museum, vol. VI, 1981; F. G. A. M. Smith, *Malacopsyllidae and Rhopalopsyllidae: An Illustrated Catalogue of the Rothschild Collection of Fleas (Siphonaptera) in the British Museum (Natural History)*, Brithish Museum, vol. VII, 1987; R. Traub, M. Rothschild, and J. Haddow, *The Rothschild Collection of Fleas: The Ceratophyllidae—Key to the Genera and Host Relationships*, Cambridge University Press, 1983.

Siphonophora

An order of the class Hydrozoa of the phylum Coelenterata, characterized by an extremely complex organization of components of several different types, some having the basic structure of a jellyfish, others of a polyp. The components may be connected by a stemlike region or may be more closely united into a compact organism. The medusalike components include swimming bells, a gas-filled float, bracts which are tough gelatinous bodies presumed to be protective, and gonophores which produce the eggs and sperms. Polyplike components include gastrozooids with mouth and one long hollow tentacle; dactylozooids without mouth and often with one tentacle; and gonozooids which bear the gonophores.

Most siphonophores possess a float and are animals of the open seas. Best known is the Portuguese man-of-war, *Physalia*, with a float as much as 16 in. (40 cm) long and tentacles which extend downward for many meters. These animals may be swept shoreward and may make swimming not only unpleasant but dangerous. Some siphonophores propel themselves by jets of water produced by the swimming bells. *See* COELENTERATA; HYDROZOA.

Sears Crowell

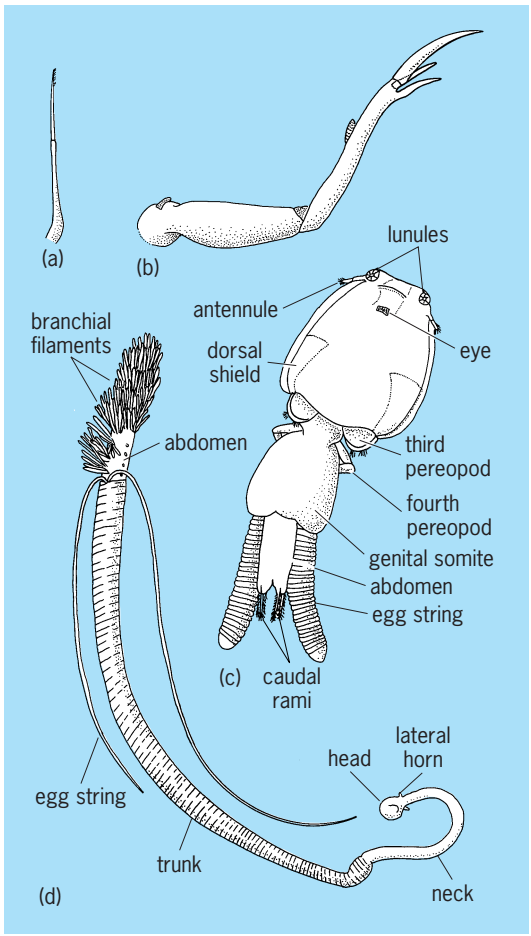
Siphonostomatoida

An order of Copepoda; all members are parasites of marine and fresh-water fishes or a variety of invertebrate hosts. In fact, it is estimated that 67% of copepod parasites of fishes belong to the Siphonostomatoida. Their obvious success in the parasitic mode of life appears at least in part to be a result of two morphologic adaptations. The first is the modification of the buccal apparatus (mouth and oral appendages). The siphonostomatoid mouth forms a tube or siphon through the partial or complete fusion of the labrum ("upper lip") and labium ("lower lip") around the esophageal opening. The resulting mouth cone has a small opening near the base through which the mandibles are free to enter. The mandibles themselves are styliform (see **illus.**) and may or may not be armed with denticles. The maxilliae are subchelate or brachiform, superficially resembling a human arm, and serve as the appendage for attachment to the host. The second adaptation is the development of a frontal filament, which is a larval organ of attachment. In those siphonostomatoids possessing a frontal filament, the brachiform second maxilla may be used to manipulate the frontal organ, and in some species fusion of those two structures forms the attachment structure.

Although siphonostomatoids all share the highly specialized character of a tube- or siphonlike mouth that is well adapted to a parasitic mode of life, other morphologic aspects are quite diverse, and so it has been speculated that the order is not in fact monophyletic. Diversity makes it impossible to give a generalized account of the order. Most familiar are the ectoparasites of fishes belonging to the family Caligidae and the mesoparasites of the Pennellidae that embed in the tissues of marine mammals (see illus.).

Those siphonostomatoids parasitic on invertebrates typically resemble most closely the cyclopoids in having primitive podoplean body segmentation and reduction of the fifth thoracic legs. The remaining siphonostomatoids, all parasites of vertebrates, are more highly modified, often bearing little resemblance to Copepoda at all. Most specializations have involved adaptations of the cephalic appendages to serve as holdfasts and body streamlining that would enhance adhesion to the body surface of the host.

It has been hypothesized that the Siphonostomatoida evolved from the podoplean line much earlier than the other major order of parasitic Copepoda,



Siphonostomatoida. (*a*) Representative siphonostomoid mandible; (*b*) brachiform maxilla (*after Z. Kabata, Parasitic Copepoda of British Fishes, Ray Society, London, vol. 152, 1979*). (*c*) Representative female caligid; (*d*) female Pennella (*after P. A. McLaughlin, Comparative Morphology of Recent Crustacea, W. H. Freeman, 1980*).

the Poecilostomatoida. Their pathway presumably involved an epibenthic, epiphytic, or epizoic ancestor of the Cyclopoida, in which the siphonlike mouth cone developed first in invertebrate parasites where the hosts were at least temporarily stationary and in contact with a firm substrate. *See* COPEPODA; CRUSTACEA; POECILOSTOMATOIDA. Patsy A. McLaughlin

Bibliography. Z. Kabata, *Parasitic Copepoda of British Fishes*, Ray Society, London, vol. 152, 1979; B. M. Marcotte, Evolution within the Crustacea, pt. 2: Copepoda, in L. G. Abele (ed.), *The Biology of Crustacea*, vol. 1: *Systematics, the Fossil Record, and Biogeography*, 1982.

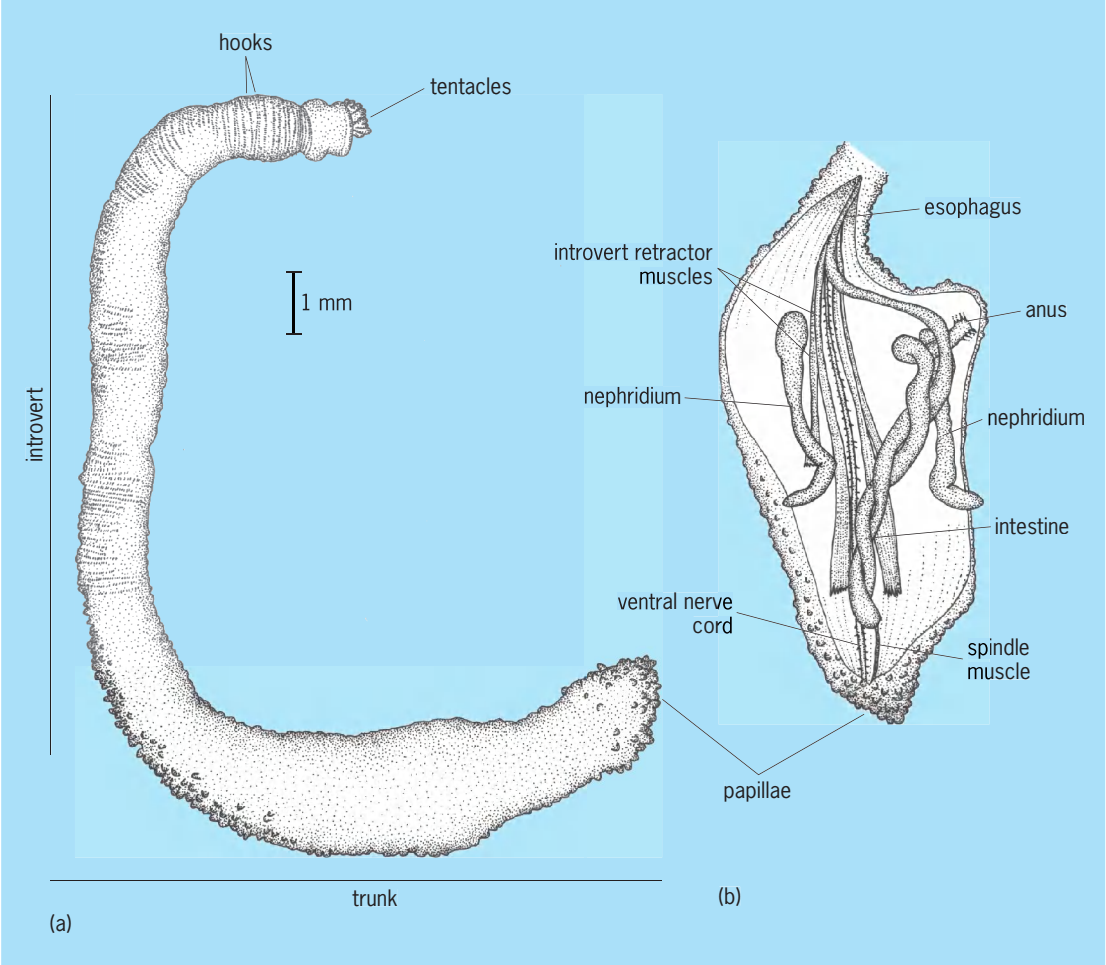
Sipuncula

A phylum of sedentary marine vermiform coelomates that are unsegmented, but possibly distantly related to the annelids; they are commonly called peanut worms. There are 17 genera and approximately 150 species living in a wide variety of oceanic habitats within the sediment or inside any protective shelter such as a discarded mollusk shell, foraminiferan test, or crevice in rock or coral.

Anatomy. Sipunculans have two body regions: the trunk and a retractable introvert (see **illustration**). The mouth is at the tip of the introvert and in one class is surrounded by tentacles of varying degrees of development and complexity. Behind the tentacular region is a zone which may bear posteriorly directed hooks either scattered or arranged in regular rings. The introvert varies from less than half the trunk length in some species to many times its length in others. Adult sipunculans range in trunk length from 2 to over 500 mm (0.08 to over 20 in.). The shape of the body ranges from almost spherical to a slender cylinder. Sipunculans have a variety of epidermal structures (papillae, hooks, or shields). Many species lack color, but shades of yellow or brown may be present.

Internal anatomy is relatively simple. The digestive tract has a straight esophagus and a double-coiled intestine extending toward the posterior end of the body and back (see illustration) terminating in a rectum, sometimes bearing a small cecum. The dorsal anus is located at the anterior end of the trunk, except in one genus (*Onchnesoma*) in which it is on the introvert. Near the anus a threadlike spindle muscle originates from the body wall and extends down the center of the gut coil, increasing its stability (see illustration). This extends to the posterior end of the trunk in many genera. A pair (only one in two genera) of simple, saclike metanephridia open ventrolaterally at the anterior end of the trunk. The longitudinal muscles on the inside of the body wall form a smooth, uniform layer in some genera, but are gathered into bundles in most. Four retractor muscles (sometimes fused to form two or rarely one muscle) control the introvert. A ventral nerve cord (see illustration) with lateral nerves and circumenteric connectives to the pair of cerebral ganglia are present. Two or four pigmented eyespots may be present on the cerebral ganglia, and a chemoreceptor (nuchal organ) is usually present.

Reproduction. The reproductive biology of sipunculans has been examined in representatives of all families. The timing and length of the reproductive period varies depending on species and geographic region. Small gametes are produced from a transient strip of tissue at the base of the ventral retractor muscles during the breeding season. They are released into the spacious coelom at an early stage, where they undergo the remainder of their growth and differentiation. One species is known to be monoecious (containing both male and female gametes), one is known to be parthenogenic (eggs develop without sperm), and two are capable of reproducing by asexual, transverse fission. Most species of sipunculans are dioecious (have two sexes) that are similar that is, they lack any sexual dimorphism. The sipunculan gametes are retained in the coelom until mature, then released through the nephridiopores, fertilization occurring externally. Free-swimming trochophore larvae generally occur, but some brooding of early stages within the maternal "home" has been



Sipunculan body regions. (*a*) External morphology and (*b*) internal anatomy of *Phascolosoma nigrescens*. (*Illustration by Anju Schulze*)

observed in *Nephasoma minuta* and may occur in other cold- or deep-water forms. The unique long-lived pelageosphera larvae are produced by some genera, allowing for transoceanic dispersal. After metamorphosis, the juveniles settle onto a suitable substratum and probably remain in that immediate vicinity throughout their lives, although there is no knowledge of growth rate or longevity.

Ecology. Most sipunculans are deposit feeders, but a few (sucnas *Themiste*) are filter feeders with elaborate tentacular crowns. Therefore, these worms play a part in the recycling of detritus and probably consume smaller invertebrates in the process. They are in turn preyed on by fishes and probably other predators (including humans in China). *See* FEEDING MECHANISMS (INVERTEBRATE).

Taxonomy. In recent years the phylogenetic relationships of the members of the phylum Sipuncula have begun to be investigated by analysis of deoxyribonucleic acid (DNA) sequence data from several different genes. While this work is currently in progress, preliminary results do support the monophyly of the phylum. However, support for all of the existing higher taxa (orders, families, and genera) is not universal, and it is likely that in the near future some changes will be proposed to the currently accepted system, which uses morphological characters to define two classes: Sipunculidea and Phascolosomatidea. Sipunculidea, with peripheral tentacles around the mouth, and simple scattered hooks when present, contains two orders (Golfingiiformes and Sipunculiformes), four families, and 11 genera. Phascolosomatidea, without peripheral tentacles, have short tentacles around the dorsal nuchal organ, and complex hooks in rings when present. It contains two orders (Phascolosomatiformes and Aspidosiphoniformes), two families, and six genera. Edward B. Cutler; A. Schulze

Bibliography. E. B. Cutler, *The Sipuncula: Their Systematics, Biology, and Evolution*, Cornell University Press, Ithaca, New York, 1994; P. E. Gibbs and E. B. Cutler, A classification of the phylum Sipuncula, *Bull. Brit. Mus. Nat. Hist. (Zoology)*, 52(1):43–58, 1987; A. B. Maxmen et al., Evolutionary relationships within the protostome phylum Sipuncula: A molecular analysis of ribosomal genes and histone H3 sequence data, *Mol. Phylogen. Evol.*, 27(3):489–503, 2003; M. E. Rice, Sipuncula, in K. G. Adiyodi and R. G. Adiyodi (eds.), *Reproductive Biology of Invertebrates*, pp. 263–280, Oxford and IBH, Kerala,

1988; M. E. Rice, Sipuncula, in F. W. Harrison and M. E. Rice (eds), *Microscopic Anatomy of Invertebrates*, vol. 12: *Onychophora, Chilopoda, and Lesser Protostomata*, pp. 238–325, Wiley-Liss, New York, 1993.

Siren (acoustics)

A sound source that is based on the regular interruption of a stream of fluid (usually air) by a perforated rotating disk or cylinder. The components of a siren are a source of air, a rotor containing a number of ports which interrupt the airflow at the desired frequency, and ports in a stator through which the air escapes. The air is supplied by a compressor, and a motor drives the rotor. The frequency of the sound wave produced by the siren is the product of the speed of rotation and the number of ports in the rotor. The shape of the rotor and stator ports determines the wave shape at the entrance of the stator port. The stator ports feed into a horn in order to improve radiation. Siren performance parameters are sound power output, acoustic pressure, and efficiency, that is, the ratio of acoustic power output to compressor power. *See* SOUND PRESSURE.

The first sirens were developed in the first half of the nineteenth century. They provided a sound source with a known and variable frequency and were used to study pitch and beats. Sirens were first used as foghorns and air warning devices in the latter half of the nineteenth century.

Sirens are used as high-intensity sound sources for sonic aerosol applications. Examples are acoustic agglomeration (the coagulation of aerosol particles in the presence of an acoustic field), combustion enhancement (for example, that of coal particles due to increased heat and mass transfer caused by the acoustic wave), and acceleration of gaseous reaction rates. Other applications of sirens include acoustic levitation (the use of radiation pressure to levitate small objects), broadband underwater sound projectors, and sonic fatigue (fatigue life and failure of structures subjected to fluctuating pressures generated by acoustic waves). *See* ACOUSTIC LEVITATION.

Electromechanical sirens use an electric motor instead of a compressed air supply to generate the acoustic signal. A second motor spins the rotor. The stator and horn increase the sound power output and efficiency. Electromechanical sirens are widely used as warning devices. *See* SOUND.

Bart Lipkens

Bibliography. R. T. Beyer, *Sounds of Our Times: Two Hundred Years of Acoustics*, Springer-Verlag, New York, 1999; T. F. Hueter and R. H. Bolt, *Sonics*, Acoustical Society of America, 2000.

Sirenia

An order of herbivorous aquatic placental mammals, commonly known as sea cows, that includes the living manatees and dugongs and the recently exterminated Steller's sea cow. The order has an extensive fossil record dating from the early Eocene Epoch, some 50 million years ago.

West Indian manatee (*Trichechus manatus*) and calf. (*U.S. Fish & Wildlife Service/Galen Rathbun*)

The earliest known sirenians were quadrupedal and capable of locomotion on land. Fossils clearly document the evolutionary transition from these amphibious forms to the modern, fully aquatic species, which have lost the hindlimbs and transformed the forelimbs into paddlelike flippers. The living species have streamlined, fusiform bodies with short necks and horizontal tail fins like those of cetaceans, but no dorsal fins (see **illustration**). The skin is thick and nearly hairless. The nostrils are separate, and the ears lack external pinnae. The apex of the heart is bifid. The thorax and lungs are elongate and the diaphragm is nearly horizontal so that the buoyant lungs lie wholly dorsal to the abdominal viscera. The skeleton typically shows both an increase in the bulk of individual bones (pachyostosis) and an increase in the proportion of compact bony tissue (osteosclerosis). These conditions provide ballast to neutralize the buoyancy of the lungs, which do not collapse during submergence in these shallow-diving animals. The manus has five digits, and bears small nails in two of the living manatee species. The organs of feeding include large, fleshy lips; numerous tactile and prehensile vibrissae; and horny pads covering the toothless anterior parts of the palate and mandibular symphysis (area of union of bones). The cheek teeth are primitively low-crowned, with two rows of rounded cusps, but the dentitions of the Recent species are divergently specialized from this primitive pattern. Where incisors are present, the first upper incisor is the largest (and in later species the only one), and it often forms an enlarged tusk. Sirenians typically feed on aquatic angiosperms, especially seagrasses (Hydrocharitaceae and Potamogetonaceae), but in ecologically marginal situations they also eat algae and even some animal material. They are normally found in tropical or subtropical marine waters, but some have become adapted to fresh water or colder latitudes. Body sizes have ranged from less than 3 m up to 9–10 m (10 to 30–33 ft). Sirenians mate and give birth in the water, bearing a single calf (occasionally twins) after about 13–14 months of gestation and then nursing it from one pair of axillary

mammae. The closest relatives of sirenians among living mammals are the Proboscidea (elephants). *See* PROBOSCIDEA.

A classification scheme is given below.

Order Sirenia
  Family: Prorastomidae
    Protosirenidae
    Trichechidae
    Subfamily: Miosireninae
      Trichechinae
  Family Dugongidae
    Subfamily: Halitheriinae
      Hydrodamalinae
      Dugonginae

Primitive sirenians. Sirenians are presumed to have originated in the Old World, along the shores of the former Tethys seaway. However, the most primitive fossils of the order have turned up by chance in Jamaica (*Prorastomus* and related forms). These were pig-sized amphibious creatures with long bodies, short legs, and narrow snouts that were not deflected from the palatal plane. They were apparently selective browsers on assorted aquatic plants. The dentition of these and other Eocene sirenians is noteworthy in having five premolars like the most primitive placental mammals. Whether this condition is really primitive or a sirenian specialization is unresolved. Some prorastomids already show weakening of the connection between the spinal column and the hindlimbs. Protosirenids, known from the Eocene of North Africa, Asia, Europe, and possibly North America, show further reduction of strength of the hindlimbs and (like all later sirenians) significant downturning of the snout (an adaptation for bottom feeding). *See* DENTITION.

Dugongids. Known from the middle Eocene to the present, dugongids have been the most diverse and successful sirenian family. The Eocene fossils of this group show a steady reduction of the hindlimb and the pelvic bones until the body takes on, by the end of the Eocene, the outward form seen in modern sea cows. Fossil dugongids have been found in tropical or formerly tropical areas around the globe, and seem always to have inhabited marine waters. Their anatomically most conservative members are placed in the subfamily Halitheriinae, which gave rise to the Dugonginae in the Oligocene Epoch and the Hydrodamalinae in the Miocene Epoch.

Dugongines apparently arose in the West Atlantic and the Caribbean regions but quickly spread to the Old World. Many of them evolved large tusks that were probably used for digging up seagrass rhizomes. They are represented today by the single species *Dugong dugon*, found in the Indian and southwest Pacific oceans. It is notable for its reduced, peglike cheek dentition: the enamel crowns of the molars wear off quickly and the chewing surface is then formed only by the roots, which in the case of the last two molars are ever growing.

Hydrodamalines evolved in the North Pacific Ocean and remained confined to that region. Over the course of the Miocene and Pliocene epochs, they adapted to cooler climates and higher-energy habitats. The lineage culminated in Steller's sea cow (*Hydrodamalis gigas*), the largest of the sirenians, which tolerated temperate and cold waters and fed mainly on kelp (large brown and red algae). It was completely toothless, and its highly modified, hooklike flippers lacked phalanges. It was exterminated by human hunting around 1768.

Trichechids. Apart from the extinct and poorly known miosirenines, which inhabited northern Europe and may have been shellfish eaters, the trichechids comprise the sirenians known as manatees. Trichechids may have arisen from primitive dugongids, in the late Eocene Epoch. The fossil record of manatees is poor, but the group seems to have evolved mainly in South America before spreading to North America, and even crossing the Atlantic to West Africa, late in the Tertiary Period. While still in South America, manatees adapted to a diet composed largely of abrasive aquatic grasses (Gramineae) by evolving the ability to replace their cheek teeth without limit. As the teeth become worn, they are shed at the front of the toothrow while replacement molars erupt at the rear, and the whole row shifts forward like a conveyor belt. This form of horizontal tooth replacement has been misleadingly likened to that of elephants, but the only mammal to have evolved a truly similar replacement process is a species of Australian wallaby. Today, manatees are represented by three species: *Trichechus manatus*, the West Indian manatee, and *T. senegalensis*, the West African manatee, both of which inhabit salt water as well as fresh water; and *T. inunguis*, the Amazonian manatee, which lives only in fresh water. Unlike dugongs, which have dolphinlike flukes, manatees have rounded, paddle-shaped tail fins, and only six cervical vertebrae instead of the normal mammalian complement of seven. The vestigial pelvis is even more reduced than in dugongids. *See* MAMMALIA.

Daryl P. Domning

Bibliography. D. P. Domning, Bibliography and Index of the Sirenia and Desmostylia, *Smithson. Contrib. Paleobiol.*, 80:1–611, 1996; H. Marsh, Our tropical siren, *Austral. Geog.*, 21:42–57, 1991; R. M. Nowak (ed.), *Walker's Mammals of the World*, 6th ed., Johns Hopkins University Press, 1999; J. E. Reynolds III and D. K. Odell, *Manatees and Dugongs*, Facts on File, 1991; W. Zeiller, *Introducing the Manatee*, University Press of Florida, 1992.

Sirius

The star α Canis Majoris, also referred to as the Dog Star, the brightest of all the stars in the night sky (apparent magnitude -1.47). Sirius owes its apparent brightness both to its close distance to Earth, only 2.64 parsecs (8.14×10^{13} km or 5.06×10^{13} mi), and to its intrinsic luminosity, which is more than

20 times that of the Sun. It is a main-sequence star of spectral type A1 with an effective temperature of about 9400 K (16,500°F). *See* SPECTRAL TYPE.

Sirius is a very interesting binary system in which the companion is a degenerate star (α CMa B). It is the brightest among the known white dwarfs (apparent magnitude 8.4) and the closest to the Sun. It orbits around the bright star once every 50 years at a mean angular separation of 7.5″, corresponding to about 20 astronomical units. The masses of α CMa A and B as computed from the orbit are 2.14 and 1.03 times the solar mass, respectively. The luminosity of the white dwarf is 400 times smaller than that of the Sun, which makes it extremely difficult to see because of the overpowering brightness of the nearby primary. Its temperature is estimated to be 24,790 K (44,654°F), much hotter than the primary. The spectrum shows hydrogen lines broadened by extreme pressure, a direct consequence of the fact that the white dwarf's mass is comparable to the Sun, yet is concentrated in a body the size of the Earth. The mean density of the material in α CMa B is about 2×10^6 times that of water, or 2 metric tons/cm^3 (nearly 40 tons/in.^{3}). *See* BINARY STAR; WHITE DWARF STAR.

Sirius has been found to have chemical anomalies. Certain elements such as barium, lanthanum, and cerium are overabundant by a factor of about 100 compared to the Sun. These anomalies may be related to the fact that Sirius is in a binary system with a white-dwarf companion. Other stars with similar peculiarities in binary systems with a white dwarf are believed to have been polluted by material shed by the companion, which was originally the more massive star in the system and evolved more rapidly. *See* STAR; STELLAR EVOLUTION. David W. Latham

Sirocco

A southerly or southeasterly wind current from the Sahara or from the deserts of Saudia Arabia which occurs in advance of cyclones moving eastward through the Mediterranean Sea. The sirocco is most pronounced in the spring, when the deserts are hot and the Mediterranean cyclones are vigorous. It is observed along the southern and eastern coasts of the Mediterranean Sea from Morocco to Syria as a hot, dry wind capable of carrying sand and dust great distances from the desert source. The sirocco is cooled and moistened in crossing the Mediterranean and produces an oppressive, muggy atmosphere when it extends to the southern coast of Europe. Rain that falls in this air is often discolored by the dust or sand which is precipitated along with the waterdrops. Under sunny conditions, the sirocco can produce temperatures in excess of 100°F (38°C) in southern Europe. Various other names are used to denote the sirocco in specific localities, such as khamsin in Egypt. *See* AIR MASS; WIND.

Frederick Sanders

Sisal

A fiber obtained from the leaves of *Agave sisalana*, produced in Brazil, Haiti, and several African countries, including Tanzania, Kenya, Angola, and Mozambique.

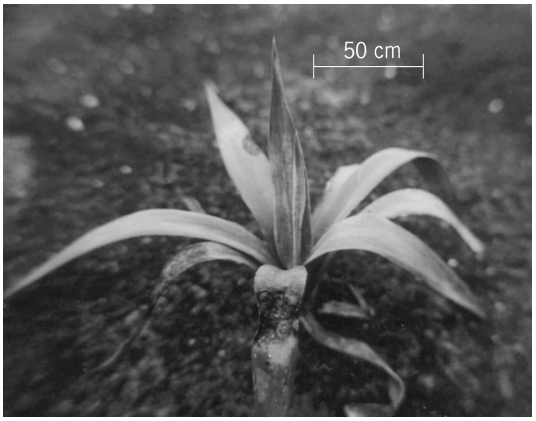
A well-developed sisal plant has a stem or trunk about 12–16 in. (30–40 cm) in diameter and 3 ft (1 m) in height at maturity. The leaves, 30–60 in. (75–150 cm) in length and 4–6 in. (10—15 cm) wide, are bright to dark green, thick, succulent, and smooth, with hardened margins that are smooth with occasional small, curved spines and a sharp terminal spike 0.4–1.2 in. (1–3 cm) long. When the plant is 6–10 years old, a flower stalk grows up from the top to a height of 10–23 ft (3–7 m). The stalk branches and produces flowers and then bulbils; seed is seldom produced.

Cultivation. Sisal grows best in the tropics where rainfall is low to moderate and the soil is calcareous with a texture that allows aeration of the roots. However, the plant does well on a fairly wide range of nonacid soils.

Land for planting should be plowed and cultivated to kill weeds. Sisal plants are propagated by bulbils or by suckers which grow from the roots of established plants. Bulbils, which tend to produce better plants, are planted in a nursery for 1 to 2 years; suckers usually are moved directly to the field location. A widely used planting pattern is a series of double rows about 3 ft (1 m) apart with a 11.5–13 ft ($3^1/_2$–4 m) avenue between each pair of rows. Plants are spaced about 3 ft (1 m) apart in the row.

Harvesting and processing. The first harvest is ready in about 3 years after field planting from the nursery stage. About 20–25 leaves having an angle of 40° or more from vertical are cut at intervals of 6–12 months, depending on growing conditions.

The fibers are found in the fleshy tissue of the leaf, mostly in association with the vascular bundles. The freshly cut leaves are brought to a central location for the defibering operation, frequently called decortication, which involves beating and scraping. The basic defibering units are wheels or drums 3–5 ft (100–150 cm) in diameter and about 12 in. (30 cm) wide. Blunt beater blades 2 in. (5 cm) high and spaced 10 in. (25 cm) apart, are attached to the face of the drum at right angles to the direction of rotation. An adjustable breastplate one-third the circumference of the drum is set so that the beater blades clear it by about the thickness of a fiber at its narrowest part, the trailing edge. The space is wider where the leaf enters at the leading edge. As the leaf passes between the breastplate and the drum, the pulp is beaten out and the fibers are scraped clean. Jets of water, directed onto the fiber as it passes through each defibering unit, wash the fiber and carry away the waste. A carrying and gripping device moves a layer of leaves sideways into one end of the machine through the first defibering unit, where the lower half of the leaves are cleaned. Then the grip is transferred to the



Leaf spot of sisal caused by *Colletotrichum agaves*.

cleaned fiber, and the layer of partly cleaned leaves continues through the second defibering unit, where cleaning is completed. The fiber is now ready to be dried on poles in the sun or in an apron conveyor dryer.

A smaller, lighter defibering machine with the same action is also used. The top of the leaf is handheld while the base end is inserted against the revolving drum to about the center of the leaf and withdrawn. The clean fiber is then grasped, and the other end is cleaned. This machine cleans fiber with less loss, but production per unit of labor is very low.

Uses. Sisal is used mainly for twine and rope, but some of the lower grades are used for upholstery padding and paper. The greatest quantity goes into farm twines, followed by industrial tying twine and rope. Most sisal-fiber ropes made in the United States are small to medium in size, intended for light duty. Sisal is sometimes used for marine cordage in Europe. *See* NATURAL FIBER.

Elton G. Nelson

Diseases. Sisal is subject to fewer diseases than most crops, with many disease symptoms being caused by nutrient deficiencies. Fungus diseases includes two bolt (stem) rots. *Aspergillus niger* has been associated with one which is a soft wet rot. The other, a dry rot, has been attributed to *Fusarium solani*. The fungus enters through the exposed base of harvested leaves and spreads inwardly.

Leaf spots are caused by the fungi *Microdiplodia agaves, Marssonina agaves,* and *Colletotrichum agaves* (see **illus.**).

Nutritional diseases include leaf banding, chlorotic leaf spot, and tip withering and mottling.

The only major insect pest of sisal is the weevil *Scyphophorus interstitialis*. It can be reduced by field sanitation and by spraying with aldrin or dieldrin.

The cause of the destructive "red rot" disease is suspected to be a virus. The fiber of the closely related plants *A. fourcroydes* and *A. cantala* is sometimes called sisal fiber; these species are affected by many of the above diseases. *See* LILIALES; PLANT PATHOLOGY; PLANT VIRUSES AND VIROIDS.

Thomas E. Summers

Skarn

A broad range of rock types made up of calc-silicate minerals such as garnet, regardless of their association with ores, that originate by replacement of precursor rocks. It was a term originally coined by miners in reference to rock consisting of coarse-grained, calc-silicate minerals associated with iron ores in central Sweden. Ore deposits that contain skarn an gangue are termed skarn deposits; such deposits are the world's premier sources of tungsten. They are also important sources of copper, iron, molybdenum, zinc, and other metals. Skarns also serve as sources of industrial minerals such as graphite, asbestos, and magnesite. *See* ORE AND MINERAL DEPOSITS; SILICATE MINERALS.

Characteristic features and origin. Skarns are composed largely of calc-silicate minerals (for example, garnet, pyroxenoid, pyroxene, olivine, and amphibole). They are relatively coarse-grained (for example, the garnets up to 1 cm (0.4 in.) in diameter and the pyroxenes up to 10 cm (4 in.) in length). Skarns are characterized by few minerals for the number of chemical components; and they are commonly zoned (consisting of zones of different minerals arranged systematically around the locus of replacement). Skarns form in geological environments characterized by igneous intrusion, metamorphism, and circulation of high-temperature (800–400°C or 1470–750°F) aqueous fluids at relatively shallow crustal levels (1–10 km or 0.6–6 mi). Skarn formation involves complex reactions between aqueous fluids and rocks (or protoliths). These reactions result in changes in the chemical composition, mineralogy, and texture of the protoliths. The most common protoliths are calcareous (for example, limestone or dolostone), although noncalcareous rocks (such as igneous rocks) also are replaced by skarn. The most common sources of skarn-forming fluids are metamorphic water released when hydrous minerals such as clays are subjected to increasing temperature; and magmatic water released during final crystallization of granitic magmas. Some magmatic waters are enriched in metals and volatile components, and these may form skarn deposits where they encounter calcareous rocks. *See* METAMORPHISM.

Classification. Based on mineralogy, three idealized types of skarn are recognized: calcic skarn characterized by calcium- and iron-rich silicates (andradite, hedenbergite, wollastonite) most commonly replaces limestone; magnesian skarn characterized by calcium- and magnesium-rich silicates (forsterite, diopside, serpentine) most commonly replaces dolostone; and aluminous skarn characterized by aluminum- and magnesium-rich calc-silicates (grossularite, vesuvianite, epidote) replaces either limestone or igneous rocks. Based on dominant process, three idealized skarn types are recognized: metamorphic skarn from isochemical recrystallization of metal-rich, calcareous sedimentary rocks; reaction skarn by local exchange of components along contacts between reactive rocks; and metasomatic skarn as the result of transfer of components into the

local environment from an external source (such as a magma). The vast majority of skarn deposits contain metasomatic skarn. *See* DOLOMITE; LIMESTONE; METAMORPHIC ROCKS.

Classification of deposits. Skarn deposits, are generally classified on the basis of dominant economic metal association. Because different metal associations reflect different sources of hydrothermal fluids or different local geological and geochemical environments, the metal classification of skarn deposits also results in a separation of deposit types on the basis of geological and mineralogical features. The most abundant of the skarn deposit types include iron deposits (Sarbai in Siberia, Russia; Eagle Mountain in California); tungsten deposits (MacMillan Pass in Yukon, Canada; King Island in Tasmania, Australia); copper deposits (Bingham, Utah; Ok Tedi, Papua New Guinea); and zinc-lead deposits (Yeonhwa-Ulchin, South Korea; EI Mochito, Honduras). *See* COPPER; IRON; LEAD; TUNGSTEN; ZINC.

M. T. Einaudi

Bibliography. M. T. Einaudi, L. D. Meinert, and R. J. Newberry, Skarn deposits, *Economy Geology 75th Anniversary Volume*, pp. 317–391, 1981; V. A. Zharikov, Skarns, *Int. Geol. Rev.*, 12:541–559, 619–647, 706–775, 1970.

Skeletal system

The supporting tissues of animals which often serve to protect the body, or parts of it, and play an important role in the animal's physiology.

Skeletons can be divided into two main types based on the relative position of the skeletal tissues. When these tissues are located external to the soft parts, the animal is said to have an exoskeleton. If they occur deep within the body, they form an endoskeleton. All vertebrate animals possess an endoskeleton, but most also have components that are exoskeletal in origin. Invertebrate skeletons, however, show far more variation in position, morphology, and materials used to construct them.

Exoskeletons. Many of the invertebrate phyla contain species that have a hard exoskeleton, for example, corals (Cnidaria); limpets, snails, and *Nautilus* (Mollusca); and scorpions, crabs, insects, and millipedes (Arthropoda). However, these exoskeletons have different physical properties and morphologies. The form that each skeletal system takes presumably represents the optimal configuration for survival. *See* ARTHROPODA; MOLLUSCA; ZOOPLANKTON.

Calcium carbonate is the commonly found inorganic material in invertebrate hard exoskeletons. The stony corals have exoskeletons made entirely of calcium carbonate, which protect the polyps from the effects of the physical environment and the attention of most predators. Calcium carbonate also provides a substrate for attachment, allowing the coral colony to grow. However, it is unusual to find calcium carbonate as the sole component of the skeleton. It normally occurs in conjunction with organic material, in the form of tanned proteins, as in the hard shell material characteristic of many mollusks. Cuticular exoskeletons are widely distributed among the invertebrates. They are formed by proteins that have been stiffened by the chemical action of phenols (the process is called tanning). Mineral salts may be incorporated within the cuticle for additional strength and stiffness.

The major advantage of a hard exoskeleton is the high degree of protection afforded to the body organs against mechanical damage and desiccation. Some disadvantages are the relatively large mass of the skeletal structure in proportion to the size of the soft tissues; the inability of some animals to remodel the skeleton and repair damage; and problems for a growing animal bounded by rigid skeletal tissues. Arthropods have solved this last problem by periodically shedding the old skeleton (molting) and replacing it with a new, larger one. Immediately after molting, the new skeleton is soft and compliant, allowing it to stretch to accommodate the increased size of the animal. During this period the animal is more vulnerable to predation. *See* MOLTING (ARTHROPODA).

Endoskeletons. Internal hard skeletons are less common among the invertebrates but are a feature of some mollusks and echinoderms. There are various degrees of skeletal reduction seen in these groups, as evidenced by the urchins which have a complete rigid skeletal test formed of calcareous ossicles sutured together, and by sea cucumbers in which the skeleton is reduced to microscopic ossicles. The vertebrate endoskeleton is usually constructed of bone and cartilage; only certain fishes have skeletons that lack bone. In addition to an endoskeleton, many species possess distinct exoskeletal structures made of bone or horny materials. This dermal skeleton provides support and protection at the body surface. *See* ECHINODERMATA.

Human Skeleton

Various structural components make up the human skeleton, including collagen, three different types of cartilage, and a variety of bone types.

Collagen. Collagen is the most abundant protein in the body. It is of fundamental importance in all organ systems and is found in all parts of the musculoskeletal system. Tropocollagen, a structure composed of three polypeptide chains, forms the basic building unit of collagen. The rodlike tropocollagen molecules polymerize into large collagen fibrils. Cross-links between adjacent molecules and also between fibrils give collagen its tensile strength. In tendons and ligaments, these fibrils are bundled together to form larger fibers. Tendons are formed from parallel bundles of fibers, an arrangement which allows tendons to support high uniaxial tensile loads to which they are subjected during activity. Although most ligament fibers are parallel, some have an oblique orientation. This reflects the more complex loading patterns applied to ligaments during movement, where small loads may be applied in a variety of directions. Collagen is stiff and strong only in tension. *See* COLLAGEN.

Cartilage. The formation of cartilage (chondrogenesis) from mesenchyme occurs in many areas of the embryo, such as the skull, limbs, and vertebral column. Most embryonic cartilage is replaced by true bone in endochondral bone formation, but in many regions the cartilage remains throughout life. The tissue consists of cartilage cells (chondrocytes) which manufacture, secrete, and maintain the extracellular organic matrix that surrounds them. This matrix contains a dense network of collagen fibrils within a concentrated solution of protein-polysaccharide molecules called proteoglycans. The three main types of cartilage are fibrocartilage, elastic cartilage, and hyaline cartilage.

Hyaline cartilage, the most abundant form of cartilage in the body, ossifies during development to become bone. In the adult it covers the articular surfaces of bones, supports the trachea and bronchi, forms the costal cartilages linking the first ten ribs to the sternum, and reinforces the nose. This type of cartilage contains extremely fine collagen fibrils that can be seen only with the electron microscope. Hyaline articular cartilage is a special form found within synovial joints. Its material characteristics are perfectly adapted to this mechanically demanding environment.

Fibrocartilage is found predominantly in the pubic symphysis, the menisci of the knee, and the intervertebral discs. This type of cartilage is very durable and can withstand large tensile and compressive forces. It also represents a transitional material found at tendon and ligament insertions into bone and at the margins of some joint cavities. Histologically it can be characterized by large collagen fibers running through the matrix.

Elastic cartilage is similar to hyaline cartilage but contains abundant elastin fibers which give it a yellowish appearance and make it very flexible, while still maintaining its strength. It is found in the outer ear, the auditory tube, and the larynx. *See* CONNECTIVE TISSUE.

Bone. This specialized rigid connective tissue comprises cells which produce an organic extracellular matrix of collagen fibers and ground substance, and inorganic materials in the form of mineral salts. In vertebrates, the mineral portion of bone consists primarily of calcium and phosphate in the form of hydroxyapatite crystals. The organic phase gives bone its resilience and flexibility, while the inorganic phase makes the bone hard and rigid.

The functions of bone are numerous, relating to the maintenance of mineral (mainly calcium) homeostasis, formation of blood cells, and mechanical requirements.

Mineral (mainly calcium) homeostasis. Plasma calcium ion (Ca^{2+}) concentrations are maintained at about 5 mEq/L. This concentration is effectively regulated by calcitonin and parathyroid hormone and is required for normal blood clotting and nerve and muscle function. Low concentrations of calcium in the extracellular fluid that bathes the parathyroid glands elicits parathyroid hormone release. This stimulates osteoclastic breakdown of bone, releasing calcium and phosphorus into the extracellular fluid. Simultaneously, calcium absorption from the gut is increased, calcium loss via the kidneys is decreased, and urinary phosphate excretion is elevated. Calcitonin, released from the thyroid gland in response to elevated calcium concentrations, inhibits calcium removal from bone and increases urinary calcium excretion.

Formation of blood cells. During embyonic development, cellular elements of the blood (red and white blood cells and platelets) are produced by a process known as hematopoiesis. Early production occurs in the vessels of the yolk sac, but following the development of other organ systems, major blood cell production occurs in the liver, spleen, thymus, and bone marrow. At about the sixth developmental month, red bone marrow becomes the major site for the production of red and white blood cells and platelets. In adults, bone is the primary site for white cell production and the only site for red cell production, with portions of the sternum, ribs, vertebrae, skull, pelvis, scapulae, and proximal femoral and humeral heads of particular importance.

Mechanical requirements (support, protection, and leverage). Bone has a protective function, particularly with regard to the central nervous system. The skeleton is the supporting framework for the body and provides stiff levers on which muscles act to generate movement. *See* CALCIUM METABOLISM; HEMATOPOIESIS; HOMEOSTASIS.

There are a number of easily recognized bone types whose microstructure can be related to the mechanical functions required of the bone. The only type of bone that does not require a preexisting surface or structure for its formation is woven bone. Woven bone is formed during early stages of development, and in adults it is encountered only in pathological conditions, including bone fracture repair. It consists of randomly oriented, small-diameter, highly mineralized collagen fibers. It is not a strong form of bone because of its unordered structure and, particularly, because much of its volume is unmineralized. Its value lies in its ability to rapidly bridge the gap between broken ends of bone to act as temporary scaffolding during the process of repair. It is later removed by bone remodeling events.

Lamellar bone is the dense hard material that constitutes most of the skeleton. In a typical bone, such as the femur, lamellar bone has two distinct types of organization. In the shaft (diaphysis) the material is deposited in layers to form compact or cortical bone. A transverse section through the diaphysis would show a tree-ring-like arrangement of bone. Bone cells (osteocytes) occupy spaces between adjacent lamellae. They interconnect with vascular spaces and one another by fine cellular extensions running in narrow channels (canaliculi). These cells maintain the integrity and normal functioning of bone. Two other main cell types important in normal bone are the bone-forming cells (osteoblasts), and the bone-destroying or -resorbing cells (osteoclasts).

Trabecular (spongy) bone is another form of lamellar bone found particularly in the expanded ends (epiphyses) of long bones and in the vertebrae.

Although it is made of the same material as cortical bone, the mechanical characteristics of trabecular bone differ as a result of the honeycomb arrangement of bone into interlacing struts (trabeculae). This spongy bone is less dense and stiff than cortical bone, and has a large surface area which makes this tissue important for the exchange of calcium between the skeleton and the bloodstream.

Plexiform or fibrolamellar bone is a medium-density, relatively strong bone. It is found at sites of rapid growth where strength is also needed, such as in the limbs of large, fast-growing animals, which include humans during the growth phase in puberty.

Haversian systems or secondary osteons are structural units of bone that are formed secondarily within preexisting bone. This intracortical remodeling involves the formation of a tunnel through bone, achieved by the action of osteoclasts generating a resorption cavity. Following a reversal phase, in which a cement lining is deposited around the perimeter of the cavity, the tunnel is refilled with concentric rings of new bone by osteoblastic activity. A small-diameter central canal containing blood vessels is left unfilled. A completed secondary osteon is a branched structure. Secondary osteons form throughout life and are implicated in skeletal adaptation, repair processes, and mineral exchange between blood and bone (**Fig. 1**). *See* BONE.

Development of the skeleton. All the components of the skeleton are derived from mesenchyme of either mesodermal or neural crest ectodermal origin. The majority of bones in the human body develop by the process of endochondral ossification; that is, the mesenchyme first forms a cartilaginous model which is subsequently replaced by true bone. The facial bones and certain bones of the cranium are formed by intramembranous ossification, in which mesenchyme is converted into skeletal elements by forming bone directly, without need of a cartilage stage. Sesamoid bones are specialized intramembranous bones that form within tendons.

Ossification. The development of bone from the embryo to the adult depends on the orderly processes of mitotic division, growth, and remodeling. These are determined largely by genetics, but are strongly influenced by hormonal action and nutrition. In endochondral ossification, the cartilaginous model is gradually calcified, resulting in cartilage cell (chondrocyte) death. Osteoblasts, together with a blood supply, invade the model and begin to secrete osteoid, which subsequently mineralizes and forms a primary ossification site. This process is repeated in the epiphyses of long bones to form secondary ossification sites. It is normal for the primary ossification site to form earlier, grow more rapidly, and cease ossification later than the secondary sites. The locations and ages at which ossification sites are active in the developing human skeleton are well documented. Hand and wrist x-radiographs are used to determine the skeletal age of individuals because of the orderly progression of ossification in these regions. Such information is important for checking whether an individual's growth rate is abnormal and whether hormone treatment is indicated. Applications also exist in the forensic field, where age at death can be accurately predicted from skeletal remains.

Two parts of developing bones remain as cartilage: epiphyseal plates and articular cartilage. Epiphyseal plates are situated between the diaphysis and the epiphysis. Longitudinal bone growth occurs by chondrocyte proliferation on the diaphyseal sides. The plates are finally obliterated by an extension of diaphyseal ossification into these regions, thus preventing further growth in the length of the bone. Articular cartilage never ossifies except in pathological situations such as osteoarthritis. *See* ARTHRITIS.

Changes in the radial dimensions of long bones occur by new material being deposited beneath the periosteum, a connective tissue membrane surrounding the bone. The endosteal membrane is the equivalent structure lining the internal surfaces of tubular bones. Bone tissue can be added or removed

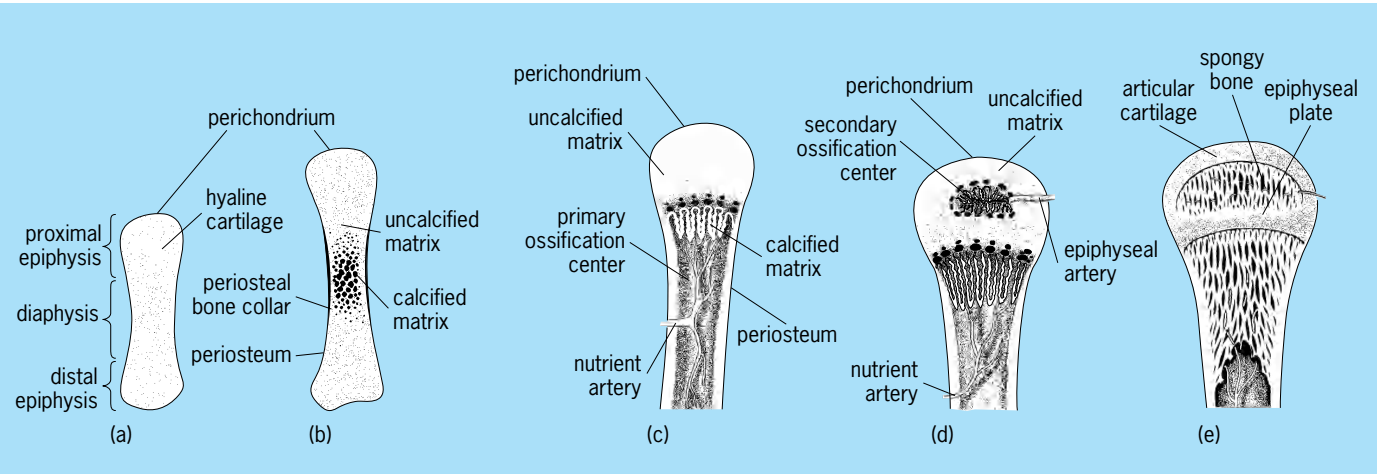


Fig. 1. Development of endochondral ossification in a long bone. (*a*) Mesenchymal cells differentiate into chondroblasts which form the hyaline cartilage model. (*b*) The cartilage model grows, and chondrocytes in the midregion calcify the matrix. (*c*) The primary ossification center and the medullary cavity form. (*d*) Postnatal development of the secondary ossification center occurs. (*e*) Remnants of hyaline cartilage as articular cartilage and epiphyseal plate persist. (*After G. J. Tortora, Principles of Human Anatomy, 6th ed., Harper Collins, 1992*)

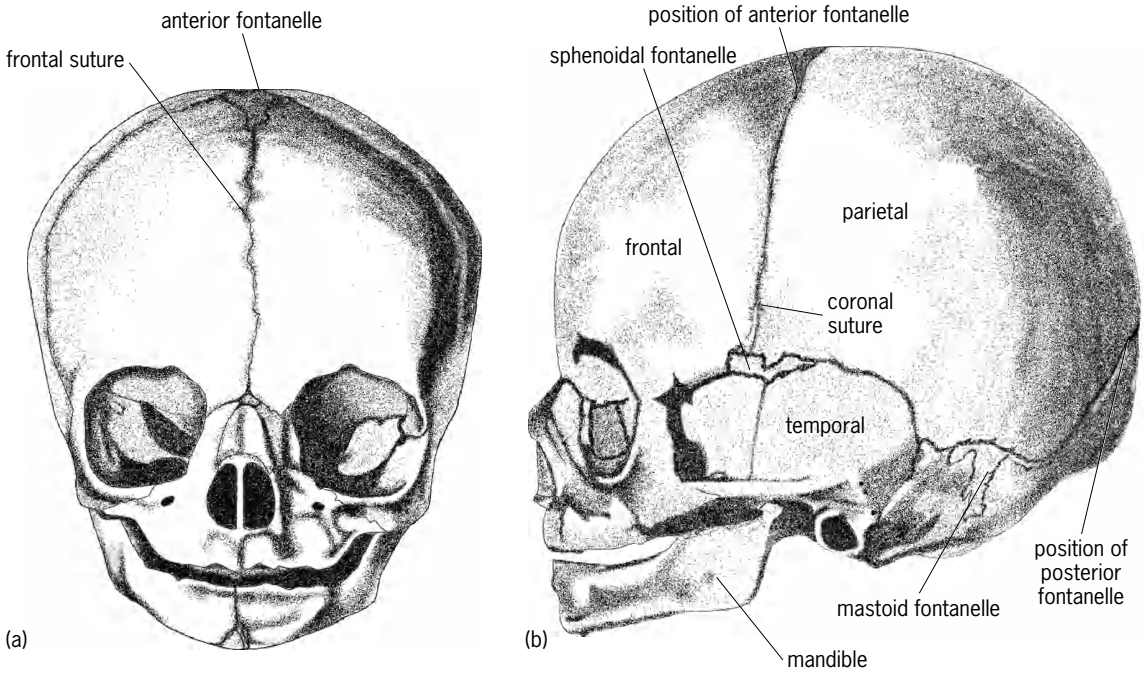


Fig. 2. Skull in a human newborn infant. Cranial sutures and areas of temporary cartilage (fontanelles) form articulations between bones. (*a*) Frontal view. (*b*) Lateral view. (*After G. J. Romanes, Cunningham's Textbook of Anatomy, 10th ed., Oxford University Press, 1964*)

at either site by the action of osteoblasts or osteoclasts, respectively. During growth and aging, there is a tendency for net bone deposition periosteally, with net resorption endosteally.

Bone growth and fracture repair can be increased by mechanical and electrical stimulation. Mechanical deformation of bone causes fluid flow through microscopic channels in bone, resulting in fluid shear stresses on cell membranes and/or production of strain-generated electrical potentials. Both effects have been shown to have osteogenic effects, with osteocytes responding to fluid flow changes, and osteoblast-mediated bone deposition occurring subsequent to mechanical loading in which strain-generated electrical potentials are recorded. In addition, electrical stimulation alone can promote osteogenesis under a wide range of conditions, including nonunion bone fractures and the maintenance of bone mass even in the absence of mechanical function. The mechanisms appear to involve insulinlike growth factors and/or transforming growth factor-β. Elevated levels of both accompany electrical stimulation, and insulinlike growth factors are capable of stimulating proliferation and differentiation of osteoprogenitor cells. Transforming growth factor-β is known to have important effects on bone formation, osteoblast proliferation and differentiation, and matrix synthesis. Furthermore, in vertebrates there are transforming growth factor-β related signaling proteins (bone morphogenetic proteins and activin) that effectively regulate animal development, can induce cartilage and bone formation, and play critical roles in modulating mesenchymal differentiation.

Joints. In the human body there are fibrous, cartilaginous, and synovial joints. Functionally, these joints can be considered as immovable (synarthrosis), partly movable (amphiarthrosis), and freely movable (diarthrosis) joints, respectively.

Immovable joints are represented by cranial sutures and epiphyseal plates prior to their ossification. Little or no movement is available at these joints. Instead, their primary function is to allow bone growth at their margin (**Fig. 2**).

Partly movable joints include the pubic symphysis that joins the two halves of the pelvis anteriorly and the fibrocartilaginous intervertebral discs. While the amount of movement between adjacent vertebrae is quite limited by disc stiffness, the vertebral column as a whole is quite flexible. Flexibility is achieved by combining each of the small movements permitted by individual discs.

Freely movable joints are the most complex and varied of the three types of joint, with their sizes and shapes matched to the functional requirements of the skeletal system at each location. At each joint, the surfaces of opposing bones are covered with a layer of articular cartilage that is a few millimeters thick. The joint is enclosed within a flexible joint capsule, the internal surface of which is lined by the synovial membrane that secretes the lubricating synovial fluid into the joint space. The main functions of articular cartilage are to distribute compressive loads over a wide area in order to reduce contact stresses, and to allow relative movement of opposing joint surfaces with minimum wear and friction. The combination of articular cartilage and synovial fluid gives these joints these remarkable properties. *See* JOINT (ANATOMY).

Articular cartilage. Collagen fibrils, enmeshed within a proteoglycan solution, form the superficial part of articular cartilage. Beneath the relatively soft

articular cartilage lies an interface (the tidemark) that indicates the start of the calcified cartilage layer. Beneath the cartilage lies a layer of subchondral bone which itself lies over the trabecular bone of the epiphysis. A gradient of increasing stiffness exists from the articular surface to the trabecular bone that helps protect the cartilage from splitting under loading conditions.

Synovial joints are subjected to an enormous range of forces. For example, the joints in the lower limb of humans have to support transient forces between two to five times body weight during running and jumping, and moderate but prolonged loading during standing. The nearly frictionless operation of joints is thought to be a function of lubricating films of synovial fluid between, and an adsorbed boundary lubricant on, the articular cartilage surfaces. The ability of cartilage to resist deformation when subjected to high stresses resides in complicated interactions between collagen fibrils, water, and proteoglycans, particularly keratin sulfate, chondroitin sulfate, and hyaluronic acid.

Damage to cartilage is thought to be a function of both the magnitude of applied forces and the speed at which forces are applied (with impact loading causing more damage). Due to its avascular nature, articular cartilage has a limited capacity for repair of such damage.

Vertebrate Skeletal Anatomy

The vertebrate skeleton consists of the axial skeleton (skull, vertebral column, and associated structures) and the appendicular skeleton (limbs or appendages). The basic plan for vertebrates is similar, although large variations occur in relation to functional demands placed on the skeleton.

Axial skeleton. The axial skeleton supports and protects the organs of the head, neck, and torso, and in humans it comprises the skull, ear ossicles, hyoid bone, vertebral column, and rib cage.

Skull. The adult human skull consists of eight bones which form the cranium, or braincase, and 13 facial bones that support the eyes, nose, and jaws. There are also three small, paired ear ossicles—the malleus, incus, and stapes—within a cavity in the temporal bone. The total of 27 bones represents a large reduction in skull elements during the course of vertebrate evolution. The three components of the skull are the neurocranium, dermatocranium, and visceral cranium. *See* EAR (VERTEBRATE).

The brain and certain sense organs are protected by the neurocranium. All vertebrate neurocrania develop similarly, starting as ethmoid and basal cartilages beneath the brain, and as capsules partially enclosing the tissues that eventually form the olfactory, otic, and optic sense organs. The basal and ethmoid plates expand to meet the olfactory and otic capsules to form a floor on which the brain rests. The optic capsule becomes the fibrous sclerotic coat of the eyeball and remains free to move independently of the skull. Further development produces cartilaginous walls around the brain. Passages (foramina) through the cartilages are left open for cranial nerves and blood vessels. The largest opening is the foramen magnum, through which the spinal cord passes. Endochondral ossification from four major centers follows in all vertebrates, except the cartilaginous fishes.

It is likely that the bones of the dermatocranium derive from the dermal armor (bony scales) of early fishes. In modern vertebrates, these bones form via intramembranous ossification from subdermal mesenchyme.

The basic structure from which the tetrapod dermatocranium probably evolved is seen in the primitive crossopterygian fishes and early amphibians. Four kinds of bones are involved. (1) Roofing bones form above and on each side of the brain and neurocranium. These comprise paired nasal, frontal, parietal, and postparietal bones that cover the dorsal parasagittal area, while prefrontal, postfrontal, postorbital, infraorbital, and lacrimal bones form a ring around the orbit. The posterolateral part of the skull is formed from the paired intertemporal, tabular, supratemporal, squamosal, and quadratojugal bones. (2) Dermal, tooth-bearing bones (premaxillae and maxillae) form the margins of the upper jaw. (3) Bones of the primary palate form the roof of the mouth in lower tetrapods and the oropharynx of fishes, and are the paired vomers, palatines, pterygoids, and ectopterygoids and the unpaired parasphenoid. (4) Opercular bones extend posteriorly from the hyoid arch to cover the gill slits. These are paired bones, represented by large opercular bones and smaller subopercular, preopercular, and interopercular bones (**Fig. 3**). *See* TETRAPODA.

Visceral skeleton. This skeleton of the pharyngeal arches is demonstrated in a general form by the elasmobranch fishes, where all the elements are cartilaginous and support the jaws and the gills. Each pharyngeal arch is typically composed of a number of cartilage elements, most of which support gills, but the first and second arches are modified to function in feeding. The mandibular (first) arch consists of two elements on each side of the body: the palatoquadrates dorsally, which form the upper jaw, and Meckel's cartilages, which join ventrally to form the lower jaw. The hyoid (second) arch has paired dorsal hyomandibular cartilages and lateral, gill-bearing ceratohyals. This jaw mechanism attaches to the neurocranium for support. Variations in this articulation occur between species and effectively determine jaw movement, and hence, the feeding abilities of fishes.

Evolution of neurocranium and dermatocranium. Modern amphibian skulls differ considerably from the primitive forms. Most changes involve the loss of roofing bones, particularly those from around the orbit, and massive reduction of the primary palate. These changes allow the eyes, which normally protrude from the head for good vision, to retract until they bulge into the oral cavity.

Early reptiles maintained the primitive pattern of the skull roof, with loss of only the intertemporal bone. Later reptiles exhibited large losses of bony elements, mostly from the temporal region and the back margin of the skull. Major changes in the roof

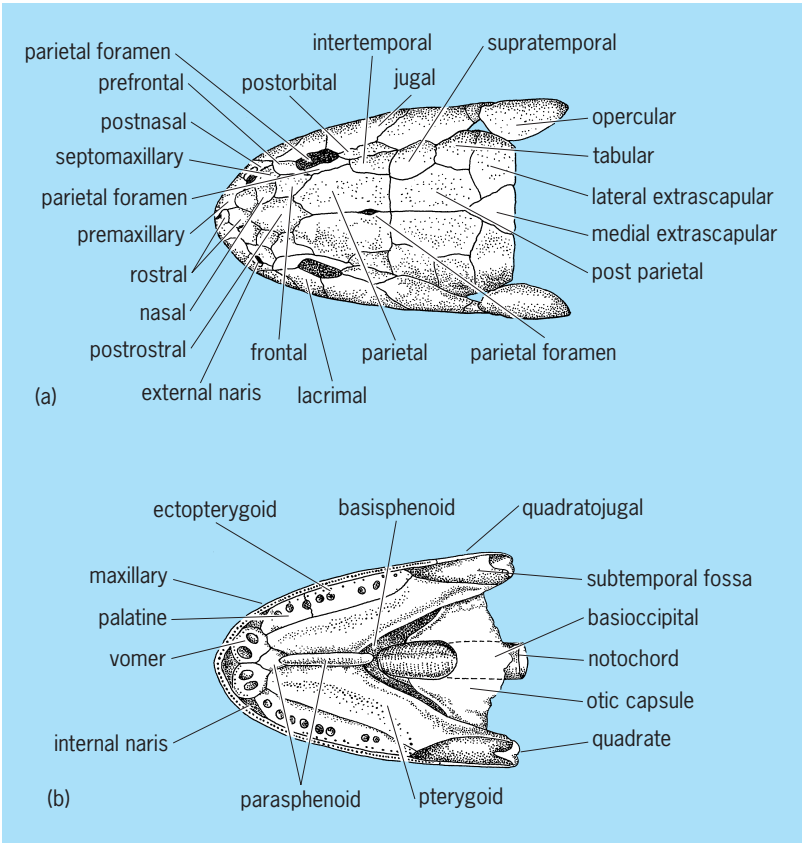


Fig. 3. Skull of *Eusthenopteron*, a generalized fossil crossopterygian fish from Devonian time. (*a*) Dorsal aspect. (*b*) Ventral aspect. (*After A. S. Romer, The Vertebrate Body, 3d ed., Saunders, 1962*)

of the skull are associated with the development of temporal fenestrae, resulting in four recognized design types: anaspids exhibit a primitive condition with a solid skull roof (turtles and stem reptiles); diaspids have two openings on each side of the skull (archosaurs and the ancestors of lizards and snakes); euryapsids have a single, dorsal opening on each side (ichthyosaurs and plesiosaurs); and synapsids have a single, lateral opening (pelycosaurs and therapsids). Birds derive from archosaurian reptiles and have, essentially, a diaspid skull design. The pelycosaur stock, with a synapsid skull, gave rise to the mammals.

A secondary palate appears for the first time in the reptiles as a horizontal shelf of bone dividing the oral cavity into separate nasal and oral passages. The remaining bones of the more primitive primary palate are located in the nasal passageway. Only the crocodilians, among the reptiles, have a complete secondary palate formed from medial extensions of the maxillae, premaxillae, pterygoid, and palatine bones. *See* REPTILIA.

The avian skull has been modified from the diaspid reptilian type in line with increased brain size and changes in life-style, particularly feeding and flight behavior. Premaxillae and dentary bones form the majority of the upper and lower beak, respectively. Some birds can raise the upper portion of the beak relative to the rest of the skull by a hinge mechanism between the frontal bone and the premaxillae, maxillae, and nasal bones. This is one of the many forms of cranial kinesis, where motion (usually of the upper jaw) is possible, partially independent of other parts of the skull. Most vertebrates other than mammals are capable of cranial kinesis, which is a feeding adaptation. *See* AVES.

The therapsids are mammallike reptiles that evolved large temporal fenestrae in the dermal bones of the cheek region. Other notable changes involved the loss, fusion, or reduction of roofing bones and a large variation in bone proportions and sizes. A secondary palate formed from the premaxillae, maxillae, and palatines separates the nasal and oral cavities in a manner that was similar to the crocodilians. This palate continues posteriorly as a fold of skin, the soft palate. The value of a divided airway is that it gives the animal the ability to breathe while the mouth is full. *See* THERAPSIDA.

Notable increases in the size of the brain during mammalian evolution have resulted in considerable changes to the neurocranium. The zygomatic arches are the true remnants of the original lateral walls of the skull. The temporal bone, as seen in humans, is a single unit, but is really a composite of a large number of fused skeletal elements of endochondral and membranous origin. The occipital bones are usually fused, and the single occipital condyle, present in reptiles and birds, divides to become a paired structure as in modern amphibians. Large changes occur in the organization of the otic capsule, with the inclusion of a new skull bone (the entotympanic) and dermal bone (derived from the angular bone of the reptilian jaw) in a protective role for the middle ear. *See* MAMMALIA.

Evolution of visceral cranium. In all jawed vertebrates except mammals, an articulation between the posterior ends of the palatoquadrate and Meckel's cartilages (which may be ossified or ensheathed in bone) occurs between the upper and lower jaws (**Fig. 4**). The bony fishes have elaborated on the primitive condition, where the upper jaw was fused to the skull and the lower jaw or mandible could move only in the manner of a simple hinge. Teleosts are able to protrude the upper and lower jaws. This motion, coupled with expansion of the buccal cavity, enables these fishes to generate the large suction forces used to draw food into the mouth.

In the course of mammalian evolution, the dentary of the lower jaw enlarged and a ramus expanded upward in the temporal fossa. This eventually formed an articulation with the squamosal of the skull. With the freeing of the articular bone (seen as an ossified posterior end of Meckel's cartilage in teleosts) and the quadrate from their function in jaw articulation, they became ear ossicles in conjunction with the columella, that is, a skeletal rod that formed the first ear ossicle, arising in the amphibia to conduct sound waves from the eardrum to the otic capsule.

The remaining visceral skeleton has evolved from jaw and gill structures in the fishes to become an attachment site for tongue muscles and to support the vocal cords in tetrapods.

Vertebral column, ribs, and sternum. The vertebral column is an endoskeletal segmented rod of mesodermal origin. It provides protection to the spinal

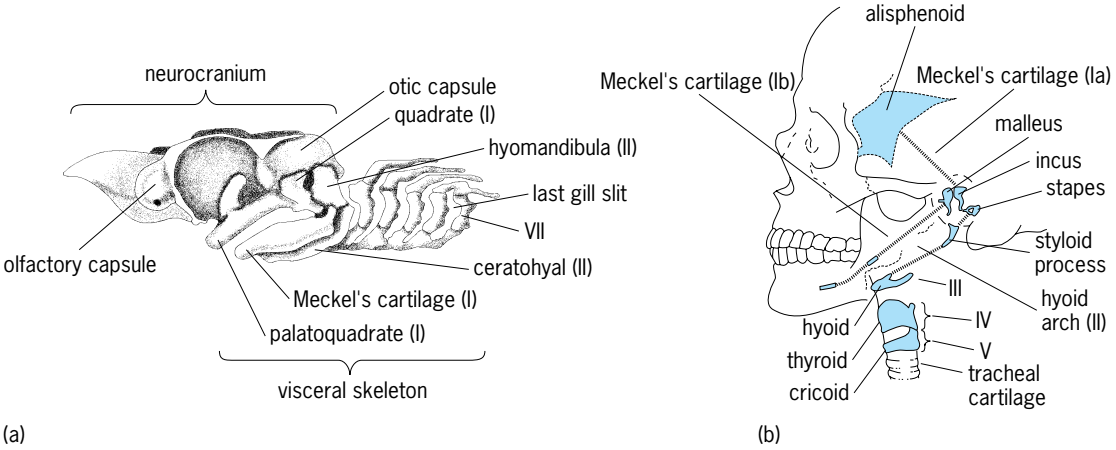


Fig. 4. Skull and visceral skeleton. (*a*) Shark. The skeleton of the first, second, and seventh pharyngeal arches is indicated. (b) Human. Shaded regions show the components of the human skull and neck that are derived from the visceral skeleton. (*After G. C. Kent, Comparative Anatomy of the Vertebrates, 7th ed., Mosby-Year Book, 1992*)

cord, sites for muscle attachment, flexibility, and support, particularly in land-based tetrapods where it has to support the weight of the body (**Fig. 5**).

1. Vertebrae. Hard, spool-shaped bony vertebrae alternate with tough but pliable intervertebral discs. Each typical vertebral body (centrum) has a bony neural arch extending dorsally. The spinal cord runs through these arches, and spinal nerves emerge through spaces. Bony processes and spines project from the vertebrae for the attachment of muscles and ligaments. Synovial articulations between adjacent vertebrae (zygapophyseal joints) effectively limit and define the range of vertebral motion.

Vertebral morphology differs along the length of the column. There are two recognized regions in fishes (trunk and caudal) and five in mammals (cervical, thoracic, lumbar, sacral, and caudal), reflecting regional specializations linked to function. Humans have seven cervical, twelve thoracic, five lumbar, five (fused) sacral, and four coccygeal vertebrae. Most amphibians, reptiles, and mammals have seven cervical vertebrae regardless of neck length (giraffes have only seven), whereas the number is variable in birds. Specific modification to the first two cervical vertebrae in most reptiles, birds, and mammals gives the head extra mobility. In humans, the occipital condyles articulate with the atlas, which in turn articulates with the axis, all via synovial joints. The atlas is a ringlike vertebra whose centrum is represented as a peg of bone fused to the axis. This odontoid process provides a pivot about which the atlas, and hence the head, swivels. The nodding movement is provided by rotational motion at the atlanto-occipital joint.

The presence of large ribs in the thoracic region often limits spinal flexibility. Birds utilize fusion of thoracic vertebrae to provide a rigid structure, adapted for the demands of powered flight. Additionally, fusion of the lumbar, sacral, and proximal caudal vertebrae produces a rigid back, the synsacrum.

In typical tetrapods, the sacral region is usually modified for support of the pelvic girdle, while the number of caudal vertebrae varies greatly (from 0 to 50) between and within animal groups. Birds have a specialized end cap to the caudal vertebrae (pygostyle) for the attachment of feathers. *See* VERTEBRA.

2. Sternum and ribs. Jawed fishes have ribs that help maintain the rigidity and support of the coelomic cavity. These ribs typically follow the connective tissue septa that divide successive muscle groups. In the caudal region, they are often small paired ventral ribs, fused on the midline to form the haemal arches. Ancestral tetrapods had ribs on all vertebrae, and their lengths varied between the vertebral regions. Modern amphibia (frogs and toads) have few thoracic ribs, and these are much reduced and never meet ventrally. Reptiles have varied rib arrangements, ranging from snakes with ribs on each vertebra (important for locomotor requirements) to turtles with only eight ribs which are fused to the inside of the carapace. Flying birds and penguins have a greatly enlarged sternum that links the ribs ventrally. This provides a large surface area for the origin of the powerful pectoralis muscles that are used to power the downstroke of the wing. In humans there are twelve pairs of ribs which form a strong but movable cage encompassing the heart and lungs.

Appendicular skeleton. This section of the skeletal system comprises the pectoral and pelvic limb girdles and bones of the free appendages. The girdles provide a supporting base onto which the usually mobile limbs attach.

Pectoral girdle. The pectoral girdle has both endoskeletal and dermal components. The dermal components are derived from postopercular dermal armor of primitive fishes, and are represented by the clavicles and interclavicles in modern vertebrates, except where they are secondarily lost. Endochondral bone forms the scapula. In fishes, the main component of the girdle (the cleithrum) is anchored to the skull by other bony elements. Increased mobility of the girdle is seen in amphibia as it becomes independent of the skull. Further development and skeletal reduction have resulted in a wide range of morphologies, culminating in the paired clavicles and scapulae of mammals.

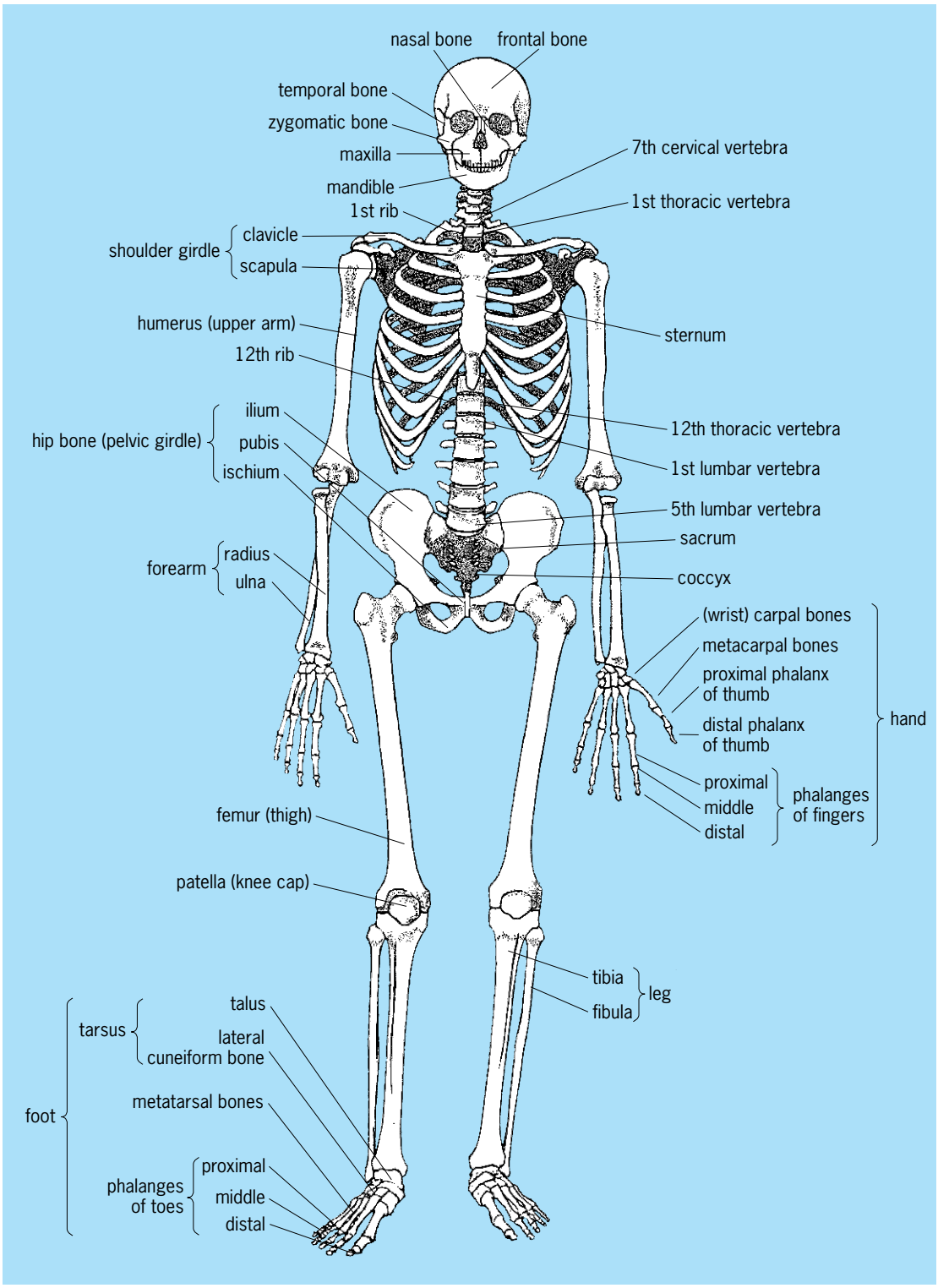


Fig. 5. Human skeleton (anterior view). (*After G. J. Romanes, ed., Cunningham's Textbook of Anatomy, 10th ed., Oxford University Press, 1964*)

Birds have fused their paired clavicles and single interclavicle to form the wishbone or furcula. This provides an increased site for the origin of the pectoral muscles, both directly and from the fascial sheet spanning the two arms. Bending of the furcula during flight is thought to assist breathing by compressing or expanding underlying air sacs. It also has a role in storing, and subsequently returning, elastic strain energy at functionally useful parts of the wing beat cycle. Clavicles have disappeared in certain groups of bounding mammals to allow greater movement of the scapula. Although humans, and most other mammals, have a coracoid process on the scapula, other tetrapods typically have a separate coracoid bracing

the scapula against the sternum and forming part of the glenoid fossa.

Pelvic girdle. The pelvic girdle forms by endochondral ossification, that is, the conversion of cartilage into bone. In the fishes, it is a small structure embedded in the body wall musculature just anterior to the cloaca. Each half of the girdle provides an anchor and articulation point for the pelvic fins. In tetrapods, the girdle attaches to the vertebral column to increase its stability and assist in the support of body weight and locomotor forces. Humans, like all other tetrapods, have a bilaterally symmetrical pelvic girdle, each half of which is formed from three fused bones: the ischium, ilium, and pubis. A part of each of these elements forms the acetabulum, the socket-shaped component of the hip joint, that articulates with the femoral head. The human pelvis is bowl-shaped, helping to support the viscera in upright stance. The large ilium (a feature shared with certain heavy-bodied animals such as cattle and horses) is the site of origin of hip extensor muscles, the gluteals. They prevent the body from bending sharply at the hip and also assist powerful thigh muscle extension during the propulsive stage of running.

All urogenital and digestive products have to pass through the pelvic outlet. This accounts for the pelvic sexual dimorphism seen in most mammals, where the pelvic opening is broader in females, because of the physical demands of pregnancy and parturition. In birds (with the exception of the ostrich and the rhea), both sexes have an open pelvic girdle, a condition also found in female megachiropteran bats (flying foxes), gophers, and mole-rats.

Paired fins and tetrapod limbs. Paired fins in fishes come in different forms, but all are involved in locomotion. In the simplest form they are fairly rigid and extend from the body, functioning as stabilizers, but they are also capable of acting like a wing to produce lift as in sharks. In many fishes, the pectoral fins have narrow bases and are highly maneuverable as steering fins for low-speed locomotion. In addition, some fishes (such as the Australian lungfish) use their pectoral and pelvic fins to walk on the river bed, while others have greatly enlarged pectoral fins that take over as the main propulsive structures (for example, rays, flatfish).

The basic mammalian pectoral limb consists of the humerus, radius, ulna, carpals, five metacarpals, and fourteen phalanges (arranged as 2-3-3-3-3, starting at the thumb; reptiles tend to have a 2-3-4-5-3 arrangement); and the pelvic limb consists of the femur, tibia, fibula, tarsal, five metatarsals, and fourteen phalanges (mammals have various digital arrangements; most reptiles have a 2-3-4-5-4 arrangement). Variation in the shape and number of elements occurs primarily in the hand (manus) and foot (pes). A typical bird pelvic limb consists of a femur, tibiotarsus (formed by fusion of the tibia with the proximal row of tarsal bones), fibula, and tarsometatarsus (formed by fusion of metatarsals II–IV), metatarsal I, and four digits (each consisting of two to five phalanges).

Locomotion and skeletal adaptations. Throughout evolution, the skeletal system has adapted to the needs of many different types of organisms. Such adaptations have been made for walking and running, speed, power, digging and burrowing, locomotion without limbs, and aerial and aquatic locomotion.

Walking and running. The undulatory side-to-side flexion of the vertebral column seen in most fishes continued to play a role in early tetrapod locomotion. It still exists in the urodeles (for example, newts) and many reptiles where limbs project laterally, with the proximal element swinging in the horizontal plane. The feet provide points of firm contact with the ground in this sprawling locomotor style, with forward progression powered by lateral trunk flexion. Most terrestrial tetrapods have raised the trunk off the ground, with the limbs projecting ventrally to support the body. Changes in joint alignment resulted in the limbs and trunk flexing and extending in the sagittal plane.

Adaptations for speed. Animals use various recognizable forms of locomotion, or gaits, for traveling at different speeds. The voluntary selection of a particular gait appears to be linked to reducing locomotor forces on the limb skeleton and minimizing the energetic cost of travel.

Obvious modifications to the basic tetrapod skeleton have accompanied the acquisition of high-speed locomotion. The lengthening of limb segments allows for longer strides which, coupled with stride frequency, determine running speed. Limb elongation has been accompanied by a reduction in the number of skeletal elements in running specialists (**Fig. 6**).

The normal pentadactyl (five-digit) limb is seen in plantigrade animals, such as primates and insectivores, where the ankles, wrists, and all bones distal to these joints contact the ground. Humans have a bipedal plantigrade posture. A plantigrade posture typically provides stability (large foot-ground contact area) and good manual or pedal dexterity (useful for climbing, and holding and manipulating objects), but does not provide particularly good running performance (**Fig. 7**).

Many of the Carnivora, Rodentia, and Lagomorpha have either lost or reduced the first digit. They are digitigrade; that is, they walk and run on four toes, with the metacarpals and tarsals raised off the ground. The functional increase in limb length contributes to their fleetness of foot, useful for catching prey or avoiding capture; and the reduction, particularly of the metacarpals, metatarsals, and digits, can be viewed as an adaptation for speed. *See* CARNIVORA; LAGOMORPHA; RODENTIA.

Hoofed (unguligrade) animals have further reduced their digits. Even-toed ungulates (sheep, cattle, camel) and odd-toed ungulates (horse, rhinoceros) walk on the tips of the remaining digits. Claws or nails are often expanded to form protective structures such as the horse's hoof. The remaining metatarsals and metacarpals are often very elongate. Digital reduction is extreme in horses, where a single digit (digit III, equivalent to the human middle finger or toe) remains. All limb reduction results in loss

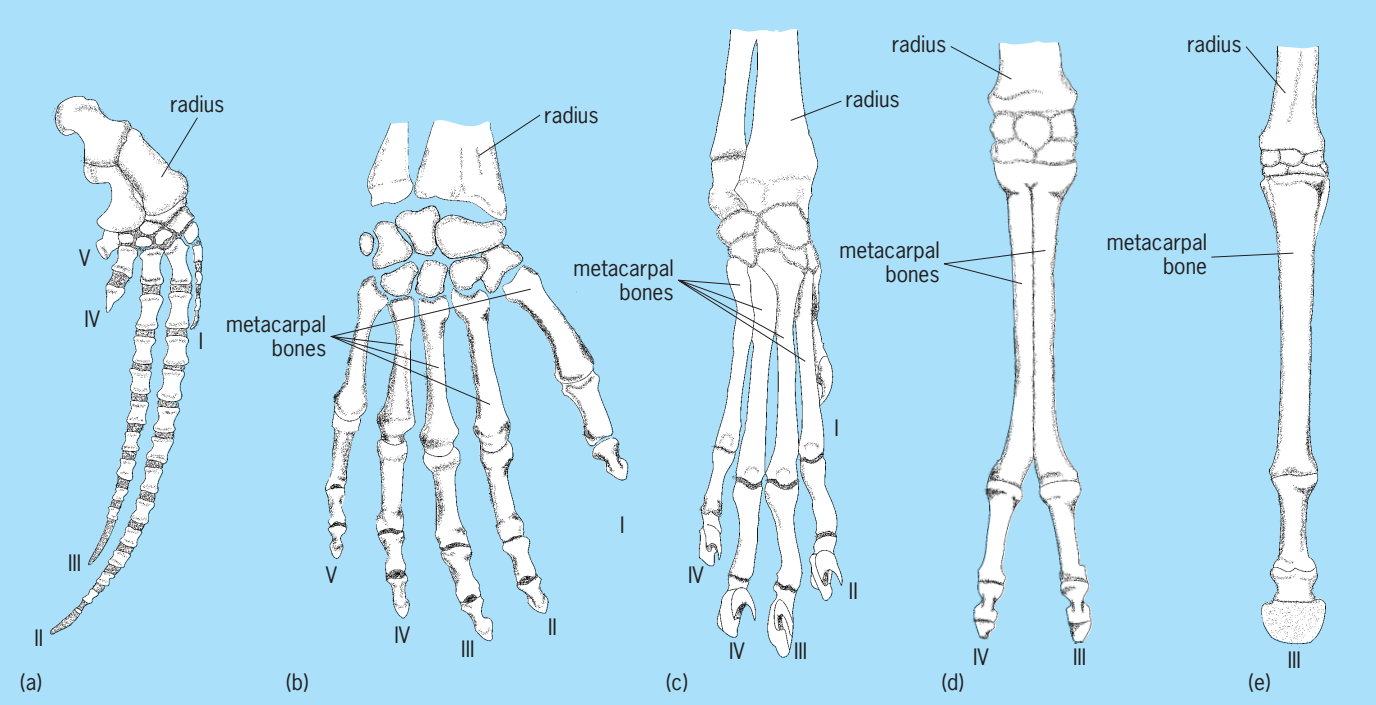


Fig. 6. Changes in the tetrapod distal limbs in relation to different locomotor requirements. (*a*) Whale (*after A. S. Romer and T. S. Parsons, The Vertebrate Body, 6th ed., CBS College Publishing, 1986*). (*b*) Human; (*c*) cat; (*d*) camel; (*e*) horse (*after G. C. Kent, Comparative Anatomy of the Vertebrates, 7th ed., Mosby-Year Book, 1992*)

of dexterity, and the horse, as an extreme example, cannot use its limbs for any function other than locomotion. However, digital reduction is extremely important for economical locomotion: multiple bones weigh more than a single element of equal strength, and thus limb mass can be reduced by reducing the number of bones without compromising limb strength.

There is a similar trend in birds. The ostrich is the only bird that has two toes (didactyl); rheas, cassowaries, emus, and others have three (tridactyl), whereas most species have four. However, variation in toe number, size, and orientation reflects more than just locomotor behavior because there are three-toed species (kingfishers, parrots) that are not runners.

Other skeletal adaptations for speed include the loss of clavicle in cats, allowing greater scapular movement which results in greater reach (greater stride length when running) and helps reduce impact forces to the axial skeleton when running; and increased flexibility of the vertebral column in the sagittal plane—another way to increase stride length.

Speed versus power. The speed of movement of limb bones is of fundamental importance for running animals. This depends on the sizes and properties of muscles and where they attach relative to the joints. A muscle that is attatched close to a joint has a small moment arm (the perpendicular distance from the line of action of the force to the joint center) and will be able to move the distal end of the bone rapidly when the muscle contracts. The larger its moment arm about the joint, the slower the movement. However, the opposite is true for exertion of force at the distal end of the bone. Thus, skeletal morphologies differ, depending on whether predominantly powerful or rapid movements are required.

Digging and burrowing. Most changes in hole-dwelling animals are related to muscular insertions for increased power in their movements. However, skeletal adaptations occur in the teeth, pectoral and pelvic limbs, and vertebral column. Major modifications include limbs becoming shorter and more robust. The

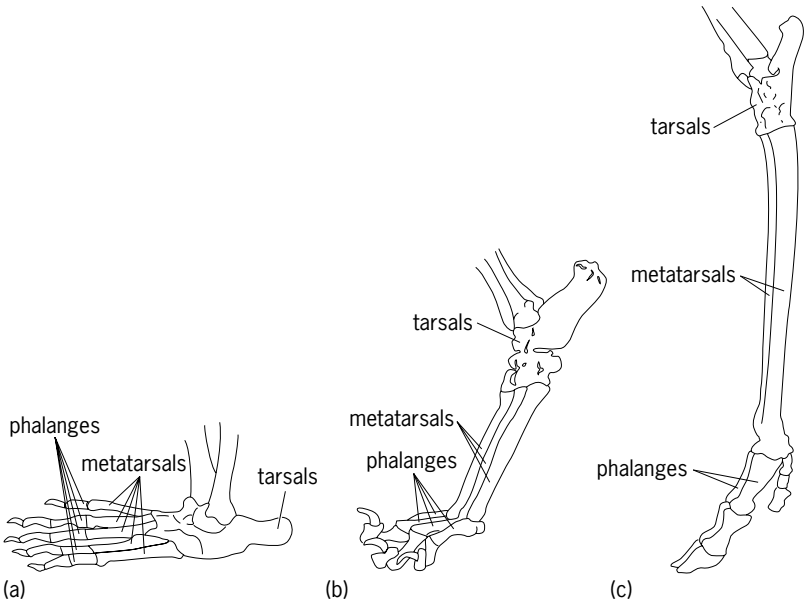


Fig. 7. Limb postures. (*a*) Plantigrade. (*b*) Digitigrade. (*c*) Unguligrade. (*After G. C. Kent, Comparative Anatomy of the Vertebrates, 7th ed., Mosby-Year Book, 1992*)

medial humeral epicondyle enlarges to accommodate an increased muscle bulk of carpal and digital flexor muscles and forearm pronators. Claws are either sharp to break hard soil or flattened to help shift soft soil, and often have large extensor tubercles on the dorsal surfaces. These act as passive bony stops to prevent overextension of the joints. Many joints are restricted to planar motion (flexion-extension) to increase their stability and make them less likely to dislocate.

Limbless locomotion. Frogs, birds, and terrestrial mammals have never produced limbless forms. Snakes (Ophidia) are the most diverse limbless group, but there are lizards (Sauria) and amphibia (Apoda) that show a reduction or loss of one or both limb pairs and girdles. The main adaptation accompanying limb loss is elongation of the body. More than 500 vertebrae have been documented in a snake. The overlapping bony scales used for protection and to gain purchase for movement are formed secondarily in the skin.

Aerial locomotion. Birds and bats (and possibly the extinct pterosaurs) are the only examples of vertebrates capable of continuous muscle-powered flight. Many others use the more passive gliding or parachuting. Flying fishes (for example, Exocoetidae, Hemirhamphidae) have enlarged pectoral fins enabling them to glide out of water to escape predators. A number of reptiles (for example, *Draco* spp.) glide with the aid of flight surfaces formed by elongated ribs.

Birds have a highly modified forelimb complex (**Fig. 8**). The humerus, ulna, and radius have surface features linked to flight, but the real adaptation of these (and other bird and pterosaur bones) is their air-filled or pneumatic character. This reduces their mass and, therefore, the muscular work involved in moving the wings up and down. In addition, the major wing bones of bats, birds, and pterosaurs have relatively large diameters and are relatively thin-walled. These traits maximize torsional strength for any given mass of bone material, which is important for flying animals in which twisting forces on the wing skeleton are commonly encountered. Support of the primary flight feathers is provided by the bones of the carpus, carpometacarpus (a fusion of carpals and metacarpals unique to birds), and manus, which in adult birds has only three digits (recently identified as digits II-III-IV). Wing length is effectively determined by the lengths of the arm, forearm, and hand skeleton, but wing shape (breadth and sharpness of wingtips) is determined by feather morphology. Secondary adaptations to flight are seen in the pelvic girdle, with its extensive connection to the synsacrum presumed to be linked to absorbing the shock of landing.

Bats have adopted a different skeletal strategy. They use highly elongated bones (humerus, radius, metacarpals, phalanges) to help support the whole wing, not only the leading edge as seen in birds. These bones are not pneumatic, but contain marrow. Weight savings do occur in distal elements due to a reduction in mineralization. Large pectoral muscles attach to ribs, and there is no massive, keeled sternum as in birds. Pelvic girdle reduction, thinner

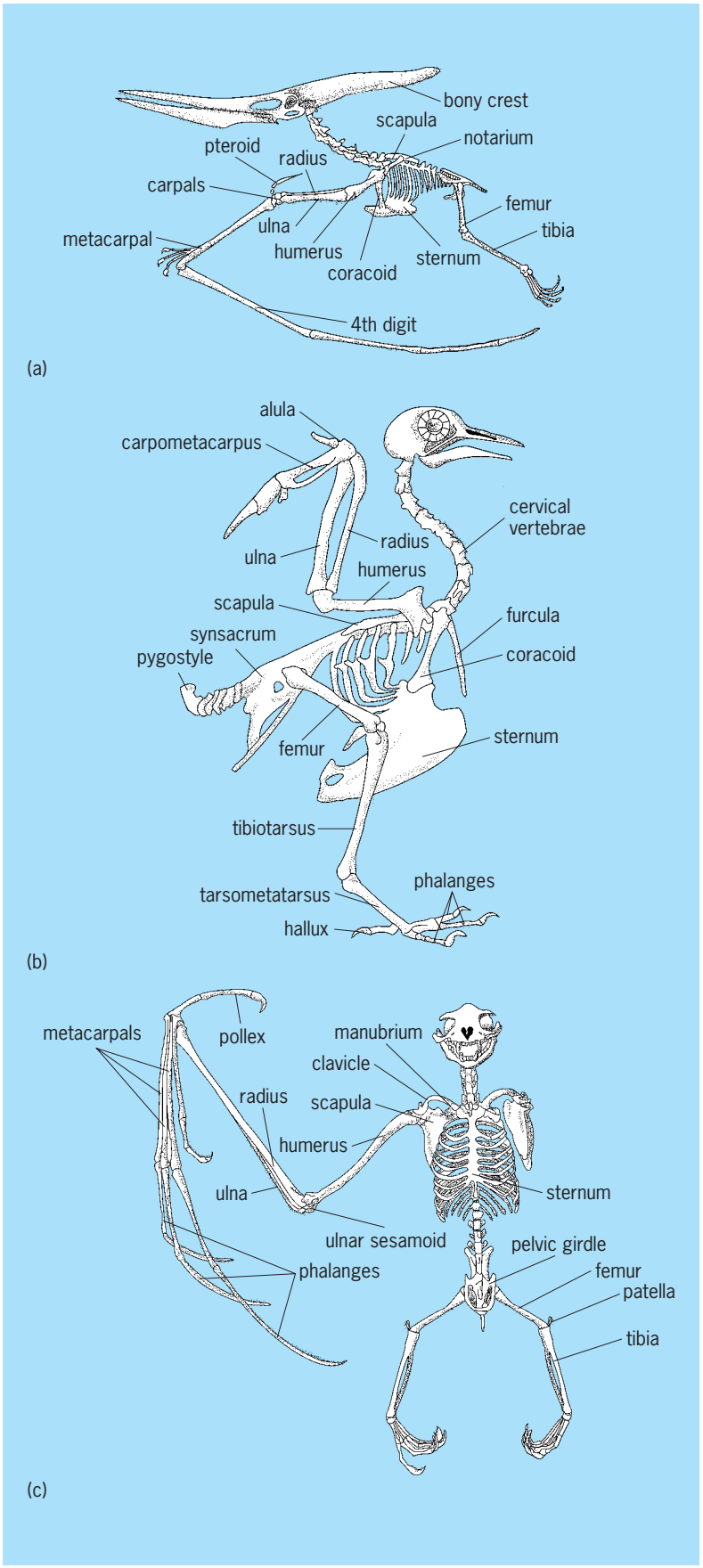


Fig. 8. Skeletons of flying vertebrates. (*a*) Extinct pterosaur (*Pteranodon*). (*b*) Typical bird skeleton. (*c*) Skeleton of a fruit bat or flying fox. (*After L. B. Halstead, Vertebrate Hard Tissues, Wykeham Publications, London, 1974*)

and straighter limb bones (they no longer have to resist large bending forces), rotation of the limb so the patella faces posteriorly, partial loss of the fibula, and possession of large claws are all skeletal adaptations for a hanging life-style. *See* FLIGHT.

Aquatic locomotion. In contrast to terrestrial locomotion, the limbs of aquatic organisms are hydrofoils or paddles. The skeletal elements are usually the same as typical tetrapods, with altered dimensions. Some specialist swimmers (such as porpoises and whales) have increased the number of phalanges in the hand. The manus is fairly stiff, the bones being bound together with connective tissue. Reduction or loss of the pelvic girdle and limbs has occurred in cetaceans and sirenians, and thrust is supplied by dorsal-ventral flexion of the vertebral column. Seals and walruses retain the pelvic structures, swimming by lateral flexion of the vertebral column.

Skeletal strength. Bones have to be stiff in order to act as levers for effective muscle action, and strong to resist failure. Although the form of the skeleton is genetically determined, the basic plan can be modified within limits by hormone action and the mechanical environment.

The limb skeleton of terrestrial tetrapods is subjected to large dynamic forces during rigorous exercise. For example, the feet of human runners generate loads of two to three times body weight. Even larger forces are applied to the skeleton by muscle action (for example, to the heel bone via the Achilles tendon). Direct measurement of bone deformation under such conditions shows that the limb bones in most animals undergo similar levels of strain (that is, change in specimen length per original length) and are typically three to five times as strong as they need to be to resist failure.

Long bones. Lamellar or cortical bone has the same mechanical properties whether taken from a mouse or an elephant. The strength of whole bones is therefore a function of the amount of bone present and its geometric arrangement. Long bones in tetrapod limbs are typically tubular, even though the marrow contained within the hollow center appears to have little functional use. Why the bones are not solid can be explained by referring to principles of biomechanics. As most bones have a degree of longitudinal curvature, they tend to bend in a predictable direction when loaded, for example, during running. One surface of the bone will tend to be under compression (squashed), and the opposite surface will be under tension (pulled). As there is a gradient across the bone running from compression to tension, there is a position between the surfaces where the forces cancel out (the neutral axis). Bone material along and close to this axis does not contribute much strength to the bone in bending. Such redundant material can be removed (forming the marrow cavity) without compromising bone strength.

Size and the skeleton. The bones of humans and elephants are proportionately more robust than those of mice. This occurs because a doubling in linear dimensions results in a fourfold increase in area and an eightfold increase in mass. If animals simply increased in size, it would result in the bones undergoing larger stresses in bigger animals. To keep bone stresses at the same level as in small animals requires that the bones be made thicker.

Another obvious feature is that small animals adopt crouching (noncursorial) postures, while larger ones are typically cursorial, standing with straighter limbs. A more upright limb posture results in a better alignment of the forces generated during locomotion with the long axes of the bones. This reduces both the forces that muscles need to generate to keep limbs extended, and the bending moments imposed on the bones. Cursorial postures lower stresses on the limb bones but reduce agility. *See* BIOPHYSICS; VERTEBRATA.

Mike Bennett

Bibliography. R. McN. Alexander, *Animal Mechanics*, 2d ed., 1983; J. D. Currev, *The Mechanical Adaptations of Bones*, 1984; M. Hildebrand et al. (eds.), *Functional Vertebrate Morphology*, Belknap Press, Cambridge, MA, 1985; A. S. King and J. McLelland, *Form and Function in Birds*, vol. 3, Academic Press, London, 1985; J. Klein-Nulend et al., Sensitivity of osteocytes to biomechanical stress in vitro, *FASEB J.*, 9:441–445, 1995; R. B. Martin and D. B. Burr, *Structure. Function and Adaptation of Compact Bone*, Raven Press, New York, 1989; F. H. Martini, *Fundamentals of Anatomy and Physiology*, 4th ed., Prentice Hall, Englewood Cliffs, NJ, 1998; M. Nordin and V. H. Frankel, *Basic Biomechanics of the Musculoskeletal System*, 2d ed., Lea and Febiger, Philadelphia, 1989; H. Zhuang et al., Electrical stimulation induces the level of TGF-β1 mRNA in osteoblastic cells by a mechanism involving calcium/calmodulin pathway, *Biochem. Biophys, Res. Comm.*, 237:225–229, 1997.

Skin

The entire outer surface of the body and the principal boundary between the external environment and the body's internal environment of cells and fluids. Skin serves as the primary barrier against the intrusion of foreign elements and organisms into the body, and also as a large and complex sense organ through which animals explore and learn about the external world. In addition, skin functions to maintain the homeostasis of the body's constituents, acting as a barrier to the loss of various ions and nutrients by diffusion. For terrestrial animals, it also serves as an effective barrier to water loss, without which most land animals would rapidly become desiccated and die.

Structure of mammalian skin. The skin of humans and other mammals can be divided into two distinct regions, the epidermis and the dermis (**Fig. 1**). The latter merges on its lower side with a looser connective tissue region called the hypodermis, which anchors the skin to deeper layers of connective tissue and is the principal site of fat deposition under the skin.

Epidermis. The outermost layer of the skin is the epidermis. It varies in thickness from 0.1 mm in

most of the protected areas of the skin to approximately 1 mm in those regions exposed to considerable friction, such as the soles of the feet and palms of the hands. The epidermis consists of a great many horizontally oriented layers of cells. The innermost layers, known as the stratum germinativum, consist of living and rapidly dividing cells. As cell division occurs in this region, the older cells are pushed toward the surface of the skin. In the process of this vertical migration, they become progressively flattened and lose their nuclei and most of their intracellular material. In effect, they die and become empty shells packed with the protein keratin. The outermost layer of the epidermis, the stratum corneum, consists of many layers of this packed cellular debris, forming an effective barrier to water loss from lower layers of the skin. Normally, its water content is exceptionally low, approximately 10% by weight. After soaking, however, the stratum corneum imbibes some external water, resulting in a wrinkled and shrunken (actually swollen) appearance. The outermost layer of the stratum corneum is continually sloughed off as small flakes. The greater the external wear and removal, the greater is the rate of cell division in the stratum germinativum and replacement of lost cells. It takes approximately 1 month for a cell to traverse the epidermis and be sloughed from its surface.

The lowest levels of stratum germinativum constitute the portion of the skin that contains melanocytes, cells that produce the dark pigment melanin. Different levels of melanin secretion are responsible for the large range of pigmentation observed among humans. Melanocytes respond with increased melanin secretion to a variety of stimuli, in particular, to ultraviolet radiation damage, resulting in tanning of the skin.

Dermis. The dermis plays a supportive and nutritive role for the epidermis. The epidermis has no blood supply of its own; that is, it does not contain any capillaries. However, nutrients and oxygen are apparently provided by diffusion from the blood supply of the underlying dermis, which extends in small, fingerlike projections (papillae) into the lower margin of the epidermis (Fig. 1). The dermis contains relatively few cells and numerous collagenous and elastic fibers, which give strength, cohesion, and resilience to the skin. The average thickness of the dermis is 1–3 mm. It is in this layer that the sebaceous and sweat glands are located and in which the hair follicles originate. The hair follicles are formed from downward invaginations of the stratum germinativum of the epidermis around dermal papillae of connective tissue. The surrounding dermis provides structural support and nutrition for the growing hair root. Associated with all hair follicles are oil-producing (sebaceous) glands, which empty their fatty contents into the hair follicle and lubricate the hair shaft. Two types of sweat glands, apocrine and eccrine, also originate in the dermis. Apocrine glands are confined to the axillary and anal-genital areas and produce viscous secretions; subsequent bacterial growth on these secretions is responsible for body odor. Eccrine glands are widely distributed over the human body and produce sweat, which has a thermoregulatory role in humans during exercise or heat stress. Both types of sweat glands have a duct running directly from the main part of the gland through the epidermis and emptying directly on the skin surface. The products of all these sets of glands are derived from the rich blood supply of the dermis. Hair, sweat glands, and mammary glands (which are modified sweat glands) are skin inclusions unique to mammals. *See* HAIR; SWEAT GLAND.

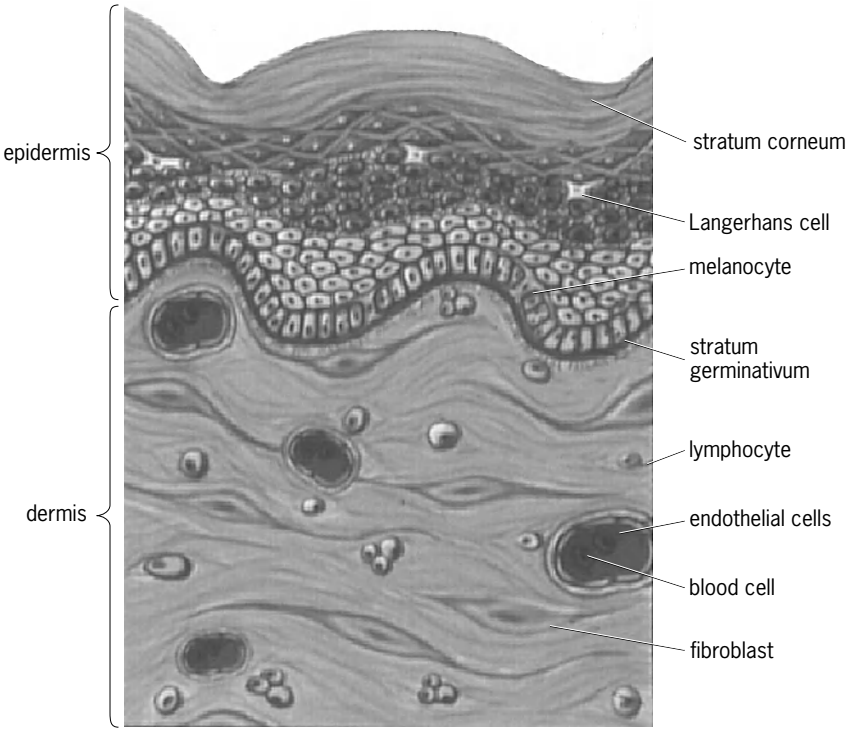


Fig. 1. Cross section of human skin. (*After Organogenesis, Inc.*).

Thermoregulation in mammals. Mammals are endothermic homeotherms; that is, their principal source of body heat is derived from their internal metabolic processes, and they maintain a constant body temperature. The skin plays a primary and essential role in thermoregulation in mammals, because it is across the skin that most heat is lost to or gained from the environment. Changes in skin insulation alone are very effective in adjusting rates of heat loss over a broad range of environmental temperatures, so that no extra metabolic heat has to be produced for thermoregulation. In this thermal neutral zone, insulation is altered by two mechanisms: piloerection and adjustment of dermal blood flow. To accomplish piloerection, a small smooth muscle fiber attaching the hair follicle to the skin contracts, raising the hair vertically. Erect hair traps a still layer of air next to the skin. This trapped air is a very efficient insulator, partially isolating an animal from its external thermal environment. The thicker and longer the hair, the more air is trapped and the more effective the insulation. Human hair on the body is too sparse to form an effective insulatory barrier, but the effects of piloerection can be seen in the

goose bumps that are produced on the skin when exposed to cold temperatures. Heat loss is also controlled by varying dermal blood flow. In warmer environments, the skin becomes flushed with blood, and metabolically produced heat can be dissipated into the environment. In cold environments, blood flow to the skin is restricted, effectively increasing the insulatory barrier between the internal source of heat production and the environment. These insulatory adjustments within the zone of thermal neutrality can be very effective at restricting heat loss. In the arctic fox, for example, an increase in metabolic rate due to shivering does not begin until environmental temperatures fall below -4°F (-20°C). *See* THERMOREGULATION.

In warm environments or during intense physical activity, which produces large amounts of metabolic heat, humans sweat. Sweat is a fluid similar in salt composition to that of blood plasma, from which it is derived. Sweating places large amounts of water directly on the surface of the skin. Some of this water evaporates, that is, is converted from a liquid to a gaseous state. Thus, metabolically produced heat may be brought to the skin through increased blood flow to the dermis, and this heat may then be dissipated to the environment through evaporation. Note, however, that the sweat must actually evaporate. If it simply drips off the body in its liquid form, no evaporative cooling is achieved. In very hot and dry environments, such as deserts, evaporation may be so rapid that humans might not be conscious of sweating, but under these conditions sweating proceeds at high rates and body dehydration can become a serious problem unless fluids are replaced regularly.

Sweating is the primary means of heat dissipation in humans, but it plays a substantially smaller role in most other mammals. While some mammals, such as horses, have widespread sweat glands for thermoregulation, many others have a very restricted distribution. In dogs, for instance, sweat glands are confined to the pads of the feet and probably serve principally to adjust pad friction during running. In mammals without a general sweat gland distribution, evaporative heat loss is undertaken by panting or by spreading saliva over the surface of the skin. In the latter case, the effect is the same in dissipating heat as if the fluid had been provided by sweating.

Evolution of vertebrate skin and scales. The most primitive living vertebrates are the agnathans, lampreys, and hagfish. While they are very different from other vertebrates in many aspects of anatomy and physiology, the basic organization of their skin is strikingly similar to that of mammals. They have an avascular, multicellular epidermis (although it contains mucus-secreting glands) underlain by a dermis containing elastic and collagen fibers and relatively few cells. There is a rich layer of melanocytes at the junction between the two layers. This basic bilayer organization of vertebrate skin must have evolved very early, or perhaps was a heritage from prevertebrate ancestors. *See* SCALE (ZOOLOGY).

Dermal structures. The first vertebrate group of which there is fossil evidence is the ostracoderms. These jawless, fishlike animals lived as bottom foragers in the seas a half billion years ago. Much of their body (especially their head and thorax) was covered with dermal bony armor. This material contained bone and denticles, the latter being composed of dentine and enamel, similar to that found in teeth in later vertebrates. This bony material was located in the dermis of the skin and was used as a defensive armor against predatory attack. A variety of similar hard dermal structures are found in some modern vertebrates. Sharks, for instance, have denticles that originate in the dermis and protrude through the epidermis, giving a rough, sandpaperlike feel to their skin. Teeth are thought to have evolved from such denticles located in the jaw region. Many fish and even some reptiles (for example, crocodiles) have bony scales (osteoderms) located in the dermis. These structures help to provide partially protective armor against external attack. Such protective elements are usually absent in mammals, although they did evolve convergently in a few groups, such as armadillos.

Scales appeared in the very first fossils of the next major vertebrate group, the placoderms. Scales are composite units of the skin, with a dermal core and overlying epidermal covering. They come in a wide variety of sizes and forms, sometimes containing bony or other supportive material, sometimes having only a connective tissue core. They are found in all descendant classes of vertebrates. They are particularly well developed in fish and reptiles. Birds, however, have well-developed scales on their legs, and their feathers were evolutionarily derived from scales. Certain groups of mammals also have scales: for example, rates and opossums have them on the tails. Most modern amphibians are scaleless; however, ancestral amphibians were almost certainly scaled, and some caecilians (apodans), an order of tropical wormlike amphibians, retain ancestral scales buried in the skin. Scales are apparently very convenient functional units in the skin. They permit a degree of protection for the underlying tissues and yet retain flexibility in the entire integument, permitting bending of the body or appendage by stretching or compressing the flexible areas between the scales. *See* FEATHER.

Functional role of skin. The functional role of the skin changed with the transition of vertebrates from aquatic to terrestrial environments. Even in aquatic environments, the skin serves as partial barrier to water entry or loss. Osmotic concentration of body fluids in most teleost fish (either marine or freshwater) is about 30% that of sea water. Thus, there is a tendency for fish to dehydrate (that is, to lose water from their body fluids) in ocean water and to become waterlogged in fresh water. Water permeability across the skin of fish, with its thick, scaly layers, is low, greatly lowering the potential water exchange with the surrounding fluid. In both marine and freshwater environments, however, water is abundant and freely available. Upon emergence into the terrestrial environment and air, primitive tetrapods found themselves in a very different hydric environment. Water was not always present and, even more significantly,

there was a continual loss of water from their bodies to the air by evaporation. Thus, any terrestrial environment in which the relative humidity is less than 100% is necessarily a desiccating environment. Much of the evolution of the tetrapod vertebrates was conditioned by the necessity of restricting water loss from the body. This was accomplished either by selecting wet and cool environments or by making a variety of physiological and morphological adaptations to retain internal water stores.

Among the tetrapods, there has been a general association between possession of a scaly skin and resistance to water loss across the integument. This association probably results from comparisons of scaleless frogs and salamanders, which have very high rates of cutaneous water loss, and scaled reptiles, such as lizards and snakes, which have the lowest cutaneous water loss rates among the vertebrates.

Water loss from amphibian skins. The problems of desiccation in the terrestrial environment are seen in sharpest relief among the amphibians, frogs, toads, and salamanders. Of all the terrestrial vertebrates, these organisms have the most permeable skins and greatest water permeabilities. Although some species have specialized adaptations to restrict cutaneous water loss, the vast majority lose water across their skins as if their bodies consisted of nothing more than an unprotected free surface of water. A skinned frog loses water at exactly the same rate as one with an intact skin, so the latter forms virtually no barrier to water loss. Most terrestrial species of amphibians have water permeabilities that are very similar to those of fully aquatic species; the adoption of terrestrial existence within this group has not been achieved by a reduction in the water permeability of the skin. Many terrestrial species would dehydrate and die in less than a day if they were placed in air and given no access to water. They are able to occupy terrestrial environments only because they behaviorally seek out and rehydrate in cool and wet areas. Thus, the entire life cycle of most amphibians, terrestrial as well as aquatic, is closely tied to the proximity to water.

Burrowing. In spite of their inherently high water permeabilities, some frogs and toads live in exceptionally hot and dry environments. For example, the spadefoot toad, *Scaphiopus couchii*, lives far from any permanent source of water in the arid desert of southeastern California, an exceedingly inhospitable place for an amphibian. It is able to survive by burrowing underground for most of its life and emerging onto the ground surface only after summer rainstorms. These rainstorms are infrequent and unpredictable, and are sometimes separated by several years in any particular local area. When a sufficient amount of rain falls to create temporary pools in the desert, the spadefoot toads emerge and rehydrate, not by drinking but by sitting in pools of rainwater and absorbing water across their skin. This process is under hormonal control, and rehydration can be exceptionally fast, with rates of water uptake of as much as 30% of their body weight per hour in small toads. The toads feast on the abundant insects available at that time, primarily termites which emerge with the rain. The toads mate and lay eggs in the temporary pools, and then dig down into the soil again. The tadpoles in the temporary rain pools develop exceptionally fast, for they are in a race to metamorphose into toadlets before the pool dries up. The entire period of surface activity for the toads may span only a couple of weeks out of the year, and in some years the toads may not emerge at all. Under the soil, the toads form cocoons of shed skin and enter a reduced metabolic state called estivation, in which their metabolic rates may fall to only one-fifth normal values. During estivation, they do not breathe: all gas exchange occurs across their skin. Their major source of fuel for the burrowing period is fat accumulated in abdominal fat bodies and throughout other organs. In addition, the toads may utilize 20–30% of their body's protein as a source of fuel. As a result of this protein breakdown, they also accumulate high concentrations of the nitrogen excretory product urea in their bodies. These high levels of urea aid water retention in the burrowed toads by reducing the tendency of water to diffuse across the skin into the surrounding soil. Further, less water is lost via urination by the storage of urea rather than its excretion.

Many species of frogs and toads use burrowing to escape from seasonally inhospitable environments. In many aspects of this response, their skin plays a crucial role, particularly during rehydration, underground respiration, and the formation of a cocoon. The cocoons are particularly important in reducing water loss to the surrounding dry soils. In some species, these consist of many layers of the shed portion of the stratum corneum. The cocoon in composed of dozens of layers of keratin and lowers the overall rate of water loss across the skin to almost immeasurable levels. The cocoons are shed one cell layer at a time, sometimes as often as once a day.

Mucus secretion. Some arboreal frogs are also exposed to very drying environments. They perch on tall grasses or shrubs, and are sometimes even exposed to direct sunlight. These frogs cannot escape dehydration by burrowing. Instead, they have developed various means to restrict water loss across their skins. Many treefrogs (for example, spring peepers, genus *Hyla*) have rates of water loss that are only one-third to one-half these of nonarboreal frogs of the same size. Part of this reduction in water loss appears to be the result of the accumulation of dried mucus on the skin. These treefrogs release mucus onto their backs when exposed to dry air, and as this evaporates, the residual protein is left on the skin surface. Water loss is further restricted when the frogs adopt a water-conserving posture of folding their limbs beneath them and pressing their bodies against the substrate. This behavior minimizes the amount of skin exposed to dry air from which evaporation could occur.

Waterproofing. In some African and South American treefrogs, the reduction in water loss across the skin has reached spectacular levels. These frogs are

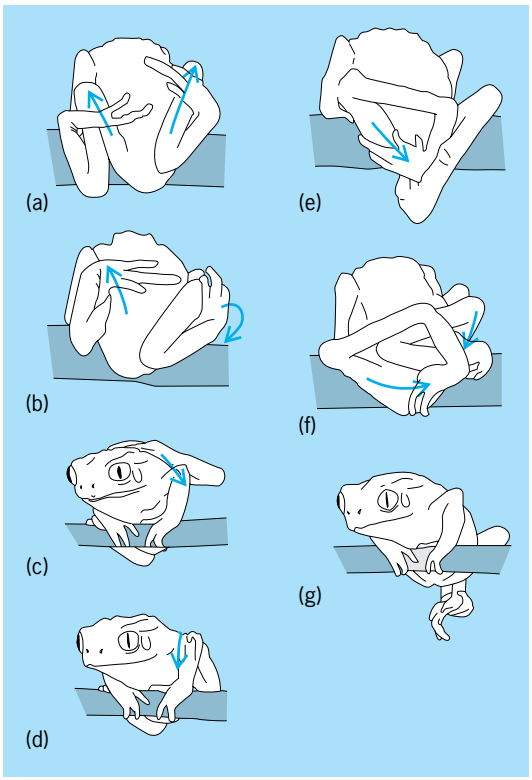


Fig. 2. The treefrog *Phyllomedusa sauvagei* wipes a waxy secretion from its skin all over its body with its feet in sequence *a–g*. Arrows show the direction of movement. After drying, this coating forms a very effective barrier to water loss from the skin. (*After L. A. Blaylock et al., Skin structure and wiping behavior of phyllomedusine frogs, Copeia, 1976:283–295, 1976*)

effectively waterproofed, reducing water loss in dry environments to levels fully equal to those of reptiles. Members of the genus *Chiromantis* live in very hot semiarid regions of southern Africa, perched on shrub branches, sometimes sitting directly in the sun. Thus, their ability to conserve water is crucial for survival. The functional basis of the waterproofing in *Chiromantis* is not known, although the stratum corneum on the backs (but not on the abdomens) of these frogs contains a unique layer immediately below the outer surface. This layer is likely involved in their low water permeability. Treefrogs of the genus *Phyllomedusa* also have extremely low rates of water loss across their skins; one species lives on exposed tree limbs in the arid Chaco region of South America and has been found to have body temperatures that are the highest ever recorded for an amphibian. These frogs are waterproofed by a unique lipid gland in the stratum corneum of their skin. The gland produces a waxy secretion that the frog wipes all over its body with its feet in a highly stereotyped behavioral sequence (**Fig. 2**). After this covering dries, the frog appears shiny, and droplets of water placed on its back bead up as on a waxed surface. This covering forms an extremely effective barrier to water loss. When body temperatures reach levels near 104°F (40°C), the treefrog remains cool by secreting water onto its back from numerous mucus glands located next to the lipid glands in its skin. The rate of this water secretion is carefully controlled so that heat gain from the environment is exactly balanced by evaporative heat loss, and body temperature remains constant. *See* INTEGUMENTARY PATTERNS; LIPID.

Albert F. Bennett

Bibliography. M. E. Feder and W. W. Burggren (eds.), *Environmental Physiology of the Amphib- ians*, 1992; W. N. McFarland et al., *Vertebrate Life*, 3d ed., 1989.

Skin disorders

The skin is subject to localized and generalized disorders, as well as those of primary occurrence in the skin and those secondary to involvement of other tissue. Diseases and disorders may affect any of the structures of the skin. They may be caused by external agents, either infectious or noninfectious, or by the abnormal accumulation of normal or abnormal skin elements, either inborn or acquired.

Infectious diseases. These are classified by the type of infectious agent—bacterial, parasitic, fungal, or viral.

Bacteria. Bacterial infections are distinguished clinically by the skin layer or appendage affected, but treatment is based upon the organism causing the infection. Impetigo and cellulitis are the most common infections of skin. Impetigo more commonly affects children and young adults, and it is usually more superficial and localized. Cellulitis is more common in older individuals and affects deeper layers in and larger areas of the skin, most often on the lower legs. Both infections may be caused by streptococcus, including group B hemolytic streptococcus, and staphylococcus, often species resistant to penicillin. Erysipelas is a rare streptococcal infection of the dermis, often of the face, characterized by high fever; it is potentially fatal if not promptly diagnosed and treated. Treatment is guided by culturing and determining the antibiotic sensitivity of the organism. Depending upon the results of the culture, treatment with antibiotics that are effective against organisms resistant to penicillin is recommended. *See* ANTIBIOTIC; DRUG RESISTANCE; STAPHYLOCOCCUS; STREPTOCOCCUS.

Skin appendages such as hair follicles may be similarly infected; the condition is known as folliculitis when individual follicles are involved, and furunculosis and carbunculosis when several or many adjacent follicles are involved in a large nodule or boil. Treatment is with antibiotics determined by culture but incision and drainage of the furuncles and carbuncles speeds treatment time. Acute infections of other skin appendages are less common, but sebaceous glands may be chronically infected in a triad of diseases including severe cystic acne, hidradenitis suppurativa, and dissecting cellulitis of the scalp. Treatment includes long-term antibiotic therapy, selective surgical intervention, and, in certain cases, retinoid drugs.

Parasites. Parasitic skin conditions are most often seen in epidemics among individuals who are in close contact, or where hygiene is poor, such as in

schools, mental institutions and other chronic care facilities, and shelters. Head, body, and pubic lice, and scabies are the most common. Pubic lice and scabies are also often transmitted by sexual contact. Scabies is also common in the general population, particularly in children and young adults. The mites of scabies tunnel within the epidermis, whereas all forms of lice live on the surface of the skin or in clothing. Treatment consists of topical pesticides, most often lindane. Treatment of allergic reactions to the insect parts or bites, particularly in scabies, is often necessary. *See* MEDICAL PARASITOLOGY; PESTICIDE.

Fungi. Fungal infections are extremely common. In the United States, most fungal infections in humans are with species incapable of infecting tissue other than keratinized epidermis. These organisms are known as dermatophytes; they cause tinea pedis (athlete's foot), tinea cruris (some forms of jock itch), and tinea capitis (a scalp condition responsible for some forms of hair loss). Regardless of the specific dermatophyte organism involved, all such infections are treated with topical or oral antifungal antibiotics that work by interfering with formations of the fungal wall. The length of treatment varies from as little as a few days of oral treatment for tinea capitis to a lengthy treatment that may require as much as 2 years for fungal toenail disease. *See* MEDICAL MYCOLOGY.

Viruses. Viral diseases often involve the skin, including conditions such as measles, chickenpox, and many others. This discussion, however, will be limited to viral diseases that affect only the skin. Viral warts (verruca vulgaris) and molluscum contagiosum are the primary examples. Both are characterized by single or multiple, somewhat contagious, skin tumors that usually are small but can in rare instances exceed 0.4 in. (1 cm) in diameter. They respond to various treatments, and they may disappear spontaneously. Both diseases can affect genital skin, and they are then much more likely to be spread by sexual contact. *See* ANIMAL VIRUS; CHICKENPOX AND SHINGLES; HERPES; MEASLES; SMALLPOX.

Inflammatory disease. Most itchy rashes are due to inflammation of the skin; they are usually known as eczema or dermatitis. Inflammatory skin diseases can occur as allergic reactions to specific chemicals in the environment, such as poison ivy; or they are caused by direct damage to the skin, such as by excessive exposure to soaps, detergents, or chemicals. In the acute stage, eczematous dermatitis is characterized by a vesicular, oozing condition. If the condition becomes chronic, the skin becomes dry, scaling, and leathery. When a specific cause can be identified, it must be avoided. Topical corticosteroids are the mainstay of treatment.

Seborrheic dermatitis is a common eczematous condition affecting primarily the areas of skin that bear sebaceous glands, that is, scalp, central face, chest, axilla, and groin. This type of dermatitis also usually responds to topical corticosteroids. Evidence has implicated the organism *Pityosporum ovale*, which explains the response to oral and topical ketoconizole (an antibiotic that kills the organism) in some cases, and the very high incidence in patients with acquired immune deficiency syndrome (AIDS), who are susceptible to many infectious diseases.

Hereditary disease. Atopic dermatitis is the skin manifestation of atopy, a clinically apparent hypersensitivity. This condition may also be associated with asthma and pollen allergies. It is hereditary and first manifests itself in infancy or early childhood. Some children outgrow it. However, it can persist into adulthood, and sometimes it does not appear until later in life. Atopic dermatitis is more common in blacks and people from Asia, particularly the Philippines. The main treatment is topical corticosteroids.

Psoriasis is a common disease of unknown etiology. It has been estimated to occur in up to 3% of the population of the United States. The disease is hereditary, but not in a simple mendelian pattern. Psoriasis is usually seen in young and middle-aged adults and it is rare in young children. Psoriasis is usually chronic for years or for the rest of the individual's lifetime. There is, however, a poststreptococcal form of the disease that is seen in adolescents and young adults; this type responds to antibiotics, and it may be cured completely, without recurrences. The typical psoriasis consists of well-defined patches and plaques of red skin with a silvery scale that often results in pinpoint bleeding when removed. The most common and persistent sites are the elbows, knees, and scalp, but any area of skin may be involved, including the palms and soles; there may be significant morbidity and disability. Nails may be involved, even in the absence of skin involvement. Psoriatic arthritis, which is similar to but distinct from rheumatoid arthritis, occurs in a small number of individuals; it may be severe enough to cause disability. There are a number of treatment modes for psoriasis, both local and systemic. Each patient requires an individual course of treatment, based on the concept that it should be adequate and not worse than the disease in terms of time, expense, discomfort, unsightliness, or potential medical complications.

Other conditions. Acne vulgaris is an extremely common skin disorder, affecting 80–90% of young adults, usually during adolescence. The condition varies widely in severity, distribution, and duration between individuals and over time in the same individual. The pilosebaceous organ of the skin is the primary target, particularly the sebaceous gland, its duct, and the infundibulum of the hair follicle. This organ is responsive to hormones and is variously stimulated by the different steroid hormones, including the sex hormones; this is the cause of the prevalence of the disease in the years immediately after puberty. In addition to increased production of sebum, bacteria contribute to the development of acne lesions. These lesions consist of open and closed comedones (blackheads), papules, pustules, and cysts. Scarring may result, particularly in cystic acne or in lesions that are physically manipulated. A large variety of treatments exist, directed at the bacterial, sebum production, and structural aspects of acne. Often several agents are used together to treat several aspects of the problem simultaneously. Each

individual must be evaluated, and a treatment regimen selected that can be modified as the condition responds or changes in severity over time.

Reactions to ingested materials, such as food or medications, often appear in the skin, usually either as a red, itchy, measleslike rash or as urticaria (hives). These reactions indicate allergy to the material, which must then be avoided. Reexposure may result in a more severe, potentially fatal, anaphylactic reaction. Treatment for these reactions consists of antihistamines and corticosteroids, both topically and, in severe cases, systemically. Under rare circumstances, usually where allergy to a potentially life-saving antibiotic or other drug exists, the agent may be given in a carefully controlled situation in the hospital or similar facility where immediate care for a severe allergic reaction can be provided. *See* ALLERGY.

Neoplasms. Skin neoplasms may be benign or malignant, congenital or acquired, and they may arise from any component of the skin. Almost all skin neoplasms are benign and acquired. The common mole (melanocytic nevus) is a neoplasm of benign melanocytes; usually it is acquired, but it may be present at birth, when it is often known as a birthmark. With the exception of giant congenital melanocytic nevi, which have approximately a 10% chance of becoming malignant and thus should be removed, nevi do not seem to have any greater risk of becoming malignant than any other part of the skin. Other common congenital nevi or birthmarks are of vascular origin, including strawberry and cavernous hemangiomas and port wine stains. With rare exception, these are of cosmetic significance only.

Malignant neoplasms may arise from cellular elements of the epidermis or dermis, or by infiltration of the skin by malignant cells arising from other tissues. By far the most common are basal cell and squamous cell cancers, which arise from basal and squamous keratinocytes of the epidermis, respectively. These constitute the most common form of cancer in the United States, with over 600,000 new cases per year. They are caused by the cumulative effects of ultraviolet radiation on the skin. Luckily, they are locally invasive in almost all cases, rarely metastasize or cause death, and are readily recognized and treated with standard outpatient procedures. They usually are characterized by a nonhealing sore, persistent red scaling or crusting patch, or a slowly growing pearly nodule on skin that has been exposed to the sun; they occur mostly on the head, neck, hands, and arms.

Malignant melanoma arises from the pigment-forming melanocyte, and thus it is usually pigmented. It is a metastasizing cancer that is often fatal if not removed surgically in the early stage. Once the cancer is advanced, chemotherapy and other interventions are rarely more than palliative. It can be recognized as a pigmented lesion, often thought to be a benign nevus initially, that increases in size, changes color, particularly with admixtures of black, blue, red, or white along with the usual tans or browns; and it becomes irregular in size and shape. Other signs and symptoms, such as bleeding, pain, and itching, are less frequent, often occurring as late manifestations. Malignant melanoma is one of the most rapidly increasing cancers—probably related to increased ultraviolet radiation exposure. *See* CANCER (MEDICINE).

Two unusual multicentric primary skin malignancies are mycosis fungoides and Kaposi's sarcoma. Mycosis fungoides is a lymphoma of the skin, usually present in several sites when first diagnosed, that may remain confined to the skin for 10 or more years before eventually spreading to internal organs and causing death. It may be extremely difficult to diagnose, both clinically, when it can appear only as eczematous patches, and histologically, for months or years. Often, multiple biopsies over time are required to make the diagnosis. It is not clear whether early aggressive chemotherapy can result in a cure or increased life-span, or whether such therapy may even shorten life. Kaposi's sarcoma occurs in two forms, the classic form seen on the legs of elderly Mediterranean men, and a form associated with HIV-1 infection and AIDS that may occur on any part of the body. It is derived from skin blood vessels, is multicentric, and usually appears as red to violet patches, plaques, or nodules. It is usually not fatal, although it may eventually spread to internal organs and may cause significant morbidity. It is treated palliatively by ionizing radiation and local surgery. When it is extensive and internal, the treatment includes chemotherapy. There are numerous other primary skin cancers, but they are rare. *See* ACQUIRED IMMUNE DEFICIENCY SYNDROME (AIDS); SKIN.

Alan N. Moshell

Bibliography. T. B. Fitzpatrick et al., *Dermatology in General Medicine*, 1987; L. A. Goldsmith (ed.), *Biochemistry and Physiology of the Skin*, 1983; W. F. Lever and G. Schaumburg-Lever, *Lever's Histopathology of the Skin*, 8th ed., 1997; W. Montagna and P. F. Parakkal, *The Structure and Function of Skin*, 1974; S. L. Moschella and H. J. Hurley, *Dermatology*, 3d ed., 1992.

Skin effect (electricity)

The tendency for an alternating current to concentrate near the outer part or "skin" of a conductor. For a steady unidirectional current through a homogeneous conductor, the current distribution is uniform over the cross section; that is, the current density is the same at all points in the cross section. With an alternating current, the current is displaced more and more to the surface as the frequency increases. The conductor's effective cross section is therefore reduced so the resistance and energy dissipation are increased compared with the values for a uniformly distributed current. The effective resistance of a wire rises significantly with frequency; for example, for a copper wire of 1-mm (0.04-in.) diameter, the resistance at a frequency of 1 MHz is almost four times the dc value. *See* ALTERNATING CURRENT; ELECTRICAL RESISTANCE.

A mathematical description of skin effect can be

derived from Maxwell's equations, but exact solutions have been obtained for only several simple shapes, including cylindrical, tubular, and flat conductors. For a plane conductor carrying a sinusoidal alternating current, the current density is a maximum at the surface and its magnitude decreases exponentially with distance into the conductor. A skin depth or penetration depth δ is frequently used in assessing the results of skin effect; it is the depth below the conductor surface at which the current density has decreased to $1/e$ (approximately 37%) of its value at the surface and is given by $\sqrt{1/(\pi f \mu \sigma)}$, where f is the frequency and σ and μ are the conductivity and permeability of the material respectively. This concept applies strictly only to plane solids, but can be extended to other shapes provided the radius of curvature of the conductor surface is appreciably greater than δ. *See* MAXWELL'S EQUATIONS.

At a frequency of 60 Hz the penetration depth in copper ($\sigma = 5.8 \times 10^7$ S m^{-1}, $\mu = 4\pi \times 10^{-7}$ H m^{-1}) is found from the above expression to be 8.5 mm (0.33 in.); at 10 GHz it is only 6.6×10^{-7} m. Waveguide and resonant cavity internal surfaces for use at microwave frequencies are therefore frequently plated with a high-conductivity material, such as silver, to reduce the energy losses since nearly all the current is concentrated at the surface. Provided the plating material is thick compared to δ, the conductor is as good as a solid conductor of the coating material. *See* CAVITY RESONATOR; MICROWAVE; WAVEGUIDE.

In the case of a cylindrical conductor carrying a sinusoidal alternating current, the solution to the skin-effect equation is determined in terms of the zero-order Bessel functions of the first kind. When the frequency is high, the exact formulation reduces to the exponential solution for a plane conductor. A useful approximation to the ratio of ac-to-dc resistance of a round conductor of radius r is $r/2\delta$, if $\delta \leq 0.1r$. For low frequencies or conductors of small radii, the exact ratio in terms of the Bessel functions must, however, be employed. At high frequencies the central part of a cylindrical conductor carries so little current that it is often removed to give a tube which saves on weight and can also be convenient for passing cooling fluid. The skin-effect equation for tubular conductors is solved in terms of Bessel functions of the first and second kind. *See* BESSEL FUNCTIONS.

Frank A. Benson

Bibliography. B. I. Bleaney and B. Bleaney, *Electricity and Magnetism*, 3d ed., 1976, reprint 1989; W. H. Hayt, *Engineering Electromagnetics*, 6th ed., 2001; M. A. Plonus, *Applied Electromagnetics*, 1978; S. Ramo, J. R. Whinnery, and T. van Duzer, *Fields and Waves in Communication Electronics*, 3d ed., 1994.

Skink

A widely dispersed reptile belonging to the family Scincidae. The family contains approximately 1200 species and is represented on every continent except Antarctica. There are over 150 species of skink in Australia alone. Characteristics include elongate bodies, moderate to long tails, and proportionately short, reduced, or absent limbs. Some scales on the head and body have cores of bone. In many species the eyelids are transparent. Skinks exhibit pleurodont dentition; that is, the teeth are attached on the side of the jaw. *See* DENTITION.

Five-lined skink (*Eumeces fasciatus*). (*Photo by George W. Robinson; © California Academy of Sciences*)

Skinks vary widely in size and longevity (some have been known to live to 20 years). They are carnivorous, eating invertebrates and, in the case of larger species, even small mammals, reptiles, and birds. Three reproductive modes can be found among skink species: oviparity (egg-laying), viviparity (bearing live young), and ovoviviparity (producing eggs that develop internally and hatch before or soon after extrusion). Three genera, *Neoseps*, *Scincella*, and *Eumeces*, occur in the United States. They are represented by species such as the federally threatened sand skink (*N. reynoldsi*) of Florida, which has reduced limbs and ear openings covered by scales that are specifically adapted for burrowing; the ground skink (*S. lateralis*), the only *Scincella* species found throughout the eastern regions; and about 15 species of the common genus *Eumeces*, the striped skinks (see **illustration**).

A number of species of the genus *Tiliqua*, represented by the largest skinks in terms of body size, inhabit Malaysia and Australia. *Tiliqua gigas* reaches a length of 2 ft (0.6 m). The blue-tongued skink (*T. scincoides*), also native to Australia, can survive a wide temperature range. Conversely, the Egyptian skink (*Scincus officianalis*), a common species of the Sahara region, is lethargic unless the temperature is near 100°F (38°C). It has a flattened snout that is used for burrowing into the sand. Its wide, thin toes resemble feathers and are well adapted for walking on sand. Another Sahara species, *Chalcides sepoides*, has a cylindrical body and four small limbs that fold against the body when the animal moves through sand, permitting a snakelike body movement. A similar species, the three-toed sand skink (*C. striatus*), common in France and the Iberian Peninsula, also folds its legs when crawling, but unlike the previous species it has depressions in the

side of the body to accommodate the folded limbs. These highly adapted species are but a few examples of the diversity of skinks worldwide. *See* REPTILIA; SQUAMATA. W. Ben Cash

Bibliography. F. H. Pough et al., *Herpetology*, 3d ed., Prentice Hall, Upper Saddle River, New Jersey, 2004; G. R. Zug, L. J. Vitt, and J. P. Caldwell, *Herpetology* Academic Press, San Diego, 2001.

Skunk

A carnivore long classified in the subfamily Mephitinae in the family Mustelidae (otters, weasels, mink, badgers, and their relatives), but since 1997 placed in the independent family Mephitidae along with the Oriental stink badger *Mydaus*. This new placement was based on convincing DNA evidence linking skunks (genera *Mephitis*, *Spilogale*, and *Conepatus*) and *Mydaus* genetically.

Skunks are found only in the New World, ranging from Canada to Chile. These mammals have short limbs with five clawed toes on each foot. The claws are curved and nonretractile. A pair of well-developed anal scent glands, used for defense against predators, are present. A short rostrum is present on the elongate and flattened skull. Carnassial teeth (used to shear flesh and bone) are well developed.

Striped and hooded skunks (Mephitis). These two species range from southern Canada to Costa Rica. Adults are about the size of a house cat, standing approximately 18 cm (7 in.) at the shoulder, with adults usually attaining a head and body length between 550 and 750 mm (21 and 29 in.), including a tail 175–300 mm (7–12 in.) long. Adults usually weigh between 1.3 and 4.5 kg (3 and 10 lb). The dental formula is I 3/3, C 1/1, Pm 3/3, M 1/2 × 2, for a total of 34 teeth. *See* DENTITION.

The relatively small head possesses small, short, rounded ears, small eyes, and a pointed muzzle. The feet are semiplantigrade with long claws on the forefeet for digging. The pelage (the coat, including fur and hairs) is long, coarse, and oily. The body is black with a narrow white stripe running up the middle of the forehead and a broad white area on the nape of the neck that usually divides into a V at about the shoulders (see **illustration**). The resulting two white stripes may continue to the base of the bushy tail. There is, however, much variation in the length and width of the stripes; some skunks have very broad, well-defined stripes, whereas in others the stripes are almost or completely lacking.

Striped skunks prefer bushy and sparsely wooded areas and are most abundant in agricultural regions where the forest gives way to open lands and farm crops. Hooded skunks occur at low elevations in desert areas. Both species are mainly nocturnal. Daylight hours are usually spent resting or sleeping in a hollow log, in wood piles, beneath an abandoned building, along a fence row, or in a burrow in the ground. Striped skunks may become dormant for prolonged periods during severe winter weather, but they do not hibernate. Striped and hooded skunks are not good climbers. When frightened or annoyed, the skunk contracts the muscles surrounding the scent glands, thus forcing the scent out in a fine, almost invisible spray. At least seven major volatile components have been identified in the secretion of the striped skunk. The fluid has a most distasteful and nauseating stench and produces intense smarting and burning if it comes in contact with the membranes of the eyes, nose, or mouth. Incongruous as it may seem, the oily yellow fluid produced by the skunk is used as a base for perfumes.

Striped skunk (*Mephitis mephitis*). (*Photo © Creatas/Punch Stock*)

Striped and hooded skunks are omnivorous and feed on insects, crayfish, fish, frogs, turtle eggs, snakes, small birds, and small mammals. During fall and winter, fruit, small mammals, carrion, grains and nuts, grasses, leaves, and buds are important.

Striped skunks are polygamous and usually mate from January through March. Most females normally produce a single annual litter. Following a gestation of 59–77 days, a litter of 3–10 blind young are born in a nest of dry leaves and grasses. They have a fine covering of hair in which the adult pattern is evident. The eyes open at 22–35 days of age, and by 7–8 weeks they are out following their mother in single file on her nightly foraging trips. They are weaned at 8 weeks. Striped skunks may live 5 to 6 years in the wild. The main predators are the great horned owl and the barred owl.

Spotted skunk (Spilogale). Spotted skunks are the smallest and most weasel like in appearance of the three genera of skunks. This genus consists of three species which range from southern British Columbia to Oaxaca, Mexico, and eastward to Pennsylvania and the southeastern United States. Spotted skunks

may be distinguished from the other two genera by their smaller size, forehead patch, and pattern of stripes and spots. The basic color pattern consists of six white stripes extending along the back and sides; these are broken into smaller stripes and spots on the rump. There is a triangular patch in the middle of the forehead, and the tail is usually tipped with white. Variations, however, are infinite. A pair of scent glands is present under the base of the tail. The dental formula is I 3/3, C 1/1, PM 3/3, M 1/2 × 2, for a total of 34 teeth. Adults are between 185 and 565 mm (7 and 22 in.) in total length, including an 80–220 mm (3–8 in.) tail. They weigh 200–1000 g (7–35 oz).

Spotted skunks inhabit a variety of brushy, rocky, and wooded habitats but avoid dense forests and wetlands. In the states, they inhabit riparian woodlands and areas of vegetation along fences. The omnivorous diet consists mainly of insects and vegetation in the summer and rodents and other small animals in the winter. Unlike striped and hooded skunks, spotted skunks are agile climbers.

Breeding is not the same in all parts of the range, with some populations mating during March and April and others mating in September and October. Gestation may range from 50 to 65 days in spring breeders to 230 to 250 days in fall breeders. The period of delayed implantation may range from 2 weeks in spring breeders to as long as 6 months in fall breeding populations. Litters average 3 to 6 young. The eyes open at 32 days, and they are weaned at about 54 days of age. Longevity in the wild is unknown, but a captive lived 9 years 10 months.

Hog-nosed skunks (Conepatus). Five species of hog-nosed skunks range from southern Colorado and eastern Texas to southern Chile and Argentina. These skunks have the coarsest fur of all skunks. This genus can be distinguished from the other two by the absence of a white head stripe and by its long head and naked muzzle, somewhat like a hog's snout. This adaptation allows the skunk to root in the ground for insects, which make up a good part of its diet. The color pattern is variable. In some, the top of the head, the back, and the tail are white, and the remainder of the animal is black. In others, the pattern resembles that of *Mephitis*, with the pelage being black except for two white stripes and a mostly white tail. Adults are 460–900 mm (18–35 in.) long, including a 160–410 mm (6–16 in.) tail. They weigh 2.3–4.5 kg (5–10 lb). Little is known concerning reproduction and longevity.

Donald W. Linzey

Bibliography. J. W. Dragoo and R. L. Honeycutt, Systematics of mustelid-like carnivores, *J. Mammal*, 78:426–443, 1997; D. W. Linzey, *The Mammals of Virginia*, McDonald & Woodward Publishing, 1998; R. N. Nowak, *Walker's Mammals of the World*, 6th ed., Johns Hopkins University Press, 1999; J. O. Whitaker, Jr., and W. J. Hamilton, Jr., *Mammals of the Eastern United States*, 3d ed., Comstock Publishing Associates, 1998; D. E. Wilson and S. Ruff (eds.), *The Smithsonian Book of North American Mammals*, Smithsonian Institution Press, 1999.

Slate

Any of the deformed fine-grained, mica-rich rocks that are derived primarily from mudstones and shales, containing a well-developed, penetrative foliation that is called slaty cleavage (see **illus.**). Slaty cleavage is a secondary fabric element that forms under low-temperature conditions (less than 540°F or 300°C), and imparts to the rock a tendency to split along planes. It is a type of penetrative fabric; that is, the rock can be split into smaller and smaller pieces, down to the size of the individual grains. If there is an obvious spacing between fabric elements (practically, greater than 1 mm), the fabric is called spaced. Slates typically contain clay minerals (for example, smectite), muscovite/illite, chlorite, quartz, and a variety of accessory phases (such as epidote or iron oxides). Under increasing temperature conditions, slate grades into phyllite and schist. *See* ARGILLACEOUS ROCKS; CLAY MINERALS; PHYLLITE; SCHIST.

Cleavage categories. There are four main categories of cleavage based on morphological characteristics: disjunctive cleavage, pencil cleavage, slaty cleavage, and crenulation cleavage.

Disjunctive cleavage. This type of cleavage is defined by an array of subparallel fabric elements, called cleavage domains, in which the original rock fabric

Slaty cleavage in folded mica-rich rock near Walland, Blount County, Tennessee. (*USGS*)

and composition have been markedly changed by the process of pressure solution. Domains are separated from one another by intervals, called microlithons, in which the original rock fabric and composition are more or less preserved. The cleavage domains cut across a preexisting fabric (usually bedding) that is preserved in the microlithons. In a horizontal bed of argillaceous (clay-rich) limestone or sandstone that is subjected to a compressive stress, dissolved ions created by pressure solution diffuse away from the site of dissolution through a fluid film that adheres to the grain surfaces. The ions precipitate at crystal faces where compressive stress is less, precipitate nearby in pressure shadows adjacent to rigid grains, or enter the pore fluid system and are carried out of the local rock environment entirely. Where it occurs, the process preferentially removes more soluble grains. Thus, in an argillaceous limestone, calcite is removed and clay and quartz are progressively concentrated. In an argillaceous sandstone, the process is effectively the same, except that quartz is the mineral that preferentially dissolves and clay alone is concentrated. Eventually, a discrete domain develops in which there is a selvage composed of mostly clay with some relict corroded calcite or quartz grains. If deformation continues, the domain further thickens by pressure solution along its edges. As a result, compositional contrasts between cleavage domains and microlithons become so pronounced that a new layering is defined in the rock.

Pencil cleavage. When a fine-grained sedimentary rock (shale or mudstone) breaks into elongate pencillike shards, it is said to have a pencil cleavage. Typically, pencils are 5–10 cm (2–4 in.) long and 0.5–1 cm (0.2–0.4 in.) in diameter. In outcrop, pencil cleavage looks like it results from the interaction of two fracture sets, but the parting reflects the internal alignment of clay grains in the rock. Because of their strong shape anisotropy, clay flakes attain a preferred orientation parallel to bedding when they settle out of water and are compacted together. This preferred orientation imparts the tendency for clay to break on bedding planes that is displayed by shale. If the shale is subjected to layer-parallel shortening, rotation and microfolding of preexisting grains and growth of new grains occur. The plane of new flakes or parts of deformed flakes is roughly perpendicular to the shortening direction (that is, at a high angle to bedding). At an early stage during this process, the new deformation fabric is comparable in degree of development to the initial bedding fabric, creating the pencils.

Slaty cleavage. Pencil cleavage may be considered a snapshot of an early stage in the process by which slaty cleavage develops. If shortening perpendicular to cleavage planes continues, clay throughout the rock displays a preferred orientation at an angle to the primary sedimentary fabric. Eventually, the deformation fabric dominates the primary fabric. Formation of this secondary fabric occurs by much the same process as the formation of disjunctive cleavage; however, the resulting domains are so closely spaced that effectively there are no uncleared microlithons, and the clay in the entire rock displays the tectonically induced preferred orientation. When a rock has this type of fabric, it is said to have slaty cleavage. Thus, slaty cleavage is defined by a strong dimensional preferred orientation of clay in a very clay rich, low-grade metamorphic rock, and the resulting rock is a slate. Slaty cleavage tends to be smooth and planar. Coupled with the penetrative nature of slaty cleavage, these characteristics enable slates to split into very thin sheets. This and the durability of the rock are reasons why slates are used in the roofing industry, in the tile industry, and in the construction of pool tables.

Crenulation cleavage. Penetrative microfolding of the original bedding foliation or a secondary foliation (such as slaty cleavage) produces the fourth morphological cleavage type, called crenulation cleavage. Crenulation cleavage is characterized by the presence of microfolds with near-parallel limbs that define the cleavage surface. *See* ROCK CLEAVAGE.

Cleavage and strain. In order to determine the relationship of cleavage to strain, it is necessary to consider strain markers in a cleaved rock, such as reduction spots. Reduction spots are small regions where the iron in a red slate is reduced, and therefore look greenish. If these spots start out as spheres, they make ideal strain markers; and geologists find that the deformed spots are flattened ellipsoids with the plane of flattening almost parallel to the slaty cleavage domains. However, cleavage is probably not perfectly parallel to the flattening plane of strain, given the evidence for shear parallel to the cleavage. Estimates of the total shortening associated with cleavage formation are in excess of 60%. There is a question whether strain resulting from cleavage formation is a volume-constant strain or a volume-loss strain. If ions of soluble minerals enter the pore water system, they can be carried out of the local rock system by movement of ground water. In some rocks the volume may have decreased by up to 50% during cleavage formation. *See* SHALE.

Ben A. van der Pluijm

Bibliography. B. A. van der Pluijm and S. Marshak, Fabrics, ch. 11 in *Earth Structure: An Introduction to Structural Geology and Tectonics*, McGraw-Hill, Dubuque, 1997.

Sleep and dreaming

Sleep is generally defined as an easily reversible, temporary, periodic state of suspended behavioral activity, unresponsiveness, and perceptual disengagement from the environment. Compared with other states of temporary unresponsiveness such as syncope or coma, sleep is easily reversible with strong or meaningful sensory stimuli (for example, the roar of a nearby tiger, or a voice speaking the sleeper's name). Sleep should not be considered a state of general unconsciousness. The sleeper is normally unconscious (but not always) of the nature of events in the surrounding environment; this is the meaning of perceptual disengagement: a lack of conscious perception and meaningful responsiveness to environmental

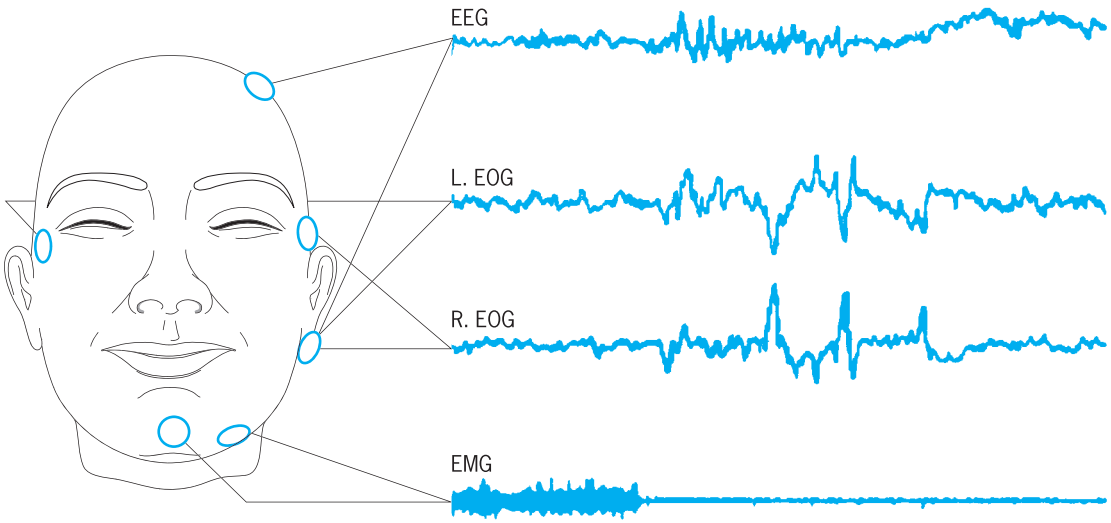


Fig. 1. Standard electrode placements for sleep research. Electroencephalogram (EEG) waves are recorded from left (C3) or right (C4) central locations, which give the best discrimination of sleep stages. Electrooculogram (EOG) electrodes are typically sited to allow both vertical and horizontal eye movements to be recorded. The standard chin (submental) electromyogram (EMG) site is chosen to show reliable suppression of muscle tone during onset of REM (rapid eye movement) sleep.

stimuli. However, the sleeper's attention may be fully engaged in experiencing a dream. And if reportability is accepted as a sufficient condition for conscious mental processes, any dream that can be recalled must be considered conscious. Dreaming, then, can be simply defined as the world-modeling constructive process through which people have experiences during sleep, and a dream is just whatever the dreamer experienced while sleeping.

Nature of sleep. In general, biological organisms do not remain long in states of either rest or activity. For example, if a cat's blood sugar level drops below a certain point (the cat's reference level for hunger-related action), the cat is motivated by hunger to venture from its den in search of a meal. It will continue to hunt until it finds food (or starves). After satisfying the urge to eat, the cat is no longer motivated to expend energy tracking down uncooperative prey; now its biochemical state motivates a return to its den, to digest in peace, conserve energy, and generally engage in restful, regenerative activities, including sleep. This example tracks a cat through one cycle of its basic rest-activity cycle (BRAC). Such cyclic processes are ubiquitous among living systems.

Sleep and wakefulness are complementary phases of the most salient aspects of the brain's endogenous circadian rhythm, or biological clock. Temporal isolation studies have determined the biological clock in humans to be slightly longer than 24 hours. Several features of sleep are regulated by the circadian system, including sleep onset and offset, depth of sleep, and rapid eye movement (REM) sleep intensity and propensity. In the presence of adequate temporal cues (for example, sunlight, noise, social interactions, and alarm clocks), the internal clock keeps good time, regulating a host of physiological and behavioral processes. *See* BIOLOGICAL CLOCKS; NORADRENERGIC SYSTEM.

Sleep is not a uniform state of passive withdrawal from the world, as scientists thought until the twentieth century. In fact, a version of the BRAC continues during sleep, showing a periodicity of approximately 90 minutes. There are two distinct kinds of sleep: a quiet phase (also known as quiet sleep or QS, slow-wave sleep) and an active phase (also known as active sleep or AS, REM sleep, paradoxical sleep), which are distinguished by many differences in biochemistry, physiology, psychology, and behavior. Recordings of electrical activity changes of the brain (electroencephalogram or EEG), eye movements (electrooculogram or EOG), and chin muscle tone (electromyogram or EMG) are used to define the various stages and substages of sleep (**Fig. 1**). *See* ELECTROENCEPHALOGRAPHY.

Sleep cycle. If sleepy enough, most people can fall asleep under almost any condition even while sitting up or being subjected to loud noise, bright lights, and painful electric shocks. However, humans usually lie down and decrease environmental stimulation when preparing to go to sleep. After lying in bed for a few minutes in a quiet, dark room, drowsiness usually sets in. The subjective sensation of drowsiness can be objectively indexed by a corresponding change in brain waves (EEG activity): formerly continuous alpha rhythms (**Fig. 2***a*) gradually break up into progressively shorter trains of regular alpha waves and are replaced by low-voltage mixed-frequency EEG activity. When less than half of an epoch [usually the staging epoch is the 20–30 seconds it takes to fill one page of polygraph (sleep recording) paper] is occupied by continuous alpha rhythm, sleep onset is considered to have occurred and stage 1 sleep is scored (Fig. 2*b*). At this stage, the EOG usually reveals slowly drifting eye movements (SEMs) and muscle tone might or might not decrease. Awakenings at this point frequently yield reports of hypnagogic (leading into sleep) imagery, which can often be extremely vivid and bizarre.

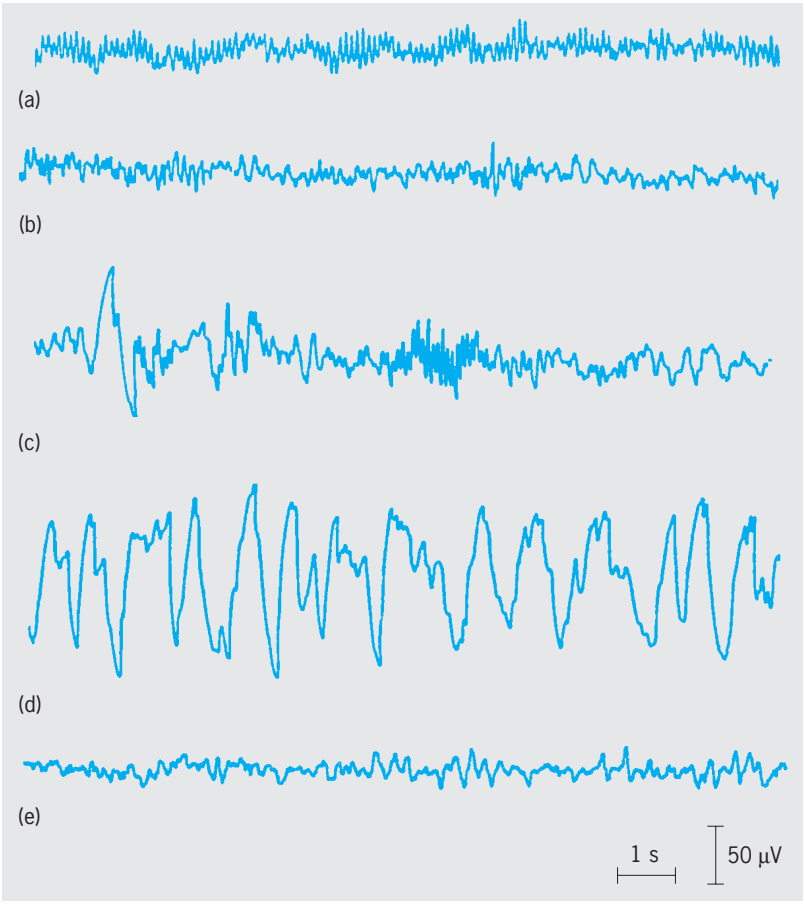


Fig. 2. Human EEG associated with different stages of sleep and wakefulness. (*a*) Relaxed wakefulness (eyes shut) shows rhythmic 8-12-Hz alpha waves. (*b*) Stage 1 non-REM sleep shows mixed frequencies, especially 3-7-Hz theta waves. (*c*) Stage 2 non-REM sleep shows 12–14-Hz sleep spindles and K-complexes. (*d*) Delta sleep shows large-amplitude (>75 μV) 0.5-2-Hz delta waves. (*e*) REM sleep shows low-amplitude, mixed frequencies with sawtooth waves.

Stage 1 is a very light stage of sleep described by most subjects as "drowsing" or "drifting off to sleep." Normally, it lasts only a few minutes before further EEG changes occur, defining another sleep stage. It is at this point that startlelike muscle jerks known as hypnic myoclonias or hypnic jerks occasionally briefly interrupt sleep. As the subject descends deeper into sleep, the EEG of stage 2 sleep is marked by the appearance of relatively high-amplitude slow waves called K-complexes as well as 12–14-Hz rhythms called sleep spindles (Fig. 2*c*). The EOG would generally indicate little eye movement activity, and the EMG would show somewhat decreased muscle tone. Reports of mental activity from this stage of sleep are likely to be less bizarre and more realistic than those from stage 1. However, light sleepers sometimes report lengthy and vivid dreams upon awakening from stage 2 sleep, especially late in the sleep cycle.

After several minutes in stage 2, high-amplitude slow waves (delta waves) gradually begin to appear in the EEG. When at least 20% of an epoch is occupied by these (1-2-Hz) delta waves, stage 3 is defined. Usually this slow-wave activity continues to increase until it completely dominates the appearance of the EEG. When the proportion of delta EEG activity exceeds 50% of an epoch, the criterion for the deepest stage of sleep, stage 4, is met. During stages 3 and 4, often collectively referred to as delta sleep (Fig. 2*d*), the EOG shows few genuine eye movements but is obscured by the high-amplitude delta waves. Muscle tone is normally low, although it can be remarkably high, as when sleepwalking or sleep-talking occurs. Recall of mental activity on arousal from delta sleep is generally very poor and fragmentary and is more thoughtlike than dreamlike. It should be noted that cognitive functioning immediately after abrupt wakening from sleep is likely to carry over some of the characteristics of the preceding sleep state. This phenomenon, known as sleep inertia, can be used as an experimental tool for studying (by inference) cognition during different stages of sleep. For example, studies have shown that stories told in response to thematic apperception test cards (a psychological test used to uncover the dynamics of personality) immediately after awakening from REM sleep are more creative than those told right after non-REM sleep. The combination of confusion, disorientation, cognitive and motor impairment, and amnesia sometimes following abrupt awakening from delta sleep is justifiably termed sleep drunkenness and can be genuinely life-threatening for people such as navy fighter pilots on alert who must fly their highly complicated supersonic jets off an aircraft carrier in the middle of the night within 5 minutes of being awakened.

After about 90 minutes, the progression of sleep stages is reversed, back through stage 3 and stage 2 to stage 1 again. But now the EMG shows virtually no activity at all (Fig. 2*e*), indicating that muscle tone has reached its lowest possible level, and the EOG discloses the occurrence of rapid eye movements—at first only a few at a time, but later in dramatic profusion. This is REM (or active) sleep. Breathing rate and heart rate become more rapid and irregular, and both males and females show signs of sexual arousal (for example, erections and increased vaginal blood flow). Brain metabolic rate increases to levels that typically exceed the waking state average. This state of intense brain activation is normally accompanied by experiences that seem vividly real while they last, but often vanish within seconds of waking. When people are abruptly awakened from REM sleep, 80–90% of the time they recall vivid and sometimes extremely detailed dreams.

While all this activity is happening in the brain, the body remains almost completely still (except for small twitches), because it is temporarily paralyzed during REM sleep to prevent dreams from being acted out. The brainstem system that causes the paralysis of REM sleep does not always inactivate immediately upon awakening. The resulting experience, known as sleep paralysis, can be terrifying, but it is quite harmless and normal if it occurs only occasionally. However, frequent sleep paralysis can be a symptom of a disorder of REM sleep called narcolepsy. *See* SLEEP DISORDERS.

After a REM period lasting perhaps 5 to 15 minutes, a young adult will typically go back through

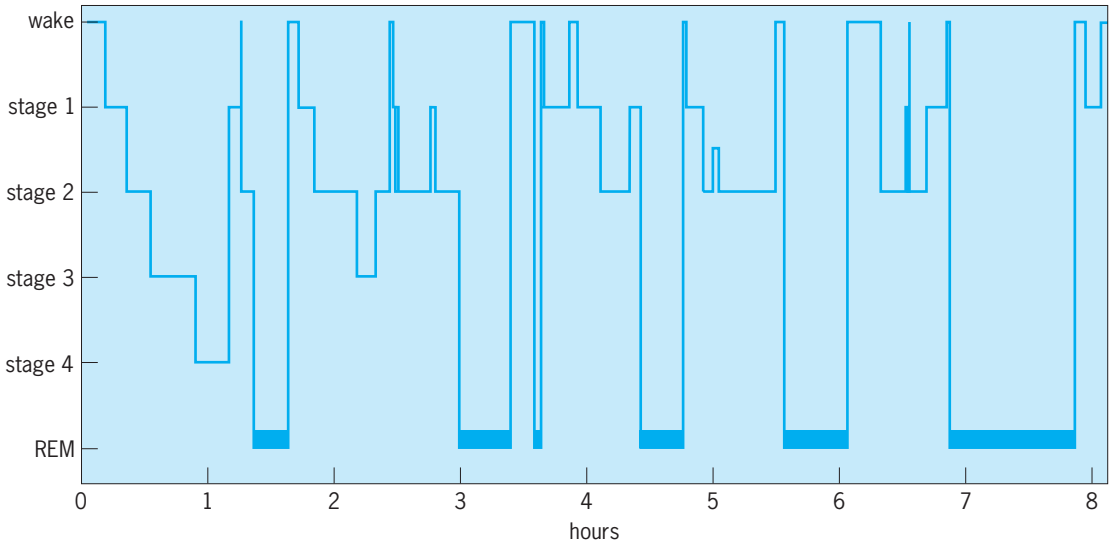


Fig. 3. Sleep histogram showing sleep cycles during a typical night. Note the increasing amount of REM sleep as the night progresses.

the preceding cycle stages, dreaming vividly three or four more times during the remainder of the night (**Fig. 3**), with two major modifications. First, decreasing amounts of slow-wave EEG activity (stages 3 and 4 or delta sleep) occur in each successive cycle. Later in the night, after perhaps the second or third REM period, no delta sleep appears on the EEG at all, only non-REM, stage-2, and REM sleep. Second, as the night proceeds, successive REM periods tend to increase in length, up to a point. While the first REM period commonly lasts less than 10 minutes, later REM periods often last 30 to 40 minutes, and even an hour or more is not exceptionally uncommon late in the sleep cycle. At the same time that the REM periods are getting longer, the intervals between them tend to decrease in length from the approximately 90 minutes characteristic of the first part of the night to as little as 20 or 30 minutes in the late morning. The fact that for humans most REM occurs in the last portion of the sleep cycle as dawn approaches suggests that REM serves a function related to preparation for waking activity.

Finally, after four or five periods of dreaming sleep, the sleeper wakes up (for perhaps the tenth time during the night) and gets up for the day. It may be difficult to believe that brief awakenings occur this frequently during an average night; however, they are promptly forgotten. This retrograde amnesia is a normal feature of sleep: information in short-term memory at sleep onset is usually not transferred into more permanent storage.

Measuring sleepiness. Probably the most obvious relationship between sleep and wakefulness is the fact that variations in the amount of sleep received predict subsequent variations in alertness the following day or days. Too little sleep one night is likely to result in increased sleepiness and decreased alertness during the following day. Variations in alertness/sleepiness form a continuum of states ranging from hyperexcitation and seizure at one extreme to coma and death at the other. Abnormal or pathological states of excessive activation are accompanied by restless hypervigilance, distractibility, agitation, insomnia, and mania. Optimal levels of activation are accompanied by feelings of energy, interest, and motivation to act. Suboptimal activation is accompanied by decreases in such feelings and motivations, along with varying levels of sleepiness, a specific drive motivating rest, inaction, and sleep. Whether sleepiness is experienced as pleasant or unpleasant depends on whether conditions allow consummation of the desire to sleep.

There are various behavioral indications of sleepiness (for example, yawning, head nodding, eye closure, and performance deficits), as well as subjective self-evaluation scales. For example, the Stanford Sleepiness Scale consists of statements that describe seven states of sleepiness/alertness varying from "wide awake" (1), through "somewhat foggy, let down" (4), to "no longer fighting sleep, sleep onset soon, having dreamlike thoughts" (7). (Of course, "asleep" is off the scale.) Subjective scales are quick, cost-effective, and easy to administer but offer relatively low reliability compared with objective measures.

The "gold standard" for quantifying sleepiness is the Multiple Sleep Latency Test (MSLT). It is based on the assumption that degrees of sleepiness are best measured by the speed with which a person falls asleep under standard conditions. Typically, the MSLT is carried out in a sleep laboratory with subjects wired for sleep recordings (EEG, EOG, and EMG). The MSLT is typically performed in five sessions, starting at 10 a.m. and after every 2 hours thereafter until 6 p.m. Care is taken to ensure that the subject is fully comfortable during each session. After taking care of any requirements for comfort (such as a drink of water and going to the bathroom) in time for the scheduled session, the subject enters a dark, quiet bedroom and the electrodes placed on the

subject's body are plugged into the polysomnograph. The subject lies on the bed, and at the scheduled start time is told to close his or her eyes and try to go sleep. Usually, prominent alpha waves are visible in the EEG at first. Sleep onset is scored at the abrupt disappearance of alpha waves, and sleep latency is measured as the time from eyes-shut to sleep onset. Immediately at sleep onset, the subject is promptly awakened and required to get out of bed, preventing the accumulation of more than a few seconds of sleep per trial. If the subject fails to fall asleep within 20 minutes, the trial ends and the sleep latency is scored as the maximum, 20 minutes.

The results from the multiple tests are averaged to provide a single measure of daytime alertness. MSLT scores greater than 15 minutes merit the label "Good Alertness"; scores more than 10 but less than 15 minutes, "Can Get By"; greater than 5 but less than 10 minutes, "Borderline"; and scores less than 5 minutes, "Twilight Zone" because memory and clarity of thinking are usually substantially impaired with this degree of sleepiness.

The quickest method of rapidly increasing sleepiness is total sleep deprivation. An MSLT study showed that young adults were in the Twilight Zone after a single night of total sleep deprivation. After two 9-hour recovery nights of sleep following two nights of total sleep deprivation, they were out of the Twilight Zone in only half the day's MSLT sessions. The fact that two full nights of sleep were insufficient to return alertness to full wakefulness suggests that a sleep debt has been incurred by the total sleep deprivation. But total sleep deprivation is not the usual source of increased daytime sleep tendency, because even partial sleep deprivation has cumulative effects. For example, after three baseline nights with 9 hours in bed per night, subjects were restricted to 5 hours in bed per night for seven consecutive nights, and the MSLT was administered on each subsequent day. Subjects showed a progressive increase in sleepiness for every night of the study. Subjects' scores dropped from Good Alertness to Can Get By after only one night of partial sleep deprivation. By night 5 of partial sleep deprivation, they reached Borderline and would have been in the Twilight Zone by night 10 if the study had continued.

The MSLT also reveals that it is typical that sleepiness reaches a peak in midafternoon. A person can be very sleepy at 4 p.m. but by 6 or 8 p.m. feel quite alert–without having received any sleep in the meantime. This is due to another factor determining alertness/sleepiness in addition to immediately prior sleep or wakefulness, namely a circadian rhythm. According to the opponent process model, there are two independent processes involved which together produce the circadian process of sleep and wakefulness. One process, whose function is to induce and maintain sleep, is called sleep homeostasis. The other process, whose function is to induce and maintain alertness, is called clock-dependent alerting.

A variation on the sleep deprivation technique is selective sleep deprivation. By waking subjects whenever they show a particular sleep stage, it is theoretically possible to determine the effects of that particular sleep stage, usually REM sleep. REM deprivation was pioneered by W. C. Dement, and the early studies produced dramatic demonstrations of the need for REM sleep. Later, more carefully controlled studies found rather more modest effects of REM deprivation on personality, cognitive function, and motivation. Probably the most consistent effect is an intensification of drives and a lowering of thresholds for cortical stimulation. What manifests as hypersexuality and hyperphagia in cats shows itself in humans as increased irritability and perhaps some tie-loosening. In fact, it is probably the threshold-lowering effect of REM deprivation that causes a single night of REM deprivation to significantly improve mood in depressed patients. *See* AFFECTIVE DISORDERS.

Both total and selective quiet-sleep deprivation produces corresponding increases of sleep-drive and intensification of delta wave activity (Fig. 2*d*). After a night of total sleep deprivation, the first few hours of subsequent sleep will be almost exclusively non-REM sleep, suggesting that quiet sleep may be even more functionally important than REM sleep.

Evolution and function of sleep. What selective forces were responsible for the evolution of sleep? The major source of daily change in local conditions on Earth derives from day and night. It is a fact of life that most organisms are adapted to either the dark or light phase of the cycle. Therefore, a biological process that limits activity to the phase of the cycle to which the organism is adapted will enhance survival. Most likely sleep developed out of the rest phase of the BRAC, allowing organisms to minimize interactions with the world during the less favorable phase, while engaging in a variety of "off-line" internal maintenance operations, including growth, repair, synthesis of biochemicals consumed during waking activity, and energy conservation, as well as memory consolidation. The fact that different species have many differences in sleep structure, process, and function fits with the idea that sleep serves the specific adaptive needs of each species. For example, quiet sleep and active sleep have different evolutionary histories and serve different biological functions. Quiet sleep appears to be an older form of sleep with simpler and more universal functions related to energy conservation, growth, and restoration. Active sleep is a mammalian invention with functions that appear to be related to specifically mammalian needs such as live birth. The portion of total sleep composed of REM is at its highest level perinatally: newborn humans spend 8 hours per day in REM sleep, with even more time during the last 6 weeks before birth. The time of maximal REM corresponds to the time of maximal growth of the brain. A number of theorists, including H. Roffwarg and M. Jouvet, suggest that this is not a coincidence but points to the main evolutionary function of REM: to provide a source of endogenous stimulation supporting the unfolding of genetic programming and self- organization of the brain. REM time decreases continuously until puberty, when it reaches the 2 hours or less it maintains throughout

the remainder of life. The fact that REM time does not decrease to zero after the full development of the nervous system suggests secondary adaptive advantages afforded by REM during adulthood, which may include facilitation of difficult learning and preparation of the brain for optimal functioning on arousal. This latter function would account for the unequal distribution of REM throughout the night, with most REM closest to the end of the night. The amount of time spent asleep by different species, and especially the amount of time spent in the deep stage of REM, varies widely as a function of lifestyle. For example, predators sleep deeply and long, while prey species sleep little with still less REM.

Both phases of sleep appear to serve a variety of distinct functions related to learning and memory. For example, quiet sleep has been shown to enhance consolidation of declarative memory, while active sleep enhances consolidation of procedural memory.

As mentioned above, the delta activity of NREM sleep is homeostatically regulated, increasing with wakefulness and decreasing during sleep. In 2004, R. Huber, G. Tononi, and colleagues found local enhancement of delta activity in specific brain areas associated with a particular motor-learning task, providing strong support for a cellular need for sleep, as well as a specific role in learning consolidation.

A number of studies have recently established that performance improvements on a variety of perceptual, motor, and cognitive tasks following a night of sleep are specifically due to sleep rather than, say, the passage of time.

Although the modern understanding of sleep and dreaming has had the benefit of half a century of scientific research, there are still no simple and conclusive answers to the questions of why we sleep. The more we learn, the more complex sleep seems.

Dreams. From the biological perspective, the basic task of the brain is to predict and control the organism's actions and regulate those actions to achieve optimal outcomes (in terms of survival and reproduction). To accomplish this task, the brain in some sense internally "models" the world. The waking brain bases the features of its world model primarily on the current information received from the senses and secondarily on expectations derived from past experience.

In contrast, the sleeping brain acquires little information from the senses. Therefore in sleep, the primary sources of information available to the brain are the current motivational state of the organism (for example, current concerns, fears, hunger, and desires) and past experience (for example, both species-specific instincts and personal memories of recent and associated past experiences of potential relevance). According to this theory, dreams result from brains using internal information to create a simulation of the external world, in a manner directly parallel to the process of waking perception, minus most sensory input. *See* PERCEPTION.

Dreaming experience is commonly viewed as qualitatively distinct from waking experience. Dreams are often believed to be characterized by lack of reflection and an inability to act deliberately and with intention. However, this view has not been based on equivalent measurements of waking and dreaming state experiences. To achieve equivalence, it is necessary to evaluate waking experience retrospectively, in the same way that dreams are evaluated. In 1995, S. LaBerge et al. published the results of a study examining cognition in dreaming and waking by sampling recall of recent experiences from both conditions and collecting responses to questions pertaining to thought processes during the period recalled. Each of 167 subjects contributed detailed reports and questionnaires from a randomly selected sample of waking experience and a corresponding sample of dreaming experience. After recording each experience, the subjects answered a series of questions about the experience described, assessing whether or not they had engaged in a number of cognitive and other activities. These included deliberate choice between alternatives, reflection, sudden distraction of attention, focused intention, self-consciousness, and emotion. Differences appeared between waking and dreaming samples for several variables: 49% of subjects reported deliberate choice from dreaming versus 74% from waking; 41% of subjects reported public self-consciousness from dreams versus 30% from waking; and 86% reported emotion from dreams versus 74% from waking. It is notable that significant differences between dreaming and waking were not evident for other cognitive activities, and none of the measured cognitive functions were absent or rare in dreams. In terms of the typical, waking and dreaming experiences have much in common, with dreams being characterized by somewhat more emotionality and somewhat less choice.

However, less typical examples reveal a wider range of qualities of consciousness in dreaming than in the usual waking state. For example, at one end of the spectrum, dreams at times exhibit cognitive and perceptual errors similar to those produced by brain damage. However, dream bizarreness also shows the creative recombinatory potential of the dreaming brain. Moreover, as illustrated by lucid dreams, dreaming consciousness can also be as volitional and rational as waking consciousness. *See* CONSCIOUSNESS.

Cognizant or lucid dreaming. Typically, people are not explicitly aware of the fact that they are dreaming while they are dreaming; however, at times a remarkable exception occurs. During such lucid dreams it is possible to freely remember the circumstances of waking life, to think clearly, and to act deliberately upon reflection or in accordance with plans decided upon before sleep, all while experiencing a dream world that seems vividly real.

Although accounts of lucid dreaming go at least as far back as Aristotle, until recently dream reports of this sort were received with considerable skepticism. In the absence of objective proof, sleep researchers doubted that the dreaming brain was capable of such a high degree of mental functioning and consciousness. A new technique involving eye movement signals, developed independently by

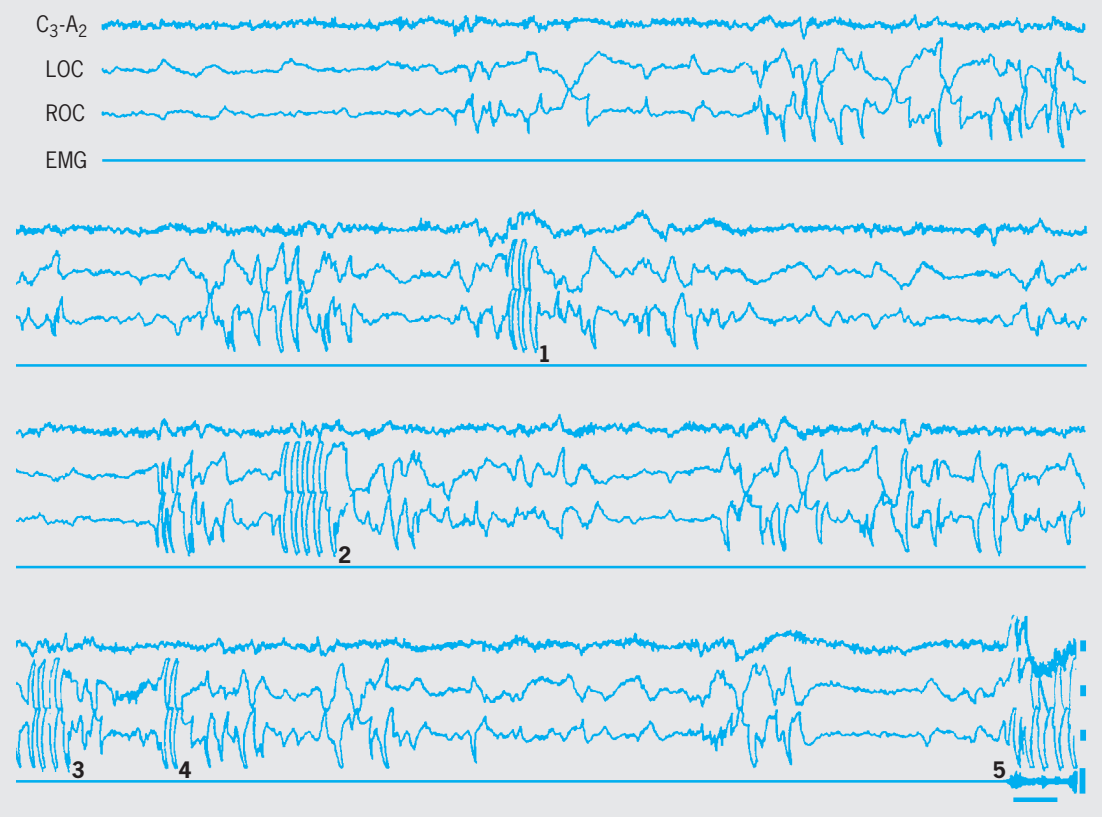


Fig. 4. Voluntary eye movement signaling marks lucid dreaming during uninterrupted REM sleep. Four channels of physiological data [central EEG (C3-A2), left and right eye movements (LOC and ROC), and chin muscle tone (EMG)] from the last 8 minutes of a 30-minute REM period are shown. Upon awakening, the subject reported having made five eye movement signals (labeled 1–5). The first signal (1, LRLR) marked the onset of lucidity. Emotional sweating artifacts can be observed in the EEG at this point. During the following 90 seconds the subject "flew about," exploring his dream world until he believed he had awakened, at which point he made the signal for awakening (2, LRLRLRLR). After another 90 seconds, the subject realized that he was still dreaming and signaled (3) with three pairs of eye movements. Realizing that this was too many, he correctly signaled with two pairs (4). Finally, upon awakening 100 seconds later, he signaled appropriately (5, LRLRLRLR). Calibrations are 50 μV and 5 seconds.

researchers in the United States and in England in the late 1970s, proved the reality of lucid dreaming. The technique was based on earlier studies which found that the directions of eye movements recorded during REM sleep sometimes exactly corresponded to the directions in which subjects reported they had been looking in their dreams. It was reasoned that if lucid dreamers can in fact act volitionally, they should be able to prove it by making a prearranged eye movement signal marking the exact time they became lucid. Using this approach, researchers verified reports of lucid dreams from five subjects by eye movement signals. All of the signals, and therefore lucid dreams, had occurred during uninterrupted REM sleep (**Fig. 4**). This result has been replicated in a number of independent laboratories around the world.

Lucid dreaming as a method for studying consciousness. The eye movement signaling experiments illustrate an important approach to the study of dreaming consciousness. The attempt to apply rigorous scientific methodology to the study of such phenomena as mental imagery, hallucinations, dreaming, and in general, conscious processes faces a major challenge: The most direct account available of the private events occurring in a person's mind is that individual's subjective report. However, subjective reports are difficult to verify objectively, and introspection is far from being an unbiased or direct process of observation. There are two strategies likely to increase the reliability of subjective reports: (1) the use of highly trained subjects (in the context of dream research, this is best achieved with lucid dreamers) who are skillful reporters of consciousness, and (2) the use of a psychophysiological approach to provide validation of subjective reports by correlating them with physiological measures. *See* HALLUCINATION.

Using these approaches in a series of studies, researchers found that various dreamed experiences (including time estimation, breathing, singing, counting, and sexual activity) produce effects on the dreamer's brain (and to a lesser extent, body) remarkably similar to the physiological effects that are produced by actual experiences of the corresponding events while awake. The results of these studies support the following picture: During REM dreaming, the events that a person consciously experiences (or seems to experience) are the results of patterns of central nervous system activity that produce, in turn, effects on the autonomic nervous system and that are body-isomorphic (similar in appearance) to the effects that would occur if a person actually experienced the corresponding events while awake. This explains, in part, why people regularly mistake

dreams for reality. To the functional systems of neuronal activity that construct the experiential world model of consciousness, dreaming of perceiving or doing something is equivalent to actually perceiving or doing it.

Dream recall. The average person remembers a dream only once or twice a week. Given the fact that every person dreams each night, that leaves at least 95% of most dreams forgotten (in REM alone; assuming five REM periods per night with perhaps two dreams per REM period). A variety of theories have been put forward suggesting often fanciful explanations of why dreams are so easily forgotten, ranging from Sigmund Freud's belief that dreams are repressed because they contain so much taboo dream thought, to F. Crick's view that the content of dreams is what the brain is trying to unlearn, and therefore ought not to be remembered. Of course, standard memory theory explains much of what is remembered of dreams and what is not, which aspects of dream content are more readily recalled and which aspects are not, and so on. *See* MEMORY; PSYCHOANALYSIS.

But why are dreams so much more difficult to recall than waking experiences? Evolution may provide the answer. As S. LaBerge reasoned in 1985, humans learn that dreams are distinct from other experiences by talking with other humans. Nonspeaking animals, however, have no way to tell each other how to distinguish dreams from reality. Explicit dream recall would thus be disorienting and maladaptive for all the nonlinguistic mammals having REM sleep during the 140 million years since REM evolved. The fact that explicit dream recall obviously is maladaptive for cats, dogs, bats, whales, and all of the rest of the mammalian dreamers except humans very likely explains why dreams are difficult for humans to recall. The dreams may be so, according to this view, because of natural selection. Like all mammals, humans and their ancestors might have been protected from dangerous confusion by the evolution of mechanisms that made forgetting dreams the normal course of affairs. *See* ORGANIC EVOLUTION. Stephen LaBerge

Bibliography. K. M. Fenn, H. C. Nusbaum, and D. Margoliash, Consolidation during sleep of perceptual learning of spoken language, *Nature*, 425:614–616, 2003; P. Fenwick et al., Lucid dreaming: Correspondence between dreamed and actual events in one subject during REM sleep, *Biol. Psychol.*, 18:243–252, 1984; S. Fischer et al., Sleep forms memory for finger skills. *Proc. Natl. Acad. Sci. USA*, 99:11987–11991, 2002; R. Huber et al., Local sleep and learning, *Nature*, 430:78–81, 2004; S. LaBerge, *Lucid Dreaming*, J. P. Tarcher, Los Angeles, 1985; S. LaBerge, Lucid dreaming: Psychophysiological studies of consciousness during REM sleep, in R. R. Bootsen, J. F. Kihlstrom, and D. L. Schacter (eds.), *Sleep and Cognition*, pp. 109–126, APA Press, Washington, DC, 1990; S. LaBerge, T. Kahan, and L. Levitan, Cognition in dreaming and waking, *Sleep Res.*, 24A:239, 1995; N. Malcolm, *Dreaming*, Routledge, London, 1959; S. C. Mednick et al., The restorative effect of naps on perceptual deterioration, *Nat. Neurosci.*, 5:677–681, 2002; A. Rechtschaffen, The single-mindedness and isolation of dreams, *Sleep*, 1:97–109, 1978; H. Roffwarg et al., Dream imagery: Relationship to rapid eye movements of sleep, *Arch. Gen. Psychiat.*, 7:235–238, 1962; C. Smith, Sleep states and memory processes, *Behav. Brain Res.*, 69:137–145, 1995; R. Stickgold, L. James, and J. A. Hobson, Visual discrimination learning requires sleep after training. *Nat. Neurosci.*, 3:1237–1238, 2000.

Sleep disorders

A reversible state during which the individual's voluntary functions are suspended but the involuntary functions, such as circulation and respiration, are uninterrupted; the sleeping subject assumes a characteristic posture with relative immobility and decreased responses to external stimuli.

The sleep state can be divided into nonrapid eye movement (NREM) and rapid eye movement (REM) sleep; NREM sleep, in turn, is subdivided into four stages based on electroencephalographic criteria. The first REM in a normal adult occurs 60–90 min after sleep onset, and there are usually four or five NREM-REM cycles, each lasting for 90–110 min. Most dreams occur during REM sleep. The sleep-wake pattern for humans follows a circadian rhythm. An average adult needs approximately 7–8 h of sleep, but in elderly people there are frequent awakenings.

Clinical disorders. The latest International Classification of Sleep Disorders (ICSD) divides sleep disorders into eight broad categories: (1) insomnias; (2) sleep-related breathing disorders, including obstructive sleep apnea syndrome; (3) hypersomnias of central origin (including narcolepsy) that are not secondary to sleep-related breathing disorders, circadian rhythm sleep disorders, or other causes of disturbed night sleep; (4) circadian rhythm sleep disorders; (5) parasomnias (abnormal movements or behavior intruding into sleep, such as sleepwalking, sleep terrors, confusional arousals, REM behavior disorders, nightmare, sleep enuresis); (6) sleep-related movement disorders, such as restless legs syndrome; (7) isolated symptoms (for example, long sleeper, short sleeper, snoring, sleep talking, etc.; these symptoms belong to the borderline between normal and abnormal sleep or are part of the continuum of normal to abnormal events in sleep), including apparently normal variants and unresolved issues due to insufficient information; and (8) other sleep disorders, which do not conform to the listing outlined above or may have features overlapping some other conditions listed in other categories. In addition, ICSD includes Appendix A, which includes among other conditions sleep-related epilepsy; and Appendix B, which includes psychiatric and behavioral disorders causing sleep dysfunction.

Most common complaints. Excessive daytime sleepiness, insomnia (sleeplessness), and abnormal movements and behavior during sleep are the major sleep complaints. Individuals with hypersomnia complain of excessive daytime sleepiness. An individual with insomnia may have difficulty initiating or maintaining

sleep; repeated awakenings or early morning awakenings; or daytime fatigue and impairment of performance. The commonest cause of insomnia is psychiatric or psychophysiologic disorder (for example, depression, anxiety, or stress), but other causes include medical disorders or pain. Circadian-rhythm sleep disorders may be associated with either insomnia or hypersomnia.

Obstructive sleep apnea syndrome. There are two forms of sleep apnea. In central apnea, both the airflow at the mouth and nose and the effort by the diaphragm decrease. In obstructive apnea, the airflow stops but the effort by the diaphragm continues. Obstructive sleep apnea syndrome is common in middle-aged and elderly obese men. The individual stops breathing for many seconds during sleep and wakes up repeatedly with a loud snore. Repeated apneas associated with reduction of oxygenation in the blood may cause serious complications, such as high blood pressure and irregularities and/or failure of the heart. Obstructive sleep apnea syndrome results from excessive relaxation of the upper airway muscles during sleep, coupled with an unknown dysfunction of the respiratory neurons.

Narcolepsy. Narcoleptic sleep attacks usually begin in individuals between the ages of 15 and 25. Narcolepsy is characterized by an irresistible desire to sleep, and the attacks may last 15–30 min. Most individuals also have cataplexy, during which there is transient loss of muscle tone that causes the individual to fall to the ground or slump the head forward for a few seconds. These attacks are often triggered by an emotional outburst. Other symptoms may include fearful dreams or feeling of loss of power at sleep onset or offset or nocturnal sleep disturbance. Narcolepsy cannot be cured.

Restless legs syndrome. This syndrome is another lifelong condition occurring during middle age. It is characterized by the urge to move the legs, usually accompanied by intensely disagreeable feelings in the legs at rest and repose, predominantly in the evening, with partial or total relief from these symptoms sometimes occurring. Most individuals with this problem also have periodic limb movements in sleep.

Diagnosis and treatment. Overnight polysomnography and multiple sleep latency tests are both important for sleep disorders. Overnight polysomnography involves recording of brain waves, muscle activities, eye movements, heart activity, airflow at the nose and mouth, respiratory effort, and oxygen saturation. Polysomnography is needed for individuals with excessive daytime sleepiness, narcolepsy, parasomnias, restless legs–periodic limb movements, and nocturnal seizures. The multiple sleep latency test is important in documenting pathologic sleepiness and diagnosing narcolepsy. The test records brain waves, muscle activities, and eye movements.

Treatment of obstructive sleep apnea syndrome consists of avoidance of sedatives or alcohol consumption, weight loss, and use of continuous positive airway pressure to open the upper airway passage so that obstructive apneas can be eliminated. The optimal continuous positive airway pressure is determined during overnight polysomnography. Narcolepsy is treated by stimulants and short naps, while cataplexy is treated with tricyclic antidepressants, selective serotonin reuptake inhibitors, and sodium oxybate. Chronic insomnia can best be treated by sleep hygiene and behavioral modification. A short-acting nonbenzodiazepine (for example, zolpidem, zaleplon, eszopiclone) or a short: or intermediate-acting benzodiazepine hypnotic may be used for transient insomnia. Most parasomnias do not require special treatment. However, psychotherapy may be helpful in some cases. Circadian rhythm disorders may be treated by bright light and chronotherapy. *See* BIOLOGICAL CLOCKS; SLEEP AND DREAMING.

Sudhansu Chokroverty

Bibliography. S. Chokroverty (ed.), *Sleep Disorders Medicine: Basic Science, Technical Considerations and Clinical Aspects*, 2d ed., 1999; *The International Classification of Sleep Disorders, Diagnosis, and Coding Manual*, 2d ed., American Academy of Sleep Medicine, 2005; M. Kryger, T. Roth, and W. Dement (eds.), *Principles and Practice of Sleep Medicine*, 4th ed., 2005.

Slider-crank mechanism

A four-bar linkage with output crank and ground member of infinite length. A slider crank (**Fig. 1**) is most widely used to convert reciprocating to rotary motion (as in an engine) or to convert rotary to reciprocating motion (as in pumps), but it has numerous other applications. Positions at which slider motion reverses are called dead centers. When crank and connecting rod are extended in a straight line and the slider is at its maximum distance from the axis of the crankshaft, the position is top dead center (TDC); when the slider is at its minimum distance from the axis of the crankshaft, the position is bottom dead center (BDC). *See* FOUR-BAR LINKAGE.

For a given crank throw, or radius, the action of the slider approaches simple harmonic motion as the length of the connecting rod is increased. Maximum accelerations occur at reversal of the slider. For constant angular velocity of the crank, slider acceleration at TDC is somewhat greater, and at BDC somewhat less, than accelerations that would occur if the motion were simple harmonic.

Although the idea of combining a crank with a connecting rod is quite old (fifteenth century), the crank was not successfully applied to a steam engine

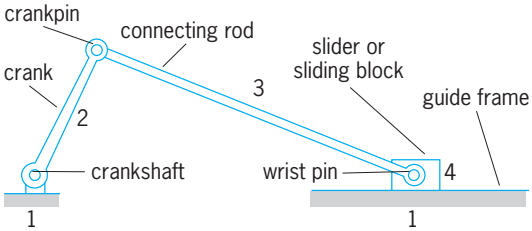


Fig. 1. Principal parts of slider-crank mechanism.

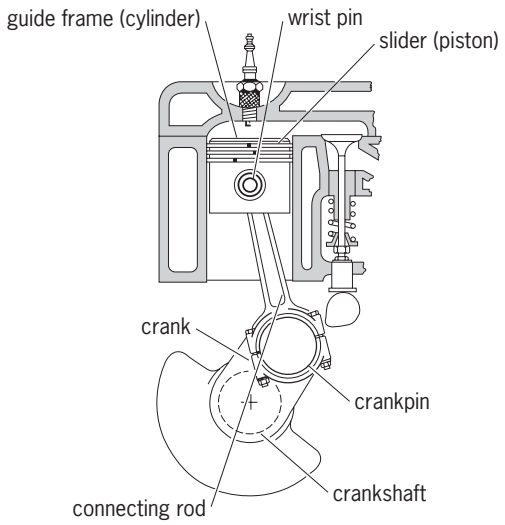


Fig. 2. Slider crank applied to internal combustion engine.

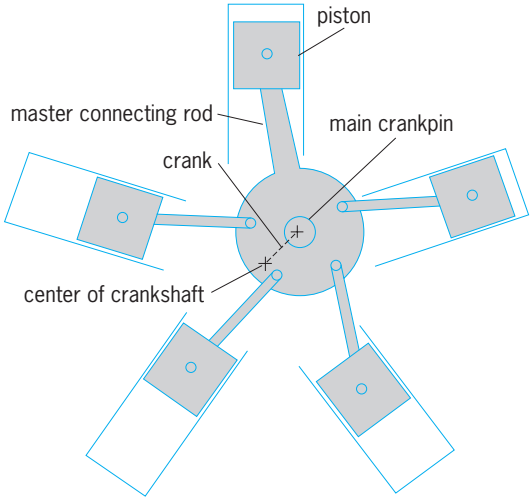


Fig. 3. Multiple-crank mechanism in radial engine.

until 1780; the completion of the linkage to include a slider had to wait for satisfactory lubrication.

Many attempts were made between 1780 and 1830 to produce rotary motion directly and thus eliminate the need for the slider-crank mechanism in prime movers. Hundreds of different rotary engine designs (many of which employed slider-crank linkage in a form not recognized by the inventor) were proposed but, for a prime mover that depends for its operation upon such a fluid as steam or air within a chamber, no arrangement has been found superior to the conventional one described here.

Internal combustion engine. The conventional internal combustion engine employs a piston arrangement in which the piston becomes the slider of the slider-crank mechanism (**Fig. 2**). Although satisfactory for engines of moderate life-span, the reversal of side thrust on the piston twice during each revolution complicates the problem of keeping the piston tight enough in the cylinder to contain the working medium in the combustion space. Because of angularity of the connecting rod, most wear occurs at the lower end of the cylinder. In some large, low-speed engines, a crosshead, differing in arrangement but similar in principle to the crosshead of a steam engine, is used to reduce cylinder wear.

Radial engines for aircraft employ a single master connecting rod to reduce the length of the crankshaft (**Fig. 3**). The master rod, which is connected to the wrist pin in a piston, is part of a conventional slider-crank mechanism. The other pistons are joined by their connecting rods to pins on the master connecting rod.

Reciprocating compressors. To convert rotary motion into reciprocating motion, the slider crank is part of a wide range of machines, typically pumps and compressors. The crankshaft may be driven through belting or by an electric motor. Portable, reciprocating compressors use pistons; stationary compressors generally employ crossheads and guides. *See* COMPRESSOR.

Other applications. Another use of the slider crank is in toggle mechanisms, also called knuckle joints. The driving force is applied at the crankpin (Fig. 1) so that, at TDC, a much larger force is developed at the slider. This is limited by practicality to the forces at which the links and pins yield.

Other forms that the slider crank can take can be studied by inversion (**Fig. 4**). Compared to a

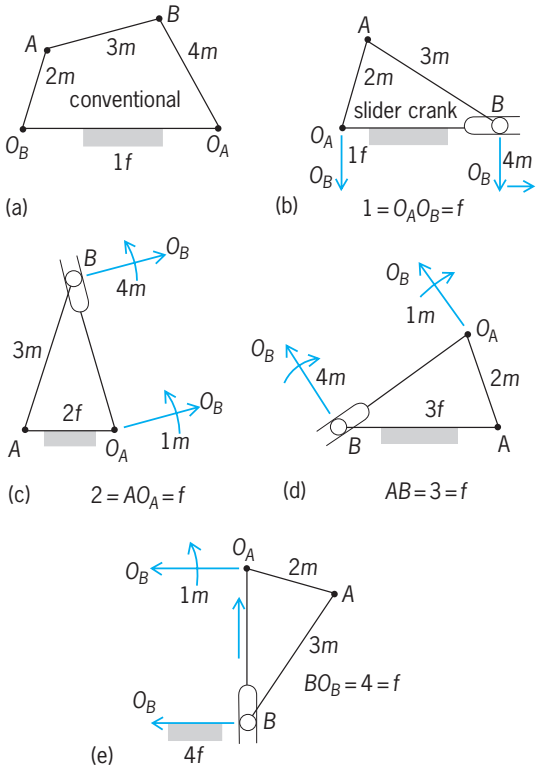


Fig. 4. Inversion of slider-crank mechanism. (*a*) Conventional four-bar linkage of which slider crank is a special case. (*b*) Link 4 made infinitely long, and link 1, which carries the slider guide, held fixed, produce the usual slider-crank mechanism. (*c*) Intermediate link 2 held fixed. (*d*) Link 3, carrying the slider head, held fixed. (*e*) Virtual link 4, to which slider *B* is pinned, held fixed. Links marked *f* are fixed; those marked *m* move.

conventional four-bar linkage (Fig. 4*a*) where link 1 is the fixed member, the slider crank replaces finite moving link 4 with the slider (Fig. 4*b*); in this form link 1 remains the fixed member. Although moving member 4 loses its separate identity, comparison of angles ABO_B in Fig. 4*a* and *b*, and comparison of angles BO_BO_A in the two diagrams, shows the actions of the two mechanisms to involve the same number of links and turning pairs within their closed loops. By successively holding each link fixed, the alternative forms of Fig. 4*c*–*e* are obtained. Douglas P. Adams

Bibliography. H. H. Mabie and C. F. Reinholtz, *Mechanisms and Dynamics of Machinery*, 4th ed., 1987; G. H. Ryder and M. D. Bennett, *Mechanics of Machines*, 2d ed., 1990; J. E. Shigley and J. J. Uiker, *Theory of Machines and Mechanisms*, 2d ed., 1994.

Slip (electricity)

A numerical value used in describing the performance of electrical couplings and induction machines. In an electrical coupling, slip is defined simply as the difference between the speeds of the two rotating members. In an induction motor, slip is a measure of the difference between synchronous speed and shaft speed.

When the stator windings of an induction motor are connected to a suitable alternating voltage supply, they set up a rotating magnetic field within the motor. The speed of rotation of this field is called synchronous speed, and is given by Eq. (1) or Eq. (2),

$$\omega_s = \frac{4\pi f}{p} \quad \text{rad/s} \tag{1}$$

$$n_s = 120\frac{f}{p} \quad \text{rev/min} \tag{2}$$

where f is the line frequency and p is the number of magnetic poles of the field. The number of poles is determined by the design of the windings. In accord with Faraday's voltage law, a magnetic field can induce voltage in a coil only when the flux linking the coil varies with time. If the rotor were to turn at the same speed as the stator field, the flux linkage with the rotor would be constant. No voltages would be induced in the rotor windings, no rotor current would flow, and no torque would be developed. For motor action it is necessary that the rotor windings move backward relative to the magnetic field so that Faraday's law voltages may be induced in them. That is, there must be slip between the rotor and the field. *See* ELECTROMAGNETIC INDUCTION; INDUCTION MOTOR.

The amount of slip may be expressed as the difference between the field and rotor speeds in revolutions per minute or radians per second. However, the slip of an induction motor is most commonly defined as a decimal fraction of synchronous speed, as in Eq. (3) or Eq. (4). Here n is the motor speed in

$$s = \frac{n_s - n}{n_s} \tag{3}$$

$$s = \frac{\omega_s - \omega}{\omega_s} \tag{4}$$

revolutions per minute, ω is its speed in radians per second, and s is the slip, or more properly the per unit slip. Typical full-load values of slip for an induction motor range from 0.02 to 0.15, depending on rotor design. Slip is sometimes expressed in percent of synchronous speed, rather than per unit. If an induction machine is driven faster than synchronous speed, the slip becomes negative, and the machine acts as a generator, forcing energy back into the electrical supply line. *See* ELECTRIC ROTATING MACHINERY. George McPherson, Jr.

Bibliography. Institute of Electrical and Electronics Engineers, *The Authoritative Dictionary of of IEEE Standards Terms (IEEE 100)*, 7th ed., 2000; G. McPherson and R. Laramore, *An Introduction to Electrical Machines and Transformers*, 2d ed., Wiley, 1990.

Slip rings

Electromechanical components which, in combination with brushes, provide a continuous electrical connection between rotating and stationary conductors. Typical applications of slip rings are in electric rotating machinery, synchros, and gyroscopes. Slip rings are also employed in large assemblies where a number of circuits must be established between a rotating device, such as a radar antenna, and stationary equipment.

Electric rotating machines. Slip rings are used in wound-rotor induction motors, synchronous motors and alternators, and rotary converters to connect the rotor to stationary external circuits. These slip rings are usually constructed of steel with the cylindrical outer surface concentric with the axis of rotation. Insulated mountings insulate the rings from the shaft and from each other. Conducting brushes are arranged about the circumference of the slip rings and held in contact with the surface of the rings by spring tension. A typical assembly is shown in **Fig. 1**. Other arrangements, such as concentric slip rings mounted on the face of an insulating disk, may be employed in special cases. Alternating-current

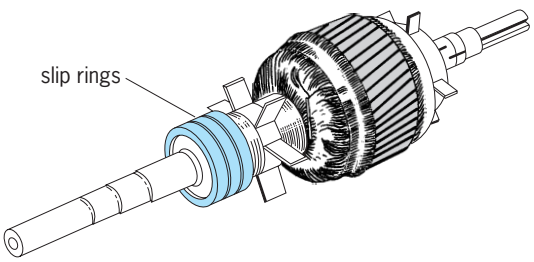


Fig. 1. Rotor of electric rotating machine.

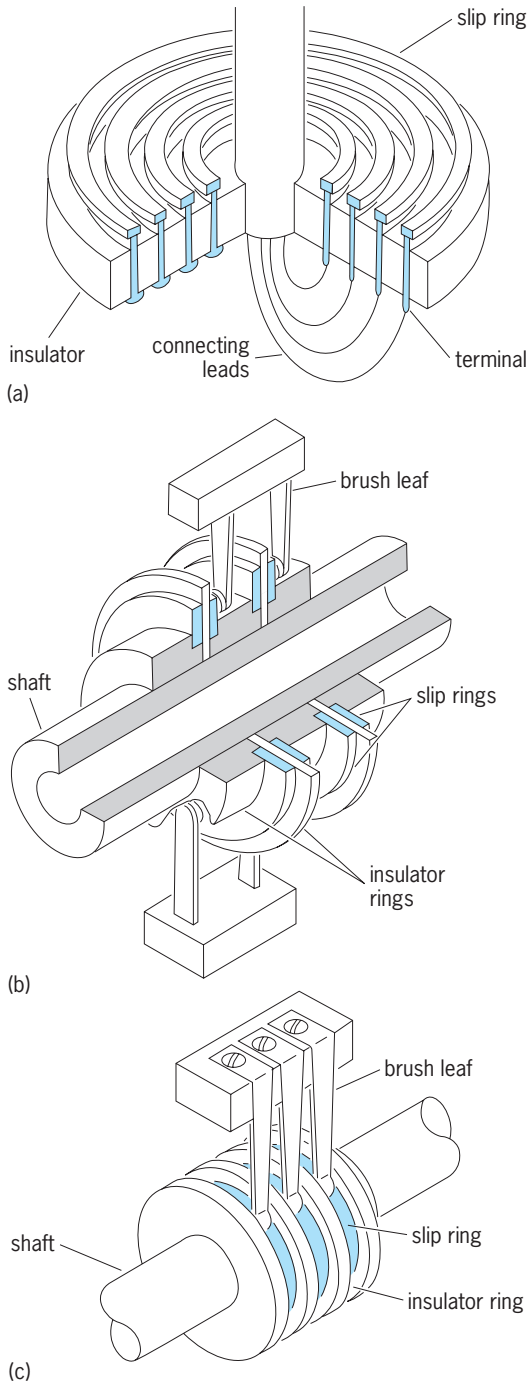


Fig. 2. Slip-ring assembly configurations. (*a*) Concentric-ring. (*b*) Back-to-back. (*c*) Drum.

windings normally require one slip ring per phase except that a single-phase winding requires two slip rings. Two slip rings are required for rotating dc field windings. *See* ELECTRIC ROTATING MACHINERY; GENERATOR; MOTOR.

Arthur B. Eckels

Slip-ring assemblies. These integral mechanical structures contain a plurality of slip rings which, in combination with self-contained brushes, provide continuous electrical connection between electric and electronic equipment mounted on stationary and rotating platforms.

Slip-ring assemblies are designed for a wide range of electric circuits. The same assembly may have circuits for power up to several hundred kilowatts; high voltage to 50 kV; power pulses for radar transmitters, and data signals, including radio frequencies to 100 MHz; strain-gage signals and thermocouple signals. Ring surface speeds range from a few feet per minute to over 15,000 ft/min (75 m/s)

Slip rings for slip-ring assemblies are made usually of coin silver, stamped from hard rolled sheet or cut from drawn tubing, or made of strip silver overlay on copper, formed into a ring, and silver brazed. Fine silver rings may be electroformed onto the insulating material. Surface finish is machined or mill rolled from 4 to 16 microinches (100 to 400 nanometers). Brushes for slip-ring assemblies are graphite combined with copper or silver in proportions suitable for the application, and may be welded or brazed to a spring-temper leaf. Leaf brush pressure is from 1.5 to 3 oz (43 to 85 g).

Large assemblies, such as those used with radar antennas, are fabricated from individual insulators and conducting materials arranged in either the concentric-ring configuration (**Fig. 2***a*) or the back-to-back ring configuration (Fig. 2*b*). Small assemblies, such as those used to transmit signals through gimbals or high-speed turbines, may be fabricated in the back-to-back ring configuration or the drum configuration (Fig. 2*c*).

Other manufacturing methods employ casting the individual rings into filled epoxy resins, electroforming fine silver into grooves machined on filled epoxy resin tubes, or molding the individual rings with electrical grades of thermosetting resins. The casting, electroforming, and molding methods have the advantage of low tolerance buildup, which is important for the synchro sizes.

Background noise for strain-gage signals should not be greater than a few microvolts; intercircuit interference (crosstalk) should not be greater than 70 dB down at 30 MHz, insertion loss at 30 MHz no greater than 0.5 dB, and brush contact resistance approximately 0.005 ohm.

Synchro slip rings are made from fine silver or gold alloy. Brushes are also made from precious-metal alloys, usually in the form of hard-drawn, spring-temper wires. Surface speeds are usually low.

Willard F. Mason

Sloan Digital Sky Survey

A large-scale imaging and redshift survey which, when complete, will map in detail over one quarter of the sky, determining the positions and brightness of over 100 million celestial objects and the redshifts of over a million galaxies and quasars. A primary objective of the survey is to obtain a map of the three-dimensional distribution of galaxies in the universe

Many decades of observational and theoretical research have converged on a unified picture of the nature of the universe. Instrumental in these

advances have been detailed measurements of the expansion rate of the universe by scientists using the Hubble Space Telescope, of the radiation from the cosmic microwave background radiation by the Wilkinson Microwave Anisotropy Probe, and of the distribution of galaxies by several redshift surveys, of which the largest to date have been the Two Degree Field Survey and the Sloan Digital Sky Survey (SDSS). *See* HUBBLE SPACE TELESCOPE; WILKINSON MICROWAVE ANISOTROPY PROBE.

These last projects probe the universe by measuring the redshifts of galaxies, that is, the shift in wavelength of spectral features from their known rest wavelengths to longer wavelengths. Redshifts are caused by the expansion of the universe. Because of this expansion, the farther away from us a galaxy lies the more rapidly it is receding from us. This relation can be turned around, for once the proportionality between distance and redshift (the Hubble "constant" or parameter) has been measured, it can be used together with a galaxy's redshift to give its distance. Thus observations of the positions on the sky and of the redshifts of a sample of galaxies can be used to measure the three-dimensional distribution of galaxies. *See* HUBBLE CONSTANT; REDSHIFT.

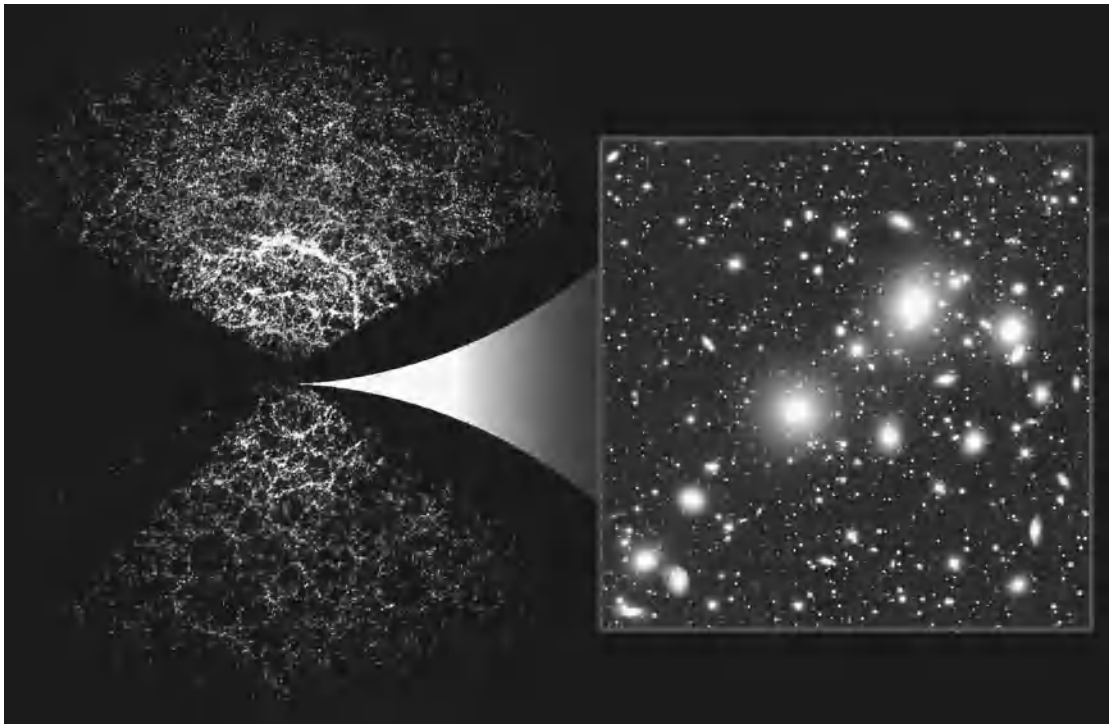
Three-dimensional galaxy distribution. Why is this of interest? Galaxies turn out not to be randomly distributed in space, but their distribution is structured in gravitationally bound clusters, sheets, and filaments. The first three-dimensional maps of significant volumes of the universe were made in the 1980s, and incorporated observations of 1000–2000 galaxies. These maps revealed a frothy, filamentary distribution of galaxies, with the largest feature being a "Great Wall" (in a map made by Margaret Geller, John Huchra, and Valerie de Lapparent) of galaxies more than 600 million light years across and extending to the boundaries of the survey volume, so that it could have been even larger.

Encoded in these maps is the result of the initial mass fluctuations in the early stages of the universe and subsequent evolution due to the expansion of the universe, galaxy formation, and gravity: Material falls into dense regions of space, which thereby accrete more mass as the universe evolves. The exciting results discussed above pointed out the need for surveys of much larger volumes of space, so that a representative sample of the universe could be studied. Current theories of the universe say that, in a statistical sense, it is uniform at all locations and in all directions. The requirement on a redshift survey that can fairly be said to sample the universe is, then, that the volume is large enough to contain many tens of examples of the largest possible structures. It was clear from the early results that to do this would involve measuring redshifts for many hundreds of thousands of galaxies. Galaxies are faint, and accumulating enough light to measure a single redshift can take of order an hour of telescope time. The possibility of conducting surveys large enough to map the local universe in the necessary detail developed with the advent of technology that could measure thousands of redshifts in a single night, and of fast, powerful computers to acquire and reduce the data automatically.

Observing strategy. The Sloan Digital Sky Survey is one of several large surveys recently finished or under way. It is unique in several important ways: It is the largest redshift survey, with more than 400,000 galaxy redshifts measured as of June 2004 and plans to measure a million; and it acquires the imaging data from which the positions of the galaxies can be measured during the same time period as it measures the redshifts.

The reason for this strategy is that the Sloan Digital Sky Survey acquires a much better image of the sky than has previously been available for more than very small areas of sky, so it is both an imaging and a redshift survey. This improved imaging is necessary to understand better the intrinsic properties of each observed galaxy (brightness, dominant stellar population, and so forth). The Sloan Digital Sky Survey images the sky through five filters essentially simultaneously, measuring five brightnesses for each object. The filters cover the ultraviolet, green, red, far-red, and near-infrared regions of the optical spectrum. A star that is very hot, and therefore blue, will be much brighter in the ultraviolet filter than in the far-red filter, and vice versa for a cool, red star. Thus objects can be sorted and selected by their colors automatically. The filters are designed to separate quasars from stars by color, for the Sloan Digital Sky Survey is also a redshift survey of 100,000 quasars. Galaxies can be separated from stars and quasars because they appear to be fuzzy rather than pointlike on images (see **illustration**). The data are acquired by electronic imagers called charge-coupled devices (CCDs), similar to those used in digital cameras. The word digital applies both to cameras and the Sloan Survey because the imager converts light to electrons, which can be directly measured and counted as integers by computers. *See* ASTRONOMICAL IMAGING; CHARGE-COUPLED DEVICES; QUASAR.

The normal way in which redshift surveys are done is to image the sky (and before Sloan Digital Sky Survey, all large-area imaging of the sky had been done by photographic plates), select the objects for which redshifts will be measured, and then measure the redshifts. To do this with the Sloan Digital Sky Survey would have been prohibitively time-consuming and expensive. All optical observing needs clear weather, but imaging needs more: pristine clarity that does not change over the night, and a steady atmosphere, so that the images of stars are as close to pointlike as possible. Even the best sites have such conditions only a small fraction of the time. They do, however, have good atmospheric conditions much more frequently. Such conditions are perfectly adequate for obtaining the spectra that give redshifts. Accordingly, the Sloan Digital Sky Survey follows a dual observing strategy: (1) imaging is done in the best weather; (2) objects are selected from the imaging according to carefully defined criteria by powerful computer software; and their spectra are observed during the less-than-perfect weather. In this way the Sloan Digital Sky Survey efficiently uses all the available



Distribution of galaxies from the Sloan Digital Sky Survey (SDSS). The image on the left shows the redshift-space distribution for galaxies in the SDSS. The dark cones to the left and right are areas of the universe which cannot be observed because they lie behind the obscuring dust of the Milky Way Galaxy. Each point represents one galaxy. The image on the right is a small piece of sky constructed from the SDSS multicolor imaging data. The extended fuzzy objects are galaxies, and the small pointlike objects are stars in the Milky Way Galaxy. (*Courtesy of M. Tegmark and R. Lupton*)

observing time and can be completed in a reasonable time (5–7 years). The spectra are observed with devices that can obtain 640 spectra simultaneously; 7 or 8 sets, or several thousand spectra, can be observed in a single night. To do this, the Sloan Digital Sky Survey uses a dedicated 2.5-m (98-in.) telescope at the Apache Point Observatory, New Mexico. The telescope is equipped with a multi-charge-coupled-device camera and two fiber-fed spectrographs, and observes every possible clear, dark night year-round except for a short maintenance period in the summer. *See* SPECTROGRAPH; TELESCOPE.

Results. The Sloan Digital Sky Survey has obtained several fundamental results. A "Great Wall" of galaxies, 1.4×10^9 light-years long and 80% longer than the "Great Wall" found by Geller, Huchra, and Lapparent, is the largest structure in the universe observed so far. The distribution of galaxies is in excellent agreement with the current cosmological model, in which only 4% of the universe is normal matter, with the rest being dark matter and dark energy. The signature of the formation of galaxies in lumps of dark matter is seen, and comparison of the distribution of the galaxies with that of the cosmic microwave background confirms the presence of dark energy. Cosmology with the Sloan Digital Sky Survey is only just beginning to yield significant information, and many more results, measuring the universe in great detail and with great accuracy, will be forthcoming in the next few years. The current map of the universe, together with a Sloan Digital Sky Survey image of a cluster of nearby galaxies, is shown in the illustration.

The Sloan Digital Sky Survey quasar work has been enormously successful. Quasars are extremely high luminosity objects, thought to be due to the infall of gas into massive black holes. Because they are so luminous, quasars can be seen to great distances, and we observe distant objects as they were a long time ago because of the light travel time. The most distant quasars are seen when the universe was about 5% of its present age. The distant quasars have allowed the discovery of the Gunn-Peterson effect, predicted in the 1960s, whereby the light from distant objects is absorbed by hydrogen if that hydrogen is atomic. This would prevent our seeing the distant universe unless the intervening hydrogen were ionized, that is, the electrons separated from the protons, in which case the atoms could no longer absorb light. The detection of the Gunn-Peterson effect therefore locates the epoch at which the dark ages of the universe ended and the hydrogen was ionized by light from newly formed stars.

A survey as powerful as the Sloan Digital Sky Survey produces results in many other areas. The Sloan Digital Sky Survey has been instrumental in discovering that the Milky Way Galaxy is surrounded by streams of stars from the tidal destruction of small satellite galaxies that used to orbit the Galaxy. The Sloan Digital Sky Survey has discovered many gravitational lenses, in which light from distant quasars is bent by the warping of space by lumps of mass (the lenses) between the Earth and the quasar. This effect is predicted by Einstein's theory of general relativity. The Sloan Digital Sky Survey has discovered the

most massive lens yet found. All masses will distort the light passing them, and this effect has allowed the Sloan Digital Sky Survey to measure the masses of galaxies directly; the more massive the galaxy, the more the light path is bent. The Sloan Digital Sky Survey has discovered the lowest-luminosity galaxy ever found, a satellite of the nearby Andromeda Galaxy. It has discovered large numbers of brown dwarfs, objects intermediate in mass between planets and stars. It has produced the largest sample of small bodies in the solar system with accurate colors, allowing their compositions to be deduced. It can confidently expected that these findings are only the beginning of decades of discovery from the Sloan Digital Sky Survey data. *See* BROWN DWARF; GRAVITATIONAL LENS; MILKY WAY GALAXY.

The data from the Sloan Digital Sky Survey are made freely available to all, with a regular schedule of data releases. As well as being available to professional astronomers in support of their research, they are available for educational purposes and come with a large set of educational tools. *See* COSMOLOGY; GALAXY, EXTERNAL; UNIVERSE. Gillian R. Knapp

Bibliography. B. S. Ryden, *Introduction to Cosmology*, Addison-Wesley, 2003; M. A. Strauss, Reading the blueprints of creation, *Sci. Amer.*, 290(2):42, February 2004; M. Tegmark et al., The three-dimensional power spectrum of galaxies from the Sloan Digital Sky Survey, *Astrophys. J.*, 606:702–740, 2004; D. G. York et al., The Sloan Digital Sky Survey: Technical Summary, *Astron. J.*, 120:1579–1587, 2000.

Slope

The trigonometric tangent of the angle α that a line makes with the x axis as in **Fig. 1**. In **Fig. 2** the slope of a plane curve C at a point P of C is the slope of the line that is tangent to C at P. If $y = f(x)$ is an equation in rectangular coordinates of curve C, the slope of C at $P(x_0, y_0)$ is the value of the derivative $dy/dx = f'(x)$ at P, denoted by $f'(x_0)$, and hence an equation of the nonvertical tangent to C at P is $y - y_0 = f'(x_0)(x - x_0)$. *See* ANALYTIC GEOMETRY; CALCULUS. Leonard M. Blumenthal

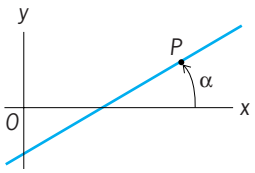


Fig. 1. Slope of a line.

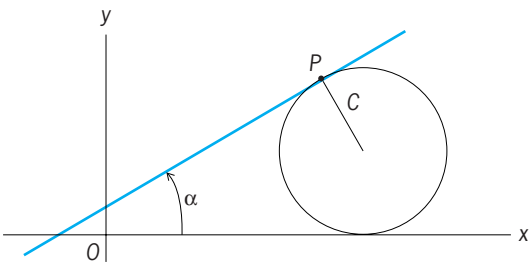


Fig. 2. Slope of a curve.

Sloth

A mammal classified in the order Xenarthra along with anteaters, tamanduas, and armadillos. The sloth differs from all other mammals by having additional articulations (xenarthrales) between their lumbar vertebrae (called xenarthrous vertebrae). They also have a double rather than a single inferior vena cava, the vein that returns blood from the posterior portion of the body to the heart. The dental formula is I 0/0, C 0/0, PM 0/0, M 5/4 × 2, for a total of 18 teeth which grow throughout life. Females have a primitive, divided uterus and a common urinary and genital tract; males have internal testes. *See* DENTITION.

Two extant families of tree sloths (Bradypodidae and Megalonychidae) inhabit the lowland and upland tropical forests of Central and South America.

Bradypodidae. The family Bradypodidae includes a single genus, the three-toed tree sloth (*Bradypus*) [see **illustration**]. This is a slender-bodied animal with a head and body length of 413–700 mm (16–27 in.) and a tail length of 20–90 mm (0.75–3.5 in.) that looks heavier because of its long, thick, coarse hair. The three-toed sloth weighs 2.25–6.20 kg (5–13.5 lb). It has three toes on each front and rear foot. The head is small and round, the tail is short and stumpy, and the eyes and ears are reduced—the animal can neither hear nor see well but depends on its senses of smell and touch. The three species in this genus have 30 ribs and long necks containing nine vertebrae (two more than most mammals). These sloths are drab brown or gray with patches of white. Males have a bright yellow spot, about the size of a silver dollar, on their back. In this spot, the hair is short and glossy.

Megalonychidae. The family Megalonychidae (two-toed tree sloths) contains a single genus (*Choloepus*) with two species. Two-toed tree sloths, or unau, with two toes on their front feet and three on their hind feet, are somewhat larger and more active than their three-toed cousins. Adults range 540–740 mm (21–28 in.) in total length and weigh 4.0–8.5 kg (9–18.5 lb). The coloration is grayish brown with a paler face. The large hooklike claws are 8–10 cm (3–4 in.) long. These sloths possess 48 ribs, six to eight vertebrae, and their tail is either absent or vestigial.

Physiology and ecology. Tree sloths are solitary and spend most of their lives hanging upside down from the upper branches of a tree. They hang by the long, hooked claws at the ends of their toes and slowly move along, hand over hand in the inverted position. They spend their waking hours slowly picking and eating leaves, especially the leaves of the Cecropia

Three-toed tree sloth (*Bradypus tridactylus*). (*Photo by Dr. Lloyd Glenn Ingles; © California Academy of Sciences*)

tree. Their stomachs contain cellulose-digesting bacteria and may retain food for as long as a month before passing it on to the small intestine. Sloths are so lethargic that two blue-green algae, which look like moss, actually grow in the grooves of their coarse hair. In the rainy season, their fur often has a distinct greenish tinge due to the luxuriant growth of algae; in times of drought, the fur turns yellowish. The greenish color serves to camouflage the animals in the tree canopy.

Sloths have a low body temperature of 30–34°C (86–93°F) which fluctuates both daily and seasonally. They regulate their temperature by shifting their position in and out of the sun. Sloths make no attempt to construct a den or shelter in the trees. They sleep and give birth while hanging upside down. They normally sleep about 18 hours a day on a vertical branch, keeping a firm hold with all four feet with their head hanging down and resting between the front limbs. Predators include jaguars and eagles, although habitat destruction and excessive hunting may prove to be the most detrimental factors affecting these species. One species (*B. torquatus*) is classified as endangered by the International Union for Conservation of Nature and Natural Resources (IUCN) and the U.S. Department of the Interior.

Reproduction and development. Most sloths are believed to breed throughout the year. A single young is born after a gestation ranging from 3.5 to 11.5 months. During its first few months, the young sloth clings flat to the mother's belly. They cease nursing between 4 and 5 weeks of age. Female two-toed sloths reach sexual maturity at 3 years of age, whereas males are sexually mature at 4–5 years of age. Life expectancy in the wild is up to about 12 years, although they have lived up to 31 years in captivity.

Donald W. Linzey

Bibliography. D. Macdonald (ed.), *The Encyclopedia of Mammals*. Andromeda Oxford Limited, 2001; R. M. Nowak, *Walker's Mammals of the World*, 6th ed., Johns Hopkins University Press, 1999.

Slow neutron spectroscopy

The use of beams of slow neutrons, from nuclear reactors or nuclear accelerators, in studies of the structure or structural dynamics of solid, liquid, or gaseous matter. Studies of the chemical or magnetic structure of substances are usually referred to under the term neutron diffraction, while studies of atomic and magnetic dynamics go under the terms slow neutron spectroscopy, inelastic neutron scattering, or simply neutron spectroscopy. The results obtained are to a considerable extent complementary to those obtained by use of optical spectroscopy and x-ray diffraction. *See* NEUTRON DIFFRACTION; NUCLEAR REACTOR; PARTICLE ACCELERATOR; SPECTROSCOPY; X-RAY DIFFRACTION.

Experiments. In a neutron spectroscopy experiment, a beam of neutrons is scattered by a specimen and the scattered neutrons are detected at various angles to the initial beam. From these measurements, the linear momenta of the incoming and outgoing neutrons (and the vector momentum changes experienced by individual neutrons) can be computed. Experiments are often carried out in a mode in which the vector momentum change $\mathbf{Q}$ is held constant and the energy transfer ϵ is employed as the independent variable for the experiment.

In general, just those neutrons which have been scattered once only by the specimen are useful for analysis; the specimen must be "thin" with respect to neutron scattering power as well as to neutron absorption. In practice, the experiments are usually intensity-limited, since even the most powerful reactors or accelerators are sources of weak luminosity when, as here, individual slow neutrons are to be considered as quanta of radiation.

Neutron spectroscopy requires slow neutrons, with energies of the order of neutrons in equilibrium with matter at room temperature, or approximately 0.025 eV. The corresponding de Broglie wavelengths are approximately 0.2 nanometer, of the order of interatomic spacings in solids or liquids. The fast neutrons emitted in nuclear or slow fission reactions can be slowed down to thermal velocities in matter which is transparent to neutrons and which contains light elements, such as hydrogen, carbon, and beryllium, by a process of diffusion and elastic (billiard-ball) scattering known as neutron moderation. By selection of those diffusing neutrons which travel in a certain restricted range of directions

(collimation), a beam of thermal and near-thermal neutrons can be obtained. The beam of slow neutrons so produced can be selected as to energy by means of their velocities, by using an electronic or mechanical velocity selector, or by means of their wavelengths, through diffraction from a specially selected and treated single-crystal monochromator. As theoretically expected, the two methods are found to be physically equivalent, but have important differences for the strategy of experiments. *See* NEUTRON; QUANTUM MECHANICS; THERMAL NEUTRONS.

Interpretation. The bulk of the observations can be accounted for in terms of scattering of semiclassical neutron waves by massive, moving-point scatterers in the forms of atomic nuclei and their bound electron clouds. The spatial structure of the scatterers, time-averaged over the duration of the experiment, gives rise to the elastic scattering from the specimen that is studied in neutron diffraction; the spatial motions of the scatterers give rise to the Doppler-shifted inelastic scattering involved in slow neutron spectroscopy.

Quasiparticles. Analysis of the scattering patterns is particularly straightforward when the neutrons can be envisaged as creating or annihilating quantum excitations called quasiparticles in the course of the scattering event. The single-quantum component then includes peaks in the scattering pattern—either sharp peaks, whose energy width reflects the energy resolution of the experiment and instrument, or broadened peaks, whose energy width is interpreted in terms of natural lifetimes for the excitations and of interactions between them. Each peak then yields, through the assumption of conservation of energy and momentum between neutron and quasiparticle, an energy ϵ and a momentum $\mathbf{Q}$ for the (created or annihilated) single quasiparticle involved. Repetition under different conditions permits deduction of the dispersion relation, $\epsilon_q = \epsilon_q(\mathbf{Q})$, of the quasiparticle concerned. Consideration of the intensity of the neutron line allows assignment of the polarization and symmetry type of the excitation concerned—for example, phonon or magnon, longitudinal or transverse, acoustic or optical. *See* DISPERSION RELATIONS; LATTICE VIBRATIONS; LIQUID HELIUM; MAGNON; PHONON.

Likewise, through analyses of their widths and shapes, the broadened peaks yield estimates of the lifetimes $\tau_q(\mathbf{Q})$ of the quasiparticle and information on its modes of interaction in the specimen. Inverse lifetimes ($1/\tau$) or frequencies (ϵ/h, where h is Planck's constant) in the range from 10^{10} to 10^{13} Hz are accessible with this method.

Diffusion phenomena. Broadened peaks may occur at zero energy transfer. This quasielastic scattering is usually the result of slow diffusionlike motions of scattering entities through the specimen. What is unusual is that distances involved are of atomic dimensions and times correspondingly short. Thus it is possible to observe self-diffusion in classical liquids on an atomic scale. The phenomena are best considered through the correlation function approach discussed below. *See* DIFFUSION.

Continuous spectra. For continuous spectra, interpretation is more difficult and indirect. Always evocative in principle (and at times useful in practice) are the Van Hove space-time correlation functions $G(\mathbf{r},t)$. These are natural extensions to space-time of the Patterson pair correlation functions of crystallography and the Zernike-Prins pair distribution functions in liquids. For a monatomic liquid, the self-correlation function $G_s(\mathbf{r},t)$ has a mathematical definition which can be approximately expressed as follows: Given an atom at position $\mathbf{r} = 0$ at time $t = 0$, then $G_s(\mathbf{r},t)$ is the probability that the same atom is at position $\mathbf{r}$ at time t. Similarly, the pair correlation function $G_p(\mathbf{r},t)$ of a monatomic liquid can be approximately defined by the statement: Given an atom at position $\mathbf{r} = 0$ at time $t = 0$, then $G_p(\mathbf{r},t)$ is the number density of other atoms at position $\mathbf{r}$ at time t. In simple cases of continuous spectra, these functions provide a conceptually simple and powerful method of direct analysis, similar to the role played by quasiparticles in the study of line spectra. This is also the case for magnetic inelastic scattering when there is only one species of magnetic atom involved.

When more than one type (j) of atom is involved, the scattering pattern is related to a weighted superposition of correlation functions (G) between pair types (j,j'). In this case, isotopic substitution (for example, deuterium for hydrogen, or ^{60}Ni or ^{62}Ni for natural nickel) can assist in sorting out the contributions. (The weights are proportional to the products of the neutron scattering amplitudes of the pair types.) When the scattering is magnetic, the three spatial components of the vector magnetic moments of the atoms can also be involved through their various self and pairwise correlations.

Applications. Just as (slow) neutron diffraction is the most powerful available scientific tool for study of the magnetic structure of matter on an atomic scale, so slow neutron spectroscopy is the most powerful tool for study of the atomic magnetic and nuclear dynamics of matter in all its phases. The direct nature of the analysis has in some cases added considerable support to the conceptual structure of solid-state and liquid-state physics and thus to the confidence with which the physics is applied. For example, neutron spectroscopy has confirmed the existence of phonons, magnons, and the quasiparticles (rotons) of liquid helium II. Detailed information has been obtained on the lattice vibrations of most of the crystalline elements and numerous simple compounds, on the atomic dynamics of many simple liquids, on the dynamics of liquid helium in different phases, and on the atomic magnetic dynamics of a great variety of ferromagnetic, ferrimagnetic, antiferromagnetic, and modulated magnetic substances.

Bertram N. Brockhouse

Bibliography. E. Balcar and S. W. Lovesey, *Theory of Magnetic Neutron and Photon Scattering*, 1989; H. Glyde, *Excitations in Liquid and Solid Helium*, 1995; J. S. Higgins and H. C. Benoit, *Polymers and Neutron Scattering*, 1994, reprint 1996; S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter*, 2 vols., 1986.

Slug (zoology)

A terrestrial pulmonate mollusk in which the shell is absent or reduced to a small internal or external rudiment. The slug form has evolved independently several times. The incorporation into the muscular foot region of the body organs (which are contained within the shell in other mollusks) results in a streamlined body shape (see **illus.**), enabling the animal to enter small holes or crevices.

***Limax maximus*, up to 8 in. (20 cm) in length. There are two pairs of tentacles on the head, and the opening to the lung is visible.**

The skin is soft and moist, and readily loses water by evaporation or gains it by absorption. Maximum water loss occurs when crawling, due to laying down the slime trail. Water loss is minimized by behavioral mechanisms; thus most slugs are abundant on the surface only on windless wet nights.

Lung respiration occurs as in other pulmonates, but skin respiration is probably at least as important.

Slugs are simultaneous hermaphrodites. Copulation follows an often elaborate courtship, and is frequently reciprocal. In many species, sperm are packaged in a spermatophore. Cross-fertilization is normal, but self-fertilization is possible and may indeed be common. Eggs are laid in moist crevices as they have little resistance to desiccation.

A few slugs are carnivorous, such as *Testacella* which eats earthworms, but the majority are catholic herbivores and may become serious horticultural and agricultural pests. *See* MOLLUSCA; PULMONATA.

Norman Runham

Smallpox

An acute infectious viral disease characterized by severe systemic involvement and a single crop of skin lesions that proceeds through macular, papular, vesicular, and pustular stages. Smallpox is caused by variola virus, a brick-shaped, deoxyribonucleic acid–containing member of the Poxviridae family. Strains of variola virus are indistinguishable antigenically, but have differed in the clinical severity of the disease caused. From at least 1157 B.C., when the disease probably killed the Egyptian Pharaoh Ramses V, hundreds of millions of cases and millions of deaths due to smallpox occurred throughout the world. On October 26, 1977, Ali Maow Maalin, a cook from Merca, Somalia, had onset of the smallpox rash; he was the last patient to contract the disease through natural transmission, although a laboratory outbreak of smallpox occurred in the United Kingdom in 1978. Following a 13-year worldwide campaign coordinated by the World Health Organization (WHO), smallpox was declared eradicated by the World Health Assembly in May 1980. Smallpox is the first human disease to be eradicated.

Epidemiology. Humans were the only reservoir and vector of smallpox. The disease was spread by transfer of the virus in respiratory droplets during face-to-face contact. Before vaccination, persons of all ages were susceptible. As the eradication program progressed, attack rates were highest in young children and older adults. Children had low vaccination coverage due to high birth rates and poor health infrastructure in developing countries; older persons, often exposed to smallpox while caring for sick family members, lost immunity if not revaccinated since youth. In 1967, when the intensified global eradication program began, 33 nations reported endemic disease, and 14 other countries reported importations of cases. Over 131,000 cases were recorded by WHO in 1967, but special surveys indicated that less than 1% of the cases were reported. *See* VACCINATION.

It is estimated that 10–15 million cases and 1–3 million deaths due to smallpox occurred yearly in the mid-1960s. All countries with endemic smallpox were in the tropics; in 1967, about 65% of the recorded cases were from India, while 78% of the smallpox-endemic countries were in Africa.

Smallpox was a winter-spring disease; there was a peak incidence in the drier spring months in the Southern Hemisphere and in the winter months in temperate climates. The spread of smallpox was relatively slow. The incubation period was an average of 10–12 days, with a range of 7–17 days. Fifteen to forty percent of susceptible persons in close contact with an infected individual developed the disease.

Manifestations. Smallpox was once one of the most feared diseases because it caused widespread skin lesions, often leaving disfiguring scars on the face. Lesions began as reddened or dark spots (macules) and evolved through papules, vesicles, and pustules over 7–14 days. The lesions were round, approximately 0.5 cm in diameter, and firmly fixed to the basal layers of the skin. Over the next 7–14 days, the pustules would dry and form crusts that would fall off, leaving pitted areas with reduced pigmentation, particularly over skin areas with a heavy concentration of sebaceous glands. In time, these areas would flatten and develop excess pigmentation. The eruption was distributed peripherally, differing from the rash of chickenpox (varicella), which has a typical distribution over the torso but was often confused with smallpox.

There were two main clinically distinct forms of smallpox, variola major and variola minor. Variola major, prominent in Asia and west Africa, was the more severe form, with widespread lesions and case fatality rates of 15–25% in unvaccinated persons, exceeding 40% in children under 1 year. From the early 1960s to 1977, variola minor was prevalent in

South America and south and east Africa; manifestations were milder, with a case fatality rate of less than 1%.

Treatment. There is no specific treatment for the diseases caused by poxviruses. Supportive care for smallpox often included the systemic use of penicillins to minimize secondary bacterial infection of the skin. When lesions occurred on the cornea, an antiviral agent (idoxuridine) was advised.

Prevention. Although inoculation with material from cowpox lesions was known in England in the mid-1700s, Edward Jenner, a British general medical practitioner who used cowpox to prevent smallpox in 1796, is credited with the discovery of smallpox vaccine (vaccinia virus). Over the next two centuries, major efforts were made to keep the fragile vaccine potent and stable in warm climates. These problems were overcome with the application of the freeze-dry technique to the manufacture of smallpox vaccines in the 1960s and with the strict quality control by WHO of vaccine used in the field. However, the intensified global smallpox eradication program did not rely only on vaccination. Although the strategy for eradication first followed a mass vaccination approach, experience in the west and central African campaign during 1968 and 1969 showed that intensive efforts to identify areas of epidemiologic importance, to detect outbreaks and cases, and to contain them would have the greatest effect on interrupting transmission.

Maintenance of eradication. Escape of variola virus from laboratories in the United Kingdom resulted in three cases of smallpox in 1973 and two cases in 1978, but these outbreaks were not due to natural transmission. WHO has since established requirements to assure strict containment of variola virus in laboratories.

In 1978, WHO established an International Commission to confirm the absence of smallpox worldwide. The 19 recommendations made by the commission included reducing the number of laboratories retaining variola virus, pursuing research on selected orthopoxviruses (such as monkeypox, an animal virus that causes a smallpoxlike eruption but only rarely infects humans and has little potential for interhuman spread), and abandoning routine vaccination except for laboratory workers at special risk. By 1990, only the Soviet Union and the United States were maintaining variola virus, which is kept under high security. Scientists and health officials in these countries are considering destroying the last remaining strains of variola virus after the viral genome has been completely mapped. *See* ANIMAL VIRUS.

Joel Breman

Bibliography. J. G. Breman and I. Arita, The confirmation and maintenance of smallpox eradication, *N. Engl. J. Med.*, 303:1263–1273, 1980; A. Deria et al., The world's last endemic case of smallpox: Surveillance and containment measures, *Bull. WHO*, 58:279–283, 1980; F. Fenner et al., *Smallpox and Its Eradication*, World Health Organization, 1988; D. A. Henderson, The eradication of smallpox, *Sci. Amer.*, 235:25–33, 1976.

Smart card

A plastic card, the size of a credit card, that contains an embedded silicon computer chip. Its primary purpose is the portable storage and retrieval of data used to authorize various types of electronic transactions. Compared to their predecessors, magnetic stripe cards (for example, credit cards), smart cards can store data relating to more than one institution, facilitate more than one type of use, carry substantially larger volumes of data, process transactions at higher rates, and provide logical and physical security to the card's data. Smart cards work by interacting with a card reader—an interface between the smart card and an institution.

History. Smart cards were developed in the late 1960s to 1970s, when inventors in France, Germany, and Japan formed the idea of cards holding microchips (**Table 1**). Throughout the 1980s, smart card implementation was mainly confined to Europe and Japan. The United States favored wired telecommunication, which supported on-line cardholder/transaction fraud detection, reducing the need to adopt stringent forms of identify verification associated with batch transaction processing. As a result, smart card developments were limited to trials by a few banks, the Department of Defense, and a few universities. Since the September 11, 2001, terrorist attacks, there has been a race to implement stringent identification procedures involving smart cards, primarily in high-liability operating environments.

Applications. Smart cards support a wide range of transactions based on their ability to identify and store information about the cardholder, card issuer, authorized accounts and their activities, authorized merchants, and authorized transactions. The validation of a cardholder's identity may be knowledge-based, where the individual knows something, such as a personal identification number (PIN). It may also be possession-based, where the presence of a "token" (for example, an ID card or a bank card) validates the person's identity. In addition, smart cards

TABLE 1. Invention and deployment of the smart card

Year	Event
1968	Jorgen Dethloff and Helmut Grotrupp (Germany) patent the idea of combining computer chips with plastic cards.
1970	Kunitaka Arimura (Japan) receives Japanese patent related to the IC card.
1974	Roland Moreno (France), journalist, receives patent in France for chip cards. Considered father of smart cards, he founded Innovatron, which issued first global licenses.
1976	Honeywell Bull (France) receives licenses as a result of DGT initiative.
1979	Flonic Schlumberger (France) and Honeywell Bull & Philips (Netherlands) receive first Innovatron licenses.
1980	GIE sponsors first smart card trials in three French cities.
1982	First USA trials held in North Dakota and New Jersey.
1996	First USA trials held on university campus.

can carry biometric information (for example, fingerprints or iris scans), providing greater security for the identification process.

Some common smart card applications include credit transactions, debit transactions, stored-value transactions, information management applications, loyalty/affinity transactions, and multiple applications.

Credit transactions. Smart cards reduce fraud by securing merchant transactions. Smart credit cards give cardholders financial purchasing flexibility in the form of preapproved cash advances and loans (initiated at the point of sale), which are associated with specific electronic "purses" contained on the card.

Debit transactions. Smart cards provide electronic direct-debit at automated teller machines and at specific points of sale. This improves the cardholder's access to specific goods and services in locations where the cardholder's identity and credit history are unknown.

Stored-value transactions. Institutions can create smart cards with a fixed value initially encoded in the card chip's memory. This stored value may be disposable (debit "read-only") or reloaded (debit/credit "rewrite"). From this sum, purchases can be deducted. These transactions do not require access to the cardholder's personal identity, so no PIN is required. Examples include telephone cards and retail merchant gift cards.

Information management applications. Smart cards can be used for the portable storage and use of cardholder's personal information such as bank credit/debit accounts, insurance information, medical history, emergency contact information, and travel documentation.

Loyalty/affinity transactions. Smart cards can apply vendor incentives (such as points, credits, discounts, or direct delivery of products or services) at the point of sale, which are tied to the purchase of goods or services.

Multiple applications. Smart cards are able to combine two or more applications on a single card (some of which are able to share PIN numbers). At present, this form of combined identity management requires the capabilities of an integrated circuit.

Physical specifications. The size of a smart card is the same as that of a typical magnetic stripe credit, debit, or ID card. The International Standards Organization specification, ISO 7810, established the size of these cards, called ID-1 cards. ISO 7810 specifies that cards be 85.7 mm (3.4 in.) wide, 55.2 mm (2.1 in.) high, and 0.76 mm (0.03 in.) deep (**Table 2**). Smart cards adhere to this standard to promote backward compatibility with the existing transaction processing infrastructure.

TABLE 2. Smart card standards

Standard	Description
ISO 7810	ID cards—physical characteristics
ISO 7811	Identification cards, magnetic stripe, embossing
ISO 7816	Design and use of contact ICC cards
ISO/IEC 10536	Contactless ICC cards
ISO/IEC 14443	Proximity cards
ISO/IEC 15693	Vicinity cards

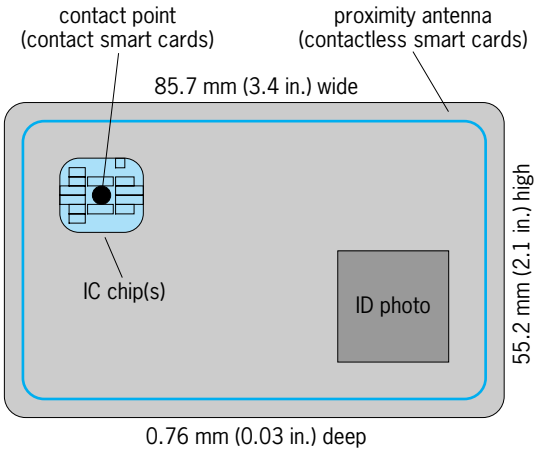


Fig. 1. Smart card showing both the contact point and the proximity antenna.

Contact versus contactless smart cards. Contact cards require that the card remain in physical contact with the reader. ISO 7816 specifies the requirements for contact cards. In general, they have a gold plate on one side of the card, covering contact points between the card's chip and the reader (**Fig. 1**). A contact smart card must be physically inserted into the reader and remain in contact with it during the session. The reader verifies that the card is properly situated and that the power supplies of both card and reader are compatible. It then supplies the card with the power necessary to complete the transaction.

The contactless card (ISO 10536, 14443, and 15693) transmits information to and from the card reader via an antenna embedded in the card (Fig. 1). Contactless smart cards rely on electrical coupling for their power. At present, contactless cards tend to be slower, less reliable, and more expensive than contact cards. However, this technology is rapidly improving and will eventually replace contact smart cards.

Integrated circuits. The computer chips embedded in smart cards can be either memory chips or microprocessors. At present, memory chips have 1–5 kilobits of memory. Less expensive than microprocessors, they depend on the card reader for processing and security. *See* INTEGRATED CIRCUITS; MICROPROCESSOR; SEMICONDUCTOR MEMORIES.

Smart cards with microprocessors are "true" smart cards. In many ways, the microprocessor makes them a miniature computer, complete with input/output ports, an operating system, and a hard drive. Microprocessor chips are available with 8-, 16-, and 32-bit architectures. They contain both short- and long-term memory cells, ranging from a few hundred bytes of read-only memory (ROM) to several megabytes of random-access memory (RAM). Usually, the operating system is loaded in the ROM, while the RAM provides a "scratch pad" for processing information.

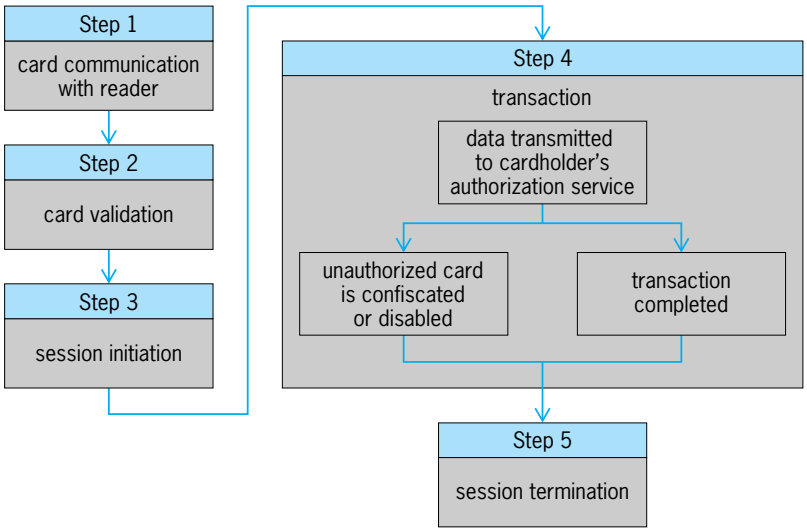


Fig. 2. Typical smart card transaction.

While smart cards are programmed using a variety of languages, the most popular is Java, which can be compiled directly on the card. Current innovations center on the development of minidatabases that can be stored on the card. As a result, smart card applications are increasingly important enablers of identify-based transactions. *See* PROGRAMMING LANGUAGES.

Transaction. A typical smart card transaction session consists of communication, validation, initiation, transaction, and termination (**Fig. 2**).

Communication. Communication with the reader may be direct or in the vicinity of the reader, depending on the type of card (contact or contactless) and the capability of the reader.

Validation. The card must be validated in order to establish a session. If the card is valid, the system captures an identification number to establish an audit trail and processes the transaction using the authorized "value" that is stored on the chip (for example, removing cash from an electronic purse). This activity is handled by a secure access module (SAM) on a chip housed in the reader. The SAM has the electronic keys necessary to establish communication with the reader's operating system.

Initiation. A reset command is issued to establish communication between the card and the reader, and the clock speed is established to control the session. In the case of a both contact-and-contactless smart card, the reader obtains the required data from the card and initiates the requested transaction.

Transaction. Data are exchanged between the card and the reader through the defined contact (input/output) points. A record of the transaction is stored on both card and reader.

Termination. For contact smart cards, the contacts are set to a stable level, the supply voltage is terminated, and the card is ejected from the terminal. For contactless cards, a termination protocol ends the session and resets the reader.

Security. There are two levels of smart card security: physical and data.

Physical security. This involves protecting the smart card against damage from cardholders, damage from the environment, and physical tampering. Bending a card, scratching its contacts, or breaking chip connections will destroy the card. New manufacturing technologies and protective sleeves are being designed to protect cards.

A range of environmental conditions can damage or destroy smart cards. Examples include electric voltage, frequency of use, humidity, light, temperature, and radiation exposure.

Various methods of tampering with smart cards have been suggested or tried. Smart card manufacturers have protected the layout of the IC chip by hiding its data pathways. Cards that show physical traces of tampering can be identified and their use blocked by readers. Manufacturers also use other technologies including holograms, hidden characters, and fingerprints to make counterfeiting unprofitable for criminals.

Data security. Smart card systems are susceptible to attempts to invade systems to steal or destroy data. This may be done by visually capturing the keystrokes of cardholders in order to learn the PIN associated with the card. A more sophisticated threat is infiltration of the system, with the goal of finding and stealing valuable data. In the case of contactless smart cards, data that are wirelessly transmitted between cards and readers potentially may be intercepted.

In response to these threats, smart cards use key-based cryptography for data encryption. There are various methods for this based on the logic of the "one-time pad," a randomly generated single-use ciphertext (encrypted text) that is decrypted by both the reader and the card and required to complete a transaction. *See* CRYPTOGRAPHY.

In a symmetric-key-encrypted transaction, the card accesses a public-key directory (stored into the reader), called a certificate authority (CA), to encrypt the public key with a "hash" (a document digest of

the key). The information used to create this hash includes version number, serial number, signature algorithm, card issuer's name, and so on. Both the card and the CA receive a decrypted copy of the hash. Next, the reader accesses the CA to get the decrypted copy of the hash. This validates and authorizes the identity of both parties, allowing the transaction to take place. The transaction recipient accesses the CA for a final hash, which is encrypted and acts as a response to the card, identifying and authorizing the reader and notifying the card of the close of the transaction.

Asymmetric encryption makes use of both public and private keys to provide additional security for sensitive transactions.

Institutions using smart cards have developed several measures to deal with insecure cards. The most popular among these is "hot listing" of insecure cards. Early on, when only contact smart cards were available, the reader was programmed to retain problematic cards. This proved problematic when readers "ate" damaged but secure cards. Hot listing of contactless cards usually involves notifying the readers of insecure cards and programming the system to refuse transactions. *See* COMPUTER SECURITY.

Katherine M. Shelfer; Kevin Meadows

Bibliography. *Open Smart Card Infrastructure for Europe V 2*, Vol. 6: *Contactless Technology*, pt. 1: *White Paper on Requirements for the Interoperability of Contactless Cards*, eESC TB 8 Contactless Smart Cards, March 2003; K.M.Shelfer et al., Smart cards, *Advances in Computer Science*, vol. 60, 2003.

Smith chart

A graphical procedure for converting impedances into reflection coefficients and vice versa. At high-frequency circuit operation, voltages and currents behave like traveling waves propagating over finite-length components. Copper traces on printed circuit boards, coaxial cables, and even simple wires become transmission lines. Changes in the length or operating frequency of these transmission lines result in periodic impedance behaviors not encountered in low-frequency circuits. Frequently, the impedance is replaced by the reflection coefficient, a more convenient way to quantify the transmitted and reflected voltage-current waves. To show how the impedance can be converted into a reflection coefficient and vice versa, P. H. Smith developed an ingenious graphical procedure based on conformal mapping principles. His approach permits an easy and intuitive display of the reflection coefficient as well as the complex impedance in a single graph. Although such a graphical procedure, nowadays known as the Smith chart, was developed in the 1930s prior to the computer age, it has retained its popularity and can be found in every data book describing passive and active high-frequency components and systems. Almost all computer-aided design programs utilize the Smith chart for the analysis of circuit impedances, design of matching networks,

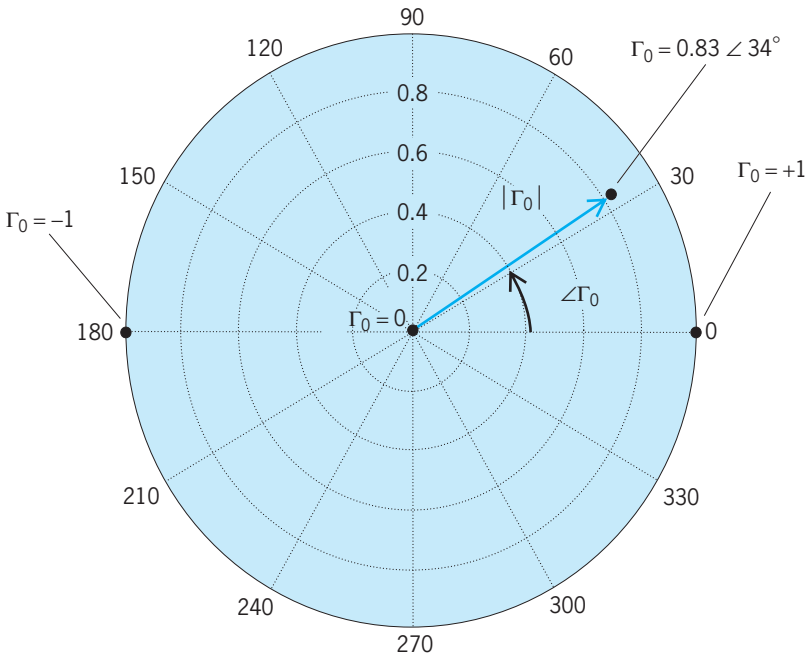


Fig. 1. The reflection coefficient Γ_0 as a complex phasor in the Γ-plane. Open- and short-circuit conditions are located at the right and left sides, while impedance matching to the characteristic impedance is achieved at the center.

and computations of noise figures, gain, and stability circles. Even instruments such as the ubiquitous network analyzer have the option to represent certain measurement data in a Smith chart format. *See* CONFORMAL MAPPING; ELECTRICAL IMPEDANCE; REFLECTION AND TRANSMISSION COEFFICIENTS; TRANSMISSION LINES.

Reflection coefficient. At its core the Smith chart displays the complex reflection coefficient in magnitude and phase. If we consider the end of a transmission line, we can define a reflection coefficient Γ_0 in the form given by the equation below, where

$$\Gamma_0 = |\Gamma_0| \angle \Gamma_0 = \frac{Z_L - Z_0}{Z_L + Z_0}$$

Z_L is a generic complex termination impedance and Z_0 is the reference impedance, typically the characteristic impedance of the transmission line. **Figure 1** depicts the reflection coefficient as a complex vector in polar form with magnitude and phase. In this example a load impedance of $Z_L = (50 + j150)\ \Omega$ is attached to a $Z_0 = 50\ \Omega$ transmission line. The reflection coefficient $\Gamma_0 = 0.83 \angle 34°$ can then be displayed as a phasor in the complex plane. It is interesting to note that open-circuit ($Z_L \to \infty$) and short-circuit ($Z_L \to 0$) conditions result in the real axis points $\Gamma_0 = +1$ and $\Gamma_0 = -1$. In other words, all values of the reflection coefficient reside within the unit circle. *See* ALTERNATING-CURRENT CIRCUIT THEORY; COMPLEX NUMBERS AND COMPLEX VARIABLES.

Impedance and reflection coefficient representation. Smith's contribution was the recognition that a generic complex impedance $Z = R + jX$ can be expressed in terms of circles that are superimposed over the reflection coefficient representation. The

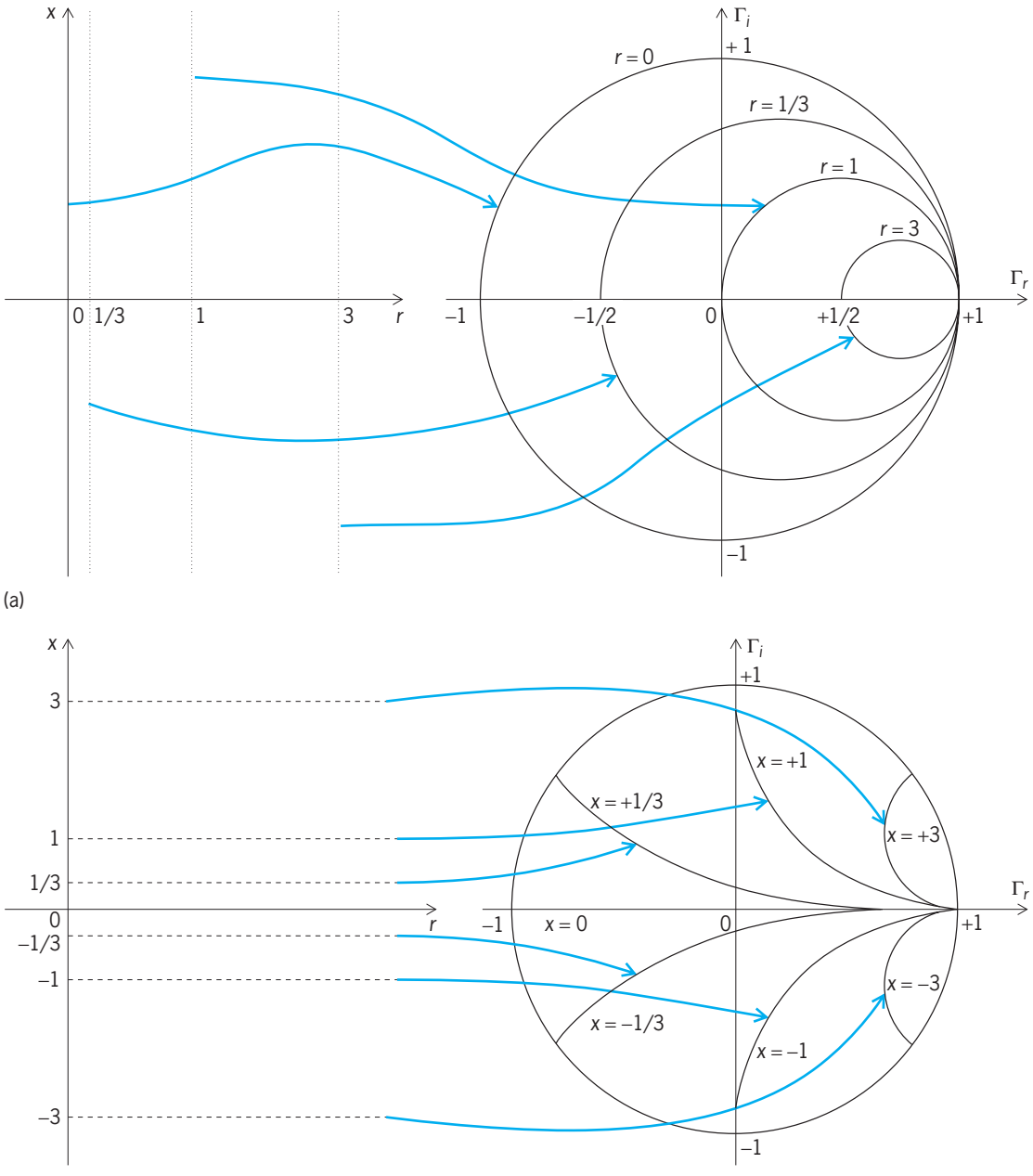


Fig. 2. Mapping of normalized resistance and reactance values into the complex reflection coefficient plane. (*a*) Mapping of normalized resistance values $r = 0$, 1/3, 1, 3 of the z-plane to corresponding r-circles in the complex Γ-plane. (*b*) Mapping of normalized reactance values $x = 0, \pm 1/3, \pm 1, \pm 3$ of the z-plane to corresponding x-circles in the complex Γ-plane.

mapping from the impedance plane into the reflection coefficient plane is done by first normalizing the impedance with respect to the characteristic impedance Z_0, $z = Z/Z_0 = r + jx$, with $r = R/Z_0$ and $x = X/Z_0$. This allows the display of all impedances within the unit circle. **Figure 2***a* depicts the mapping of r-values into shifted circles along the real axis of the reflection coefficient plane. As $r \to \infty$, the shifted circles get smaller and begin to collapse to a single point at the right side. Figure 2*b* shows a similar pattern for the reactance. However, unlike the resistance circles, the shifted reactance circles reside in the upper or lower plane, consistent with the fact that a circuit can exhibit inductive or capacitive behavior. *See* REACTANCE.

The trick is now to superimpose both the resistance and reactance circles over the complex unit circle plane, resulting in the final form of the Smith chart, shown in **Fig. 3**. In this figure it is evident how various complex normalized impedances can be interpreted as corresponding reflection coefficients. For instance, the four impedances of $(50 + j150)\ \Omega$, $(150 + j50)\ \Omega$, $(16.66 - j16.66)\ \Omega$, and $(150 - j150)\ \Omega$ are normalized with respect $Z_0 = 50\ \Omega$, yielding the normalized impedances of $(1 + j3)$, $(3 + j1)$, $(1/3 - j1/3)$, and $(3 - j3)$. These four points in the z plane are mapped into corresponding resistance and reactance circles in the complex Γ plane, as seen in Fig. 3. The intersecting resistance and reactance circles are identified as the

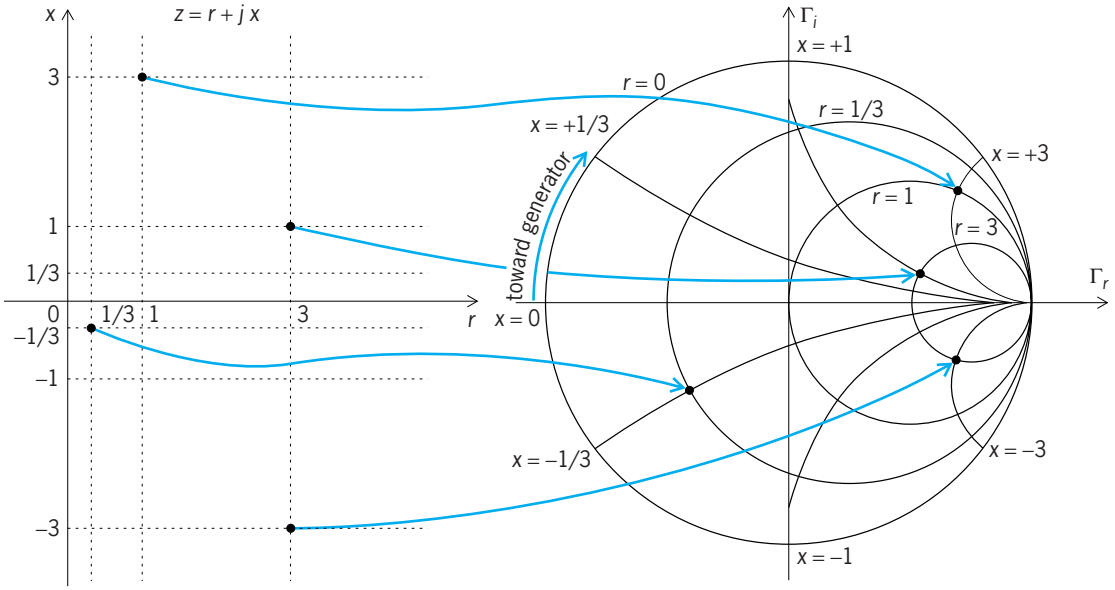


Fig. 3. The Smith chart is simply an overlay of resistance and reactance circles in the complex Γ-plane.

endpoints of vectors which are equivalent reflection coefficient phasors.

An important location is the normalized impedance value of $z = 1$; this spot maps into the center of the Smith chart, that is, $\Gamma = 0$. In the context of the above equation for the reflection coefficient, at this point the impedance is matched to the characteristic line impedance, $Z_0 = 50\ \Omega$. Matching to 50 Ω is often a key design requirement when dealing with conditioning the input and output ports of circuits like amplifiers, mixers, and oscillators.

The identification of an impedance point in the Smith chart and its corresponding reflection coefficient is but one consideration. Conversely, a reflection coefficient with its magnitude and phase can also be identified in terms of the normalized impedance value. Most radio-frequency (RF) devices are recorded in data sheets based on their scattering or S-parameters. For instance, to obtain the so-called input and output reflection coefficients S_{11} and S_{22}, it is sufficient to simply identify the intersecting resistance and reactance circles in order to obtain the equivalent normalized input and output impedances of the device.

Computations with the Smith chart. A typical example of how the Smith chart is employed to find the input impedance of a variable length transmission line (coaxial cable) is shown in **Fig. 4**. Here a frequency-independent load impedance of $Z_L = (30 + j60)\ \Omega$ is attached to the transmission line with characteristic impedance of $Z_0 = 50\ \Omega$. At a fixed operating frequency f of 2 GHz and based on a given phase velocity of $v_p = 0.5c$ (that is, 50% of the speed of light c), the coaxial cable is varied in length d. As the length is increased from 0 to 3 cm, the input impedance Z_{in} varies from $(30 + j60)\ \Omega$ to $(12.4 + j15.5)\ \Omega$. This is due to the fact that the reflection coefficient $\Gamma_0 = 0.63\angle 72^\circ$ begins to rotate in the mathematical negative sense from $\Gamma_{in} = \Gamma_0$ to the new location $\Gamma_{in} = \Gamma_0\angle(-2\beta d) = 0.63\angle -216^\circ$. Here $\beta d = 2\pi f d/v_p = 2.51$ rad (or 144°) is also known as the electric length of the coaxial cable. The rotation involves twice the electric length owing to the fact the voltage-current waves have to travel from the input of the coaxial cable to the load and back again. Reading off the normalized resistance and reactance circles, it is found that $r = 0.248$ and $x = 0.31$. Multiplying these values by $Z_0 = 50\ \Omega$ provides the final result of $Z_{in} = (12.4 + j\ 15.5)\ \Omega$.

Additional applications. Besides impedance transformations, the Smith chart finds widespread use for a number of RF applications; it is routinely utilized for noise figure, gain, and voltage standing-wave ratio (VSWR) computations. As an example, for the design of a low-noise amplifier (LNA) the optimal reflection coefficient for the minimum noise figure can typically be found in the data sheets of the

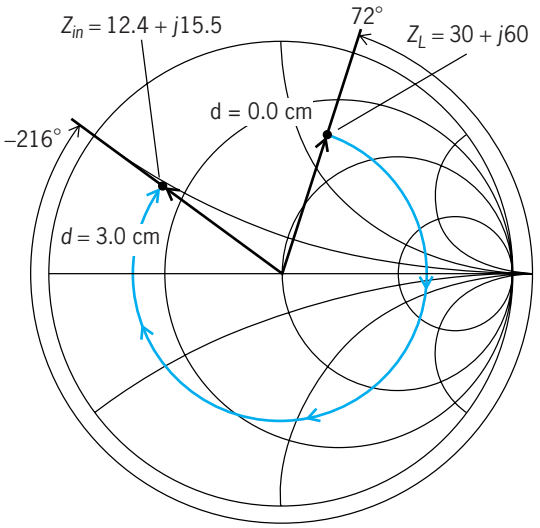


Fig. 4. Input impedance transformation of a coaxial cable from Z_L (at length $d = 0$) to Z_{in} (at $d = 3.0$ cm) at 2 GHz. The reflection coefficient of fixed magnitude is rotated from its original angular location of 72° to −216° in clockwise (or mathematical negative) direction toward the generator.

RF device. From the knowledge of the corresponding impedance value, the design engineer can then noise-match the device by designing an appropriate network whose impedance is matched to the conjugate complex input impedance of the device. Besides being useful for the design and simulation of RF systems, the Smith chart is frequently found in a number of measurement devices, notably the network analyzer. Here, the RF system under test is attached to the input and output ports of the network amplifier; a frequency sweep then allows the display of the input and output reflection coefficients as traces in the Smith chart. *See* MICROWAVE IMPEDANCE MEASUREMENT; RADIO-FREQUENCY AMPLIFIER.

Reinhold Ludwig

Bibliography. K.-C. Chan and A. Harter, Impedance matching and the Smith chart—The fundamentals, *RF Design Mag.*, pp. 52–66, July 2000; R. Ludwig and P. Bretchko, *RF Circuit Design, Theory and Applications*, Prentice Hall, 2000; B. Schweber, The Smith chart: More vital after all these years, *EDN Mag.*, pp. 50–52, March 19, 1999; P. H. Smith, *Electronic Applications of the Smith Chart*, 2d ed., Crestone Technical Books/Noble Publishing, 1995; P. H. Smith, Transmission-line calculator, *Electronics*, 12:29–31, 1939.

Smithsonite

A naturally occurring rhombohedral zinc carbonate ($ZnCO_3$), with a crystal structure similar to that of calcite ($CaCO_3$). The structure consists of alternating layers of zinc ions (Zn^{2+}) and carbonate ions (CO_3^{2-}) oriented perpendicular to the *c* crystallographic axis. Smithsonite has a hardness on the Mohs scale of $4^1/_2$, has a specific gravity of 4.30–4.45, and exhibits perfect rhombohedral cleavage. Optically, smithsonite is uniaxial negative, which means that it has a single optic axis and two principal refractive indices referred to as the ordinary refractive index and the extraordinary refractive index. Smithsonite is classified as optically negative because the ordinary refractive index is greater than the extraordinary index. ($N_o > N_e$). *See* CRYSTAL STRUCTURE; CRYSTALLOGRAPHY; HARDNESS SCALES.

Smithsonite most commonly forms as an alteration product of the mineral sphalerite (ZnS) during supergene enrichment of zinc ores in arid or semiarid environments. During this process, primary zinc minerals are oxidized and dissolved near the surface of the Earth. Although zinc ion is significantly more soluble than many other metals present in sulfide ore deposits (for example, copper, iron, lead, and silver), if solutions rich in zinc react with carbonate-rich surrounding rocks such as limestone, smithsonite may precipitate. Smithsonite thus deposited seldom occurs as well developed crystals but forms colliform, or botryoidal, encrusting masses, usually brownish or grayish white in color. It is seldom pure zinc carbonate, commonly containing other divalent metal ions such as manganese (Mn^{2+}), ferrous iron (Fe^{2+}), magnesium (Mg^{2+}), calcium (Ca^{2+}), cadmium (Cd^{2+}), cobalt (Co^{2+}), or lead (Pb^{2+}) substituting for zinc ion. Substitution of other elements for zinc may result in different colors such as blue-green (copper), yellow (cadmium), and pink (cobalt). Minerals frequently associated with smithsonite are hemimorphite [$Zn_4Si_2O_7(OH)_2 \cdot H_2O$], sphalerite (ZnS), galena (PbS), calcite ($CaCO_3$), and other secondary, supergene minerals.

Smithsonite has been mined as an ore of zinc and has also been used as an ornamental stone. Notable localities include Laurium, Greece; Sardinia, Italy; the Broken Hill Mine, Zambia; Tsumeb, Namibia; Broken Hill, New South Wales, Australia; and Kelly, New Mexico. *See* CALCITE; CARBONATE MINERALS; GALENA; HEMIMORPHITE; SPHALERITE; ZINC.

John C. Drake

Bibliography. L. L. Y. Chang, R. A. Howie, J. Zussman, *Rock-Forming Minerals*, vol. 5B: *Non-Silicates: Sulphates, Carbonates, Phosphates, Halides*, 1995; C. Klein and C. S. Hurlbut, Jr., *Manual of Mineralogy*, 21st ed., 1993; C. Palache, H. Berman, and C. Frondel, *The System of Mineralogy*, 7th ed., vol. 2, 1951; R. J. Reeder (ed.), *Reviews in Mineralogy*, vol. 11: *Carbonates: Mineralogy and Chemistry*, 1983.

Smog

The noxious mixture of gases and particles commonly associated with air pollution in urban areas. Harold Antoine des Voeux is credited with coining the term in 1905 to describe the air pollution in British towns. *See* AIR POLLUTION.

The constituents of smog affect the human cardiorespiratory system and pose a health threat. Individuals exposed to smog can experience acute symptoms ranging from eye irritation and shortness of breath to serious asthmatic attacks. Under extreme conditions, smog can cause mortality, especially in the case of the infirm and elderly. Smog can also harm vegetation and likely leads to significant losses in the yields from forests and agricultural crops in affected areas.

The only characteristic of smog that is readily apparent to the unaided observer is the low visibility or haziness that it produces, due to tiny particles suspended within the smog. Observation of the more insidious properties of smog—the concentrations of toxic constituents—requires sensitive analytical instrumentation. Technological advances in these types of instruments, along with the advent of high-speed computers to simulate smog formation, have led to an increasing understanding of smog and its causes. In much of the industrially developed world, this understanding has been used to design and implement pollution control measures to mitigate smog's harmful effects. The smoggiest cities today are mostly found in the developing world (for example, Chengdu, China; Mexico City, Mexico; and Santiago, Chile) where pollution controls are generally given a lower priority than economic growth. However, even in the developed world smog remains a significant environmental problem. For example, in the United States about 70 million people are

estimated to live in areas where smog poses a threat to human health. Moreover, while smog is generally most severe in and around urban centers, it is by no means only an urban problem. Under the appropriate conditions, smog can extend over thousands of kilometers and encompass agricultural and forested areas as well as urban areas. Some data even suggest that smog affects the composition of the atmosphere over the entire Northern Hemisphere.

Role of meteorology. Smog is an episodic phenomenon because specific meteorological conditions are required for it to accumulate near the ground. These conditions include calm or stagnant winds which limit the horizontal transport of the pollutants from their sources, and a temperature inversion which prevents vertical mixing of the pollutants from the boundary layer into the free troposphere. (The boundary layer refers to the lowest 1–2 km, or 0.6–1.2 mi, of the atmosphere, where winds are slowed by frictional resistance with the Earth's surface. The free troposphere refers to that part of the atmosphere lying between the boundary layer and the stratosphere.) *See* STRATOSPHERE; TROPOSPHERE.

The importance of meteorology also explains why some areas experience especially severe and frequent smog episodes and others do not. In cities such as Chengdu, Los Angeles, Vancouver, and Mexico City, surrounding mountain ranges restrict wind flow and prevent pollutant ventilation. In coastal cities such as Houston, the combination of nighttime offshore land breezes and daytime onshore sea breezes recycles urban pollutants and helps maintain smoggy conditions. *See* METEOROLOGY.

Smog can be classified into three types: London smog, photochemical smog, and smog from biomass burning.

London smog. Becoming prevalent in the midnineteenth century, London smog arises from the by-products of coal burning. These by-products include soot particles and sulfur oxides. During cool damp periods (often in the winter), the soot and sulfur oxides can combine with fog droplets to form a dark acidic fog.

The generation of London smog is not unique to modern times. For example, Theophrastus, in the third century B.C., wrote of the noxious quality of air exposed to coal burning. During the industrial revolution, the growing use of coal coupled with urbanization fostered more frequent and severe smog episodes in many European cities and especially in London. The first known scientific treatise on air pollution was produced by John Evelyn in 1661 and focused specifically on London smog. Probably the worst smog episode occurred in London in 1952 with an estimated 4000 deaths. Other documented killer smogs include episodes occurring in Glasgow in 1909, over the Meuse Valley in Belgium in 1930, in Donora, Pennsylvania, in 1948, in New York City in 1953, and in London in 1962. *See* ACID RAIN; COAL.

Photochemical smog. As nations switch from coal to cleaner-burning fossil fuels such as oil and gas as well as alternate energy sources such as hydroelectric and nuclear, London smogs cease. Unfortunately, photochemical smog then tends to appear. Photochemical smog is more of a haze than a fog and is produced by chemical reactions in the atmosphere that are triggered by sunlight. Because photochemical smog requires high solar intensities, its occurrence is rare when the dark foggy conditions of a London-type smog are present. *See* FOSSIL FUEL; PHOTOCHEMISTRY.

The phenomenon of photochemical smog was discovered and studied in the 1950s in Los Angeles when a growing urban population coupled with an increasing use of the automobile correlated with a discernible increase in hazy skies and an increase in the concentrations of compounds known as oxidants. At the same time, there appeared to be an increasing incidence of damage to crops in the areas surrounding Los Angeles.

A. J. Hagen-Smit first unraveled the chemical mechanism that produces photochemical smog. He irradiated mixtures of volatile organic compounds (VOC) and nitrogen oxides (NO_x) in a reaction chamber. (Note that both VOC and NO_x are emitted into the atmosphere by automobiles and other industrial processes, and NO_x represents the sum of nitric oxide or NO and nitrogen dioxide or NO_2.) After a few hours, Hagen-Smit observed the appearance of cracks in rubber bands stretched across the chamber. Knowing that ozone (O_3) can harden and crack rubber, Hagen-Smit correctly reasoned that photochemical smog was caused by photochemical reactions involving VOC and NO_x, and that one of the major oxidants produced in this smog was O_3. *See* NITROGEN OXIDES; OZONE.

Since Hagen-Smit's work, scientists have used more sophisticated reaction chambers, called smog chambers, along with advanced chemical monitors and computer simulation models to more fully elucidate the chemical reactions that form smog. A typical reaction sequence might include (1)–(8). Here, RH

$$RH + OH \longrightarrow R + H_2O \quad (1)$$

$$R + O_2 + M \longrightarrow RO_2 + M \quad (2)$$

$$RO_2 + NO \longrightarrow RO + NO_2 \quad (3)$$

$$RO + O_2 \longrightarrow HO_2 + \text{carbonyl} \quad (4)$$

$$HO_2 + NO \longrightarrow OH + NO_2 \quad (5)$$

$$2 \times (NO_2 + h\nu \longrightarrow NO + O) \quad (6)$$

$$2 \times (O + O_2 + M \longrightarrow O_3 + M) \quad (7)$$

$$RH + 4O_2 + 2h\nu \longrightarrow 2O_3 + \text{carbonyl} + H_2O \quad (8)$$

represents a hydrocarbon (that is, a type of VOC), R is a hydrocarbon radical or compound fragment (for example, CH_3CH_2 for RH = ethane), $h\nu$ represents a photon from the Sun, and M is a nonreactive, energy-absorbing molecule referred to by chemists as a third body. In the atmosphere, this third body is most often molecular nitrogen (N_2) or molecular oxygen (O_2). Carbon monoxide (CO) can also act as

a VOC in photochemical smog. In this case the reaction sequence is initiated by (9) and (10) and then followed by (5)–(7).

$$CO + OH \longrightarrow CO_2 + H \quad (9)$$

$$H + O_2 + M \longrightarrow HO_2 + M \quad (10)$$

In these reaction sequences, VOC and CO are oxidized and consumed, while NO_x acts as a catalyst. Moreover, the by-product labeled "carbonyl" is itself a VOC and can react and produce more O_3. The OH radical, which triggers the sequence by reacting with VOC or CO, is initially generated in the atmosphere from O_3 as in reactions (11) and (12). Here O^* is an

$$O_3 + h\nu \longrightarrow O^* + O_2 \quad (11)$$

$$O^* + H_2O \longrightarrow 2OH \quad (12)$$

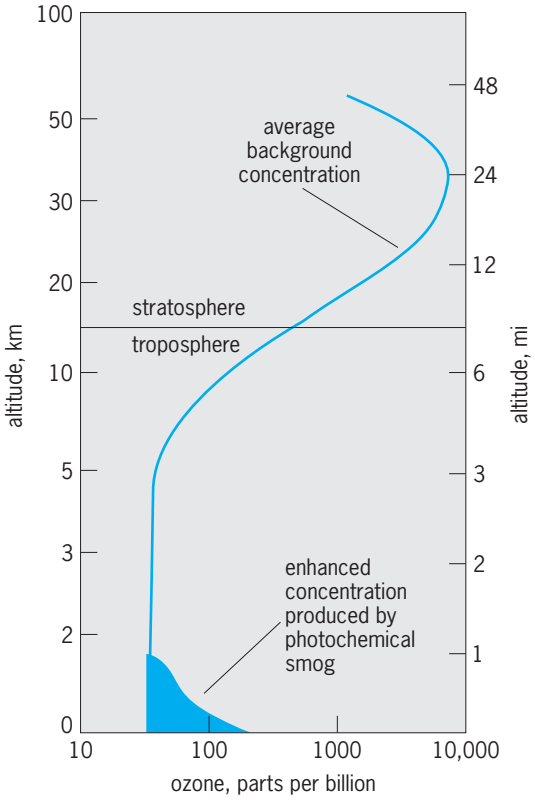
electronically excited atom of oxygen. Because the OH consumed by reaction (1) or (9) is regenerated by reaction (5), and because more OH can be generated during the breakdown of the carbonyl produced by reaction (4), the reaction sequences above are net sources of O_3. *See* ATMOSPHERIC CHEMISTRY.

Noxious constituents. While generally not as dangerous as London smog, photochemical smog contains a number of noxious constituents. Ozone, a strong oxidant that can react with living tissue, is one of these noxious compounds. Another is peroxyacetyl nitrate (PAN), an eye irritant that is produced by reactions between NO_2 and the breakdown products of carbonyls. Particulate matter having diameters of about 10 micrometers or less is another harmful smog component. Particles in this size range are of concern because they can penetrate into the human respiratory tract during breathing and have been implicated in a variety of respiratory ailments.

Mitigating photochemical smog. Because VOC and NO_x are required to generate O_3, they are referred to as the precursors of O_3. Human-induced NO_x is produced in high-temperature combustion, and human-induced VOC arises from the incomplete combustion of fossil fuels and the evaporation of solvents. Volatile organic carbons are also emitted into the atmosphere naturally by trees and other vegetation. Because the two precursors have different sources, controls that limit VOC emissions usually do not reduce NO_x emissions, and vice versa. Thus, before initiating a program to mitigate photochemical smog and O_3 pollution, it must first be determined which O_3 precursors to target for emission reductions.

Since both VOC and NO_x are required to produce O_3, it might be concluded that smog can be reduced by any one of three strategies, based on lowering (1) the emissions of VOC, (2) NO_x emissions, or (3) both. However, this conclusion would be incorrect. The photochemical smog mechanism is complex, and the relative effectiveness of pollution control strategies based on lowering VOC or NO_x emissions can vary depending upon local conditions. In various locales, VOC controls might be effective and NO_x controls ineffective, and vice versa. Ozone mitigation can be further confounded by natural emissions of VOC. In sylvan cities, such as Atlanta, Georgia, large emissions of hydrocarbons from trees (primarily as isoprene) render pollution control strategies based on lowering human-induced VOC emissions relatively ineffective and necessitate a greater focus on lowering NO_x emissions. Another complication arises from the fact that, under some circumstances, lowering NO_x emissions can actually lead to a temporary O_3 increase and occurs because of high concentration NO_x scavenger OH radicals, thereby preventing initiation of O_3 production via reactions (1) and (9). This is referred to as a NO_x disbenefit. It is also the case that strategies that effectively lower O_3 concentrations may not reduce the concentrations of other noxious smog components such as particulate matter. For these reasons, determining the best pollution control strategy for a given area has proven to be a difficult task, requiring the application of sophisticated air quality simulation models.

Tropospheric and stratospheric ozone. In addition to being a component of photochemical smog, O_3 is a natural constituent of the atmosphere (see **illus.**). In the stratosphere, where it is most abundant, O_3 actually protects life by filtering out potentially harmful ultraviolet radiation from the Sun. In the troposphere (the lowest 10–12 km or 6–7.5 mi of the atmosphere), O_3 arises naturally from the downward transport of stratospheric O_3 as well as photochemical production. Interestingly, the same photochemical



Average background concentration of ozone in the atmosphere as a function of altitude. Note the enhancement of O_3 near the Earth's surface that can occur in polluted areas as a result of photochemical smog.

reactions that produce O_3 in photochemical smog also produce tropospheric O_3. In the latter case, methane (CH_4) and carbon monoxide (CO) act as VOC precursors, while NO_x is primarily supplied by lightning. It is likely that pollution has accelerated the rate of production of tropospheric O_3 on a hemispheric scale by adding to the total atmospheric VOC, CO, and NO_x burdens. Some data suggest that present-day O_3 concentrations throughout much of the Northern Hemisphere are a factor of 2 or more larger than they were a hundred years ago.

Smog from biomass burning. Probably the oldest type of smog known to humankind is produced from the burning of biomass or wood. It combines aspects of both London smog and photochemical smog since the burning of biomass can produce copious quantities of smoke as well as VOC and NO_x.

Biomass-burning smog was most likely ubiquitous in preindustrial times. For example, in 1542 Juan Rodriguez Cabrillo, upon first arriving at the Los Angeles harbor, noted the heavy smoke from Indian campfires and named the place Bahia de los Humos (Bay of Smokes).

Even in modern times, biomass-burning smog can pose a serious environmental threat. In the 1980s, Nobel prize winner Paul Crutzen and his colleagues documented the large-scale degradation in air quality over South America from agricultural burning. During the summer of 1997, forest fires and agricultural burning in Indonesia produced a thick blanket of smoke-laden smog over much of Southeast Asia. The smog caused widespread economic disruptions and agricultural losses. The human toll was also high, with tens of thousands falling ill and many deaths. Indeed, appreciation of the dangers of biomass-burning smog led Crutzen and John Birks in 1982 to warn that a nuclear exchange and the resulting fires could cause a "nuclear winter" with catastrophic environmental consequences for the planet. William L. Chameides

Bibliography. J. Fishman and R. Kalish, *Global Alert: The Ozone Pollution Crisis*, Plenum, New York, 1990; National Research Council, *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Academy Press, Washington, DC, 1991; R. P. Turco, *Earth Under Siege: From Air Pollution to Global Change*, Oxford University Press, New York, 1997.

Smoke

A dispersion of small, solid particles and liquid droplets suspended in a gaseous medium, usually produced by a combustion process. Smoke particles start as groups of carbon atoms or small fragments of the burning material which grow in size by coagulation and condensation as they are carried from the combustion zone by the buoyancy of the hot gases. These particles are usually found to have diameters in the range 0.01–5.0 micrometers. The solid particles tend to be somewhat irregular in shape and are often seen as clusters or long chains of dark material. Smokes from flaming combustion consist mainly of solid particles. *See* COMBUSTION.

Liquid particles are always spherical and consist of water and high-molecular-weight compounds which have condensed on a small, solid nucleus. These smokes usually appear gray or white but may also appear to be lightly colored by the condensed compounds. Smokes produced by pyrolysis or smoldering combustion are largely made up of liquid particles.

Smokes can also be produced by chemical reactions or by mechanical generators such as atomizers or nebulizers. Chemical smokes use such compounds as phosphorus pentoxide (P_2O_5), ammonium chloride (NH_4Cl), titanium tetrachloride ($TiCl_4$), or aluminum chloride ($AlCl_3$). Mechanically generated smokes can be produced from various oils such as dioctylpthalate, corn oil, or silicones, or by atomizing salt solutions (for example, sodium chloride or potassium chloride) or sugar solutions (glucose or sucrose), leaving solid particles of a fairly uniform size after evaporation of the surrounding liquid.

These chemical smokes are used for such things as airflow visualization, theatrical effects, vision obscurants for crowd control or by the military, and daylight signaling devices. Mechanically generated smokes are often used for measuring the efficiency of filters because their sizes are fairly well defined by the material and production technique employed. *See* AIR FILTER; SCREENING SMOKE. Richard W. Bukowski

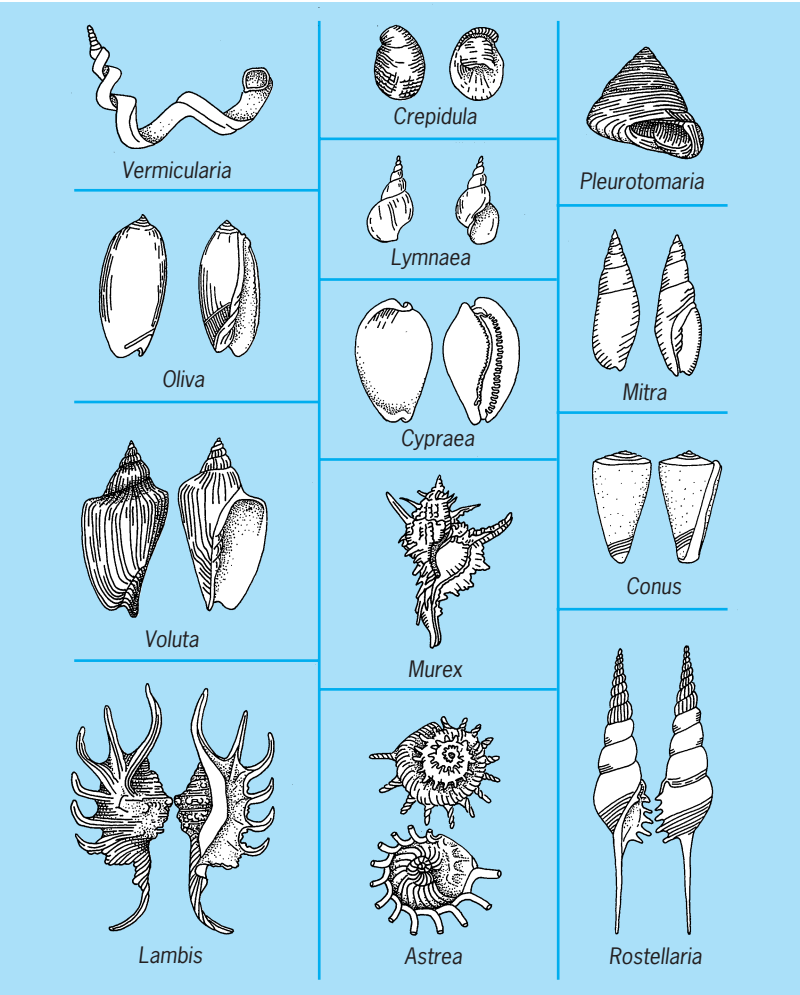
Bibliography. J. K. Beddow (ed.), *Particle Characterization in Technology*, 2 vols., 1984; R. Clift, *Particulate Technology*, 1991; H. K. Hasegawa (ed.), *Characterization and Toxicity of Smoke*, 1990.

Snail

Any of the approximately 74,000 species in the class Gastropoda of the phylum Mollusca or, alternatively, any of the 12 or so species of land pulmonate gastropods used as human food.

General characteristics. The shell of snails is in one piece and typically turbinate (see **illus.**), but may be planospiral or limpet-shaped, or may be secondarily lost (as in land slugs and marine nudibranchs). In development, gastropods have undergone torsion (the visceral mass and the mantle-shell covering it have become twisted through 180° in relation to the head and foot) so that the mantle cavity is placed anteriorly above the head.

A few primitive snails retain two aspidobranch (featherlike) gills (the characteristic ctenidia which are structurally and functionally homologous throughout the Mollusca). Most gastropods show considerable asymmetry and simplification with reduction to a single aspidobranch ctenidium or, more often, to a pectinibranch ctenidium (or comblike "half-gill"). This is accompanied by corresponding changes that reduce paired internal organs to a single kidney, a single gonad, and a single cardiac auricle. True ctenidia are lost in most of the shellless marine nudibranchs, and in the fresh-water and



Diversity of snail shells. Except for *Lymnaea* (a fresh-water pulmonate), all genera represented are marine prosobranchs. (*After R. R. Shrock and W. H. Twenhofel, Principles of Invertebrate Paleontology, 2d ed., McGraw-Hill, 1953*)

terrestrial snails of the subclass Pulmonata, in which the mantle cavity is an air-breathing lung.

In feeding, all snails use a characteristic rasping tongue or radula. This is a chitinous ribbon bearing teeth which is moved over a supporting protrusible "tongue" (the odontophore) with a to-and-fro action. The radular apparatus has a twofold function: it serves both for rasping off food material (mechanically like an inverted version of the upper incisor teeth of a beaver) and for transporting the food back into the gut like a conveyor belt. Different patterns of radular tooth rows are found in the major gastropod groups; in general, carnivorous snails have fewer (but more differentiated) teeth per row.

The class Gastropoda is divided into three somewhat unequal subclasses, the Prosobranchia, Opisthobranchia, and Pulmonata. The Prosobranchia, made up of marine gastropods with a few nonmarine forms, are the largest and most diverse. The Opisthobranchia, entirely marine, consist largely of shell-less sea slugs and the highly colorful nudibranchs. The Pulmonata are basically air-breathers that use the mantle cavity as a lung, and include the most highly evolved gastropods, with representatives on land, in fresh waters, and (by secondary reversion) on the seashore. *See* OPISTHOBRANCHIA; PROSOBRANCHIA; PULMONATA.

Both fresh-water and terrestrial snail species serve as vectors (first or sole intermediate hosts) in the transmission of flukes (Trematoda) infecting humans or domestic animals. For further discussion of these parasitic diseases *see* DIGENEA; SCHISTOSOMIASIS.

Edible snails. Although many marine gastropods are eaten by humans, including periwinkles (*Littorina*), moon snails (*Polinices*), and "conchs" of various genera, the "snails" or *Schnecken* or *escargots* of gastronomy belong to a very few species of larger pulmonate land snails in the family Helicidae. Certain species belonging to genera such as *Cepaea*, *Otala*, and *Helix* are widespread in many continents as a result of human cultural migrations. All are capable of long periods of dormancy (with the shell aperture sealed by a series of membranes, the epiphragm, and metabolism reduced over twentyfold), and this facilitates human transport.

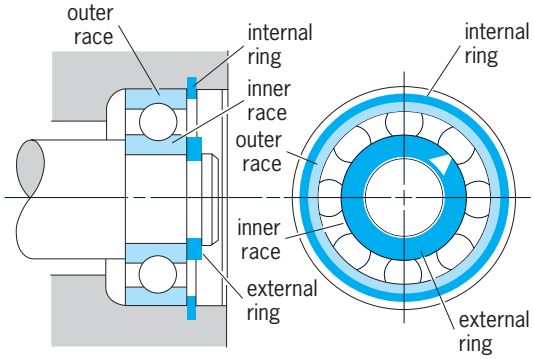
Two such European land snails, *C. hortensis* and *C. nemoralis*, are widespread and abundant. Recent work on their population genetics includes some of the most convincing demonstrations of natural selection in action. Both species are markedly polymorphic in shell colors and patterns. The basic shell color may be bright yellow, dull brown, or any shade from very pale fawn through pink and orange to red. Up to five longitudinal dark bands may be present, and all possible combinations of presence or absence of them, and of fusion between adjacent bands, have been described. All of these variables are genetically determined.

Most natural populations of *Cepaea* consist of two or more different varieties living together. The proportions of varieties in each colony correlate with the general class of background in the habitat where the snails are living. Predators include songbirds and small mammals, and in some cases it is possible to have direct quantification of selection pressures by comparing the proportions of each variety eaten to the proportions present in the snail colony. Considerable efforts of population geneticists are being directed to the elucidation of other factors involved in maintenance of this complex polymorphism in natural populations of *Cepaea*. *See* GASTROPODA; MOLLUSCA; POLYMORPHISM (GENETICS).

W. D. Russell-Hunter

Snap ring

A form of spring used principally as a fastener. Piston rings are a form of snap ring used as seals. The ring is elastically deformed, put in place, and allowed to snap back toward its unstressed position into a groove or recess. The snap ring may be used externally to provide a shoulder that retains a wheel or a bearing race on a shaft, or it may be used internally to provide a construction that confines a bearing race in the bore of a machine frame (see **illus.**). The size of the ring and its recess determines its strength under



Snap rings hold ball-bearing race in place. Internal ring supports axial thrust of axle. External ring aligns inner race against shoulder of shaft.

load. Sufficient clearance and play is needed in the machine so that the ring can be inserted and seated. *See* ANTIFRICTION BEARING; INTERNAL COMBUSTION ENGINE; SPRING (MACHINES). L. Sigfred Linderoth, Jr.

Bibliography. H. Carlson, *Spring Designers Handbook*, 1978.

Snow

Frozen precipitation resulting from the growth of ice crystals from water vapor in the Earth's atmosphere.

Formation and growth. As ice particles fall out in the atmosphere, they melt to raindrops when the air temperature is a few degrees above 32°F (0°C), or accumulate on the ground at colder temperatures. At temperatures above −40°F (−40°C), individual crystals begin growth on icelike aerosols (often clay particles 0.1 micrometer in diameter), or grow from cloud droplets (10 μm in diameter) frozen by similar particles. At lower temperatures, snow crystals grow on cloud droplets frozen by random molecular motion. At temperatures near 25°F (−4°C), crystals sometimes grow on ice fragments produced during soft hail (graupel) growth. Snow crystals often grow in the supersaturated environment provided by a cloud of supercooled droplets; this is known as the Bergeron-Findeisen process for formation of precipitation. When crystals are present in high concentrations (100 particles per liter) they grow in supersaturations lowered by mutual competition for available vapor. **Figure 1** shows the vapor pressure of water and ice and vapor pressure excess of supercooled water over ice at different temperatures. The difference has a maximum near 9.5°F (−12.5°C).

Ice crystals growing under most atmospheric conditions (air pressure down to 0.2 atm or 20 kilopascals and temperatures 32 to −58°F or 0 to −50°C) have a hexagonal crystal structure, consistent with the arrangement of water molecules in the ice lattice, which leads to striking hexagonal shapes during vapor growth. The crystal habit (ratio of growth along and perpendicular to the hexagonal axis) changes dramatically with temperature. Both field and laboratory studies of crystals grown under known or controlled conditions show that the crystals are platelike above 27°F (−3°C) and between 18 and −13°F (−8 and −25°C), and columnlike between 27 and 18°F (−3 and −8°C) and below −13°F (−25°C). Aspect ratios can be large—hexagonal plates being 50 μm thick and several millimeters across, and hexagonal columns being 100 μm across and several millimeters long. Individual crystals fall in the atmosphere at velocity up to 0.5 m s^{-1} (1.6 ft s^{-1}). As crystals grow, they fall at higher velocity, which leads, in combination with the high moisture availability in a supercooled droplet cloud, to sprouting of the corners to form needle or dendrite skeletal crystals. These are more evident for growth

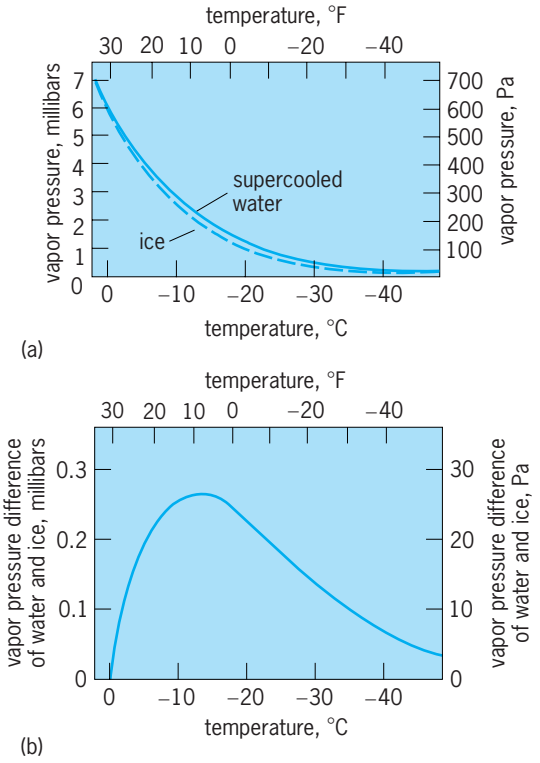


Fig. 1. Graphs showing variation of vapor pressure with temperature. (*a*) Vapor of supercooled water and ice. (*b*) Vapor pressure difference between supercooled water and ice. Maximum occurs at 9.6/12.5°F (−12.5°C).

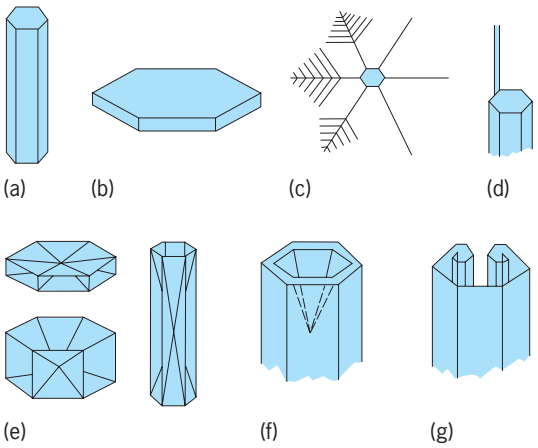


Fig. 2. Different forms of snow crystals. (*a*) Prism. (*b*) Plate. (*c*) Dendrite. (*d*) Needle. (*e*) Hopper-type crystals. (*f*) Hollow prism. (*g*) Scroll.

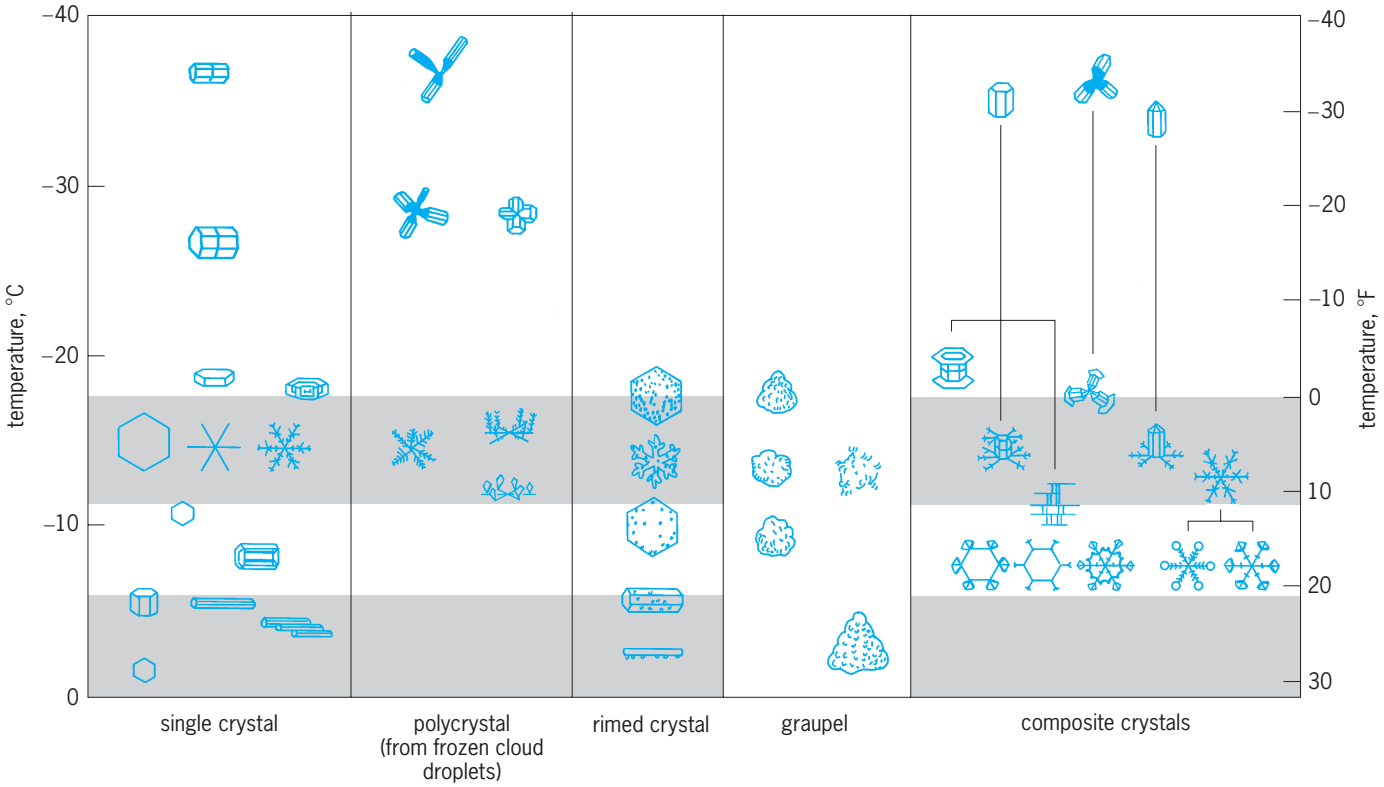


Fig. 3. Diagram showing forms of snow crystals and the temperature zones in which they occur (shaded areas). Lines indicate paths in which one type of composite crystal grows into another. (*After J. Hallett, How snow crystals grow, Amer. Sci., 72:582–589, November–December 1984*)

between 25 and 23°F (−4 and −5°C) and 10 and 3°F (−12 and −16°C) [**Fig. 2**].

As crystals fall through the atmosphere into levels of different temperature and moisture availability, their growth form changes with the environmental conditions; crystals of composite growth form (plates or dendrites with columns at the corners, a column with plates or dendrites on the end) demonstrate the subtleties of these changes. When crystals grow on cloud droplets frozen at low temperature (below 14 to −4°F or −10 to −20°C), crystals of several orientations grow radially outward to give a radial array of columns, plates, or dendrites (**Fig. 3**).

Through its shape, a snow crystal represents a record of the temperature–moisture–fall velocity history of its growth (**Fig. 4**). Two crystals that appear visually identical would have to have an identical origin and to follow an identical path through the atmosphere to reproduce temperature, moisture, and fall velocity. In view of the turbulent nature of the atmosphere, the likelihood of actually observing two crystals that followed an identical path is quite remote.

Recent concerns about global warming resulting from carbon dioxide (CO_2) production by the combustion of fossil fuel have emphasized the importance of high-level ice clouds (cirrus) forming at some 10–15 km (6–9 mi) in the atmosphere, both for their role in scattering back sunlight to space and thermal radiation back to the lower atmosphere and the ground. Such clouds are composed of complex

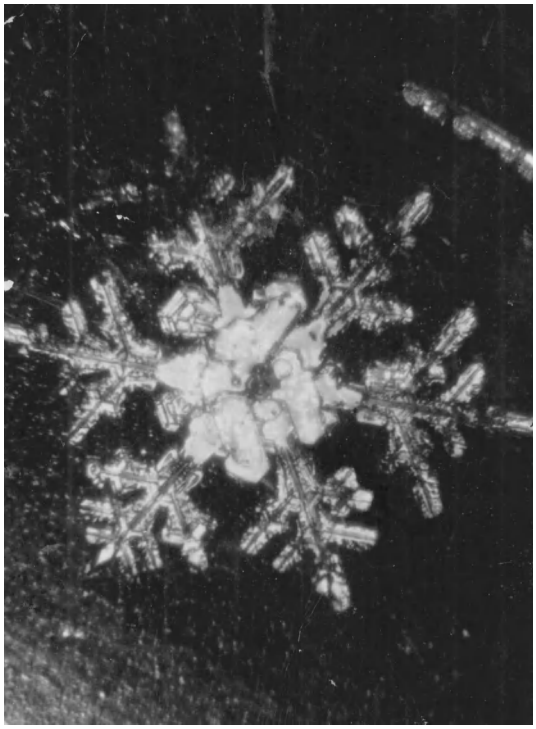


Fig. 4. Dendrite snow crystal, showing that its initial growth (center) began as a plate.

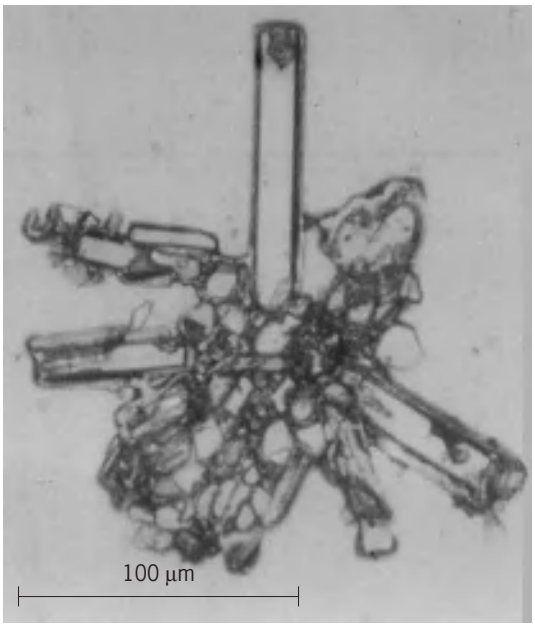


Fig. 5. Replica of an ice crystal collected by an aircraft during flight through a cirrus cloud. It is in the form of a rosette of individual columns, suggesting that the nucleating droplet froze as a polycrystal.

column-type crystals togather with plates (**Fig. 5**). The relative proportion depends on the nucleation processes (**Fig. 6**; cloud droplets may freeze as a single crystal or polycrystal) and the supersaturation as determined by the release of water vapor by local updraft and the removal by the growing crystals—which in turn depends on the crystal concentration. The effect on radiation depends either on crystal thickness or on crystal area, which depends in turn on the absorption at different wavelengths and the size and shape which influences scattering. Satellite measurements of temperatures of these clouds, used as a forecast tool, require interpretation in terms of the properties of high-level crystal cirrus shields. *See* GREENHOUSE EFFECT; WEATHER MODIFICATION.

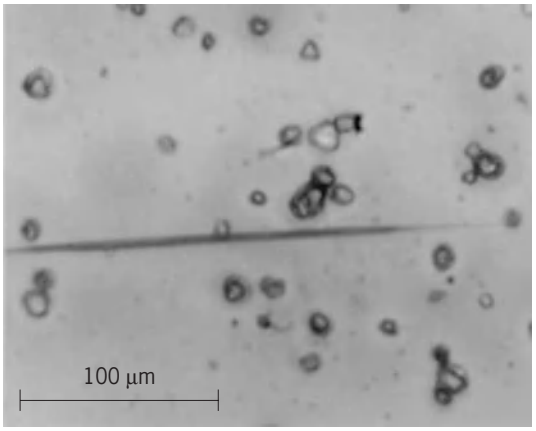


Fig. 6. Replica of crystals collected in high-level cirrus at about −50°C (−122°F) showing the simultaneous presence of hexagonal and triangular plates, a long column, and also more complex but equiaxed crystals. The habit variability suggests sequential nucleation events and growth under varying conditions. As such crystals fall to lower levels where vapor is available for growth, they form the core for larger snow crystals.

Frost crystals. Frost crystals differ from snow crystals in that the latter grow on a fixed surface and are ventilated by ambient wind. The frost crystals usually fail to have the exquisite symmetry of falling snow crystals since shielding of one side of the growing crystal often occurs at its point of attachment.

Crystal orientation. Small crystals tens of micrometers across are usually randomly oriented during their fall, but as they grow they become oriented by asymmetric flow around the falling particle, ultimately with their longest dimension horizontal by lee eddies formed by the airflow around the crystal. Larger crystals, greater than a few millimeters across (depending on shape and density), oscillate as they fall along a helical path as these eddies are shed. These orientations lead to complicated optical effects, often colored, such as haloes and arcs in sunlight and sometimes moonlight. *See* HALO.

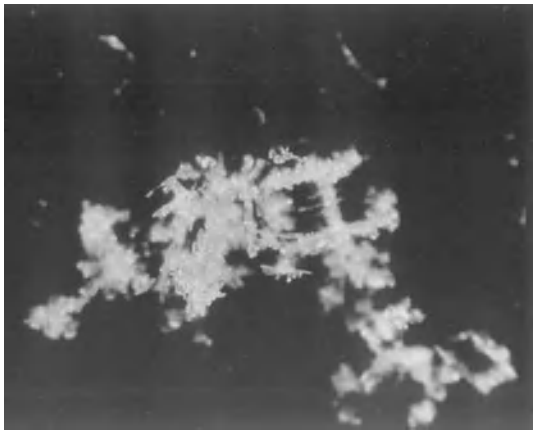


Fig. 7. Snowflake made up of several hundred individual crystals.

Snowflakes. Under some conditions crystals aggregate to give snowflakes. This happens for the dendritic crystals that grow near 5°F (−15°C), which readily interlock if they collide with each other, and for all crystals near 32°F (0°C). Snowflakes typically contain several hundred individual crystals (**Fig. 7**) and fall at 1–2 m s^{-1} (3–6 ft s^{-1}). Casual observation of falling snowflakes often reveals a complex path, the flake oscillating and spinning, which results from the shedding of eddies as the flake falls. The shape of the flake tends to be planar (rather than spherical), which represents molding by aerodynamic forces as the flake falls through the air.

Snowpack. When snow reaches the ground, changes take place in the crystals leading to properties of the snowpack that are important for skiing conditions and avalanche formation. At temperatures near 32°F (0°C) the crystals rapidly lose the delicate structure acquired during growth, sharp edges evaporate, and the crystals take on a rounded shape, some 1–2 mm (0.04–0.08 in.) in diameter. These grains sinter together at their contact points to give snow some

structural rigidity. The specific gravity varies from ~0.05 for freshly fallen "powder" snow to ~0.4 for an old snowpack. At lower temperatures, crystals retain their shape and the snowpack retains a low density, unless compressed at depth. Strong vertical gradients of temperature in the snowpack caused by low air temperatures above the pack lead to an evaporation-growth process in the pack itself. Crystals grow in a skeletal form as hollow columns (hopper crystals; Fig. 2*f*). These forms lead to low density and planes of weakness in the pack, which lead to avalanche release under high loading rates. For this to occur most readily, a strong vertical gradient of temperature (~2.3°F/in. or 0.5°C/cm) must coincide with the 25°F (−4°C) isotherm in the snowpack. A similar effect occurs when frost crystals have grown on the snowpack from an overlying fog near this temperature and are subsequently buried by a snowfall. *See* AVALANCHE.

Snowfall rates tend to be maximum in association with adequate supplies of moisture from a warm winter storm (surface temperature near 32°F or 0°C), and are enhanced when associated with mountain ranges where air rising over the mountain makes additional moisture available. These combinations give typical snowfall rates of 0.5 cm h^{-1} (0.2 in. h^{-1}) water equivalent. Higher rates of ice precipitation usually are associated with soft hail formed by riming of falling snow particles as they collide with these supercooled cloud droplets during their fall. *See* CLOUD PHYSICS; HAIL; PRECIPITATION (METEOROLOGY).

Growth mechanism and crystal shape. Crystals grow in different shapes as a result of interplay of two processes, namely heat-moisture transport between the growing crystal and its environment, and the molecular processes that take place at the crystal-air interface.

At low supersaturations (lower than usually occur for extended periods in the atmosphere) crystalline defects (departures from regular molecular packing) may be necessary for growth of specific crystal faces of high molecular density. By contrast, under most atmospheric conditions growth of faces proceeds by nucleation of layers at corners of a growing crystal which grow back toward the face center. In the case of a plate, as the layer grows out a new one is nucleated and the process repeats; in the case of a dendrite, growing in the presence of higher vapor content or at higher ventilation, layers grow only a short distance before a new one is nucleated so complete faces do not form. The reason for columns or plates is related to the properties of water molecules absorbed on the different faces, giving rise to different rates of nucleation and layer growth. Confirmation of these differences awaits development of laboratory techniques such as ellipsometry for studying the properties of a surface with a high vapor pressure, for such studies cannot make use of high vacuum techniques utilized in studies of similar crystals of technological importance. *See* CRYSTAL GROWTH; INTERFACE OF PHASES.

John Hallett

Bibliography. W. A. Bentley and W. J. Humphreys, *Snow Crystals*, 1931, reprint 1962; J. R. Eagleman, *Meteorology: The Atmosphere in Action*, 2d ed., 1992; J. Hallett, How snow crystals grow, *Amer. Sci.*, 72:582–589, November–December 1984; P. V. Hobbs, *Ice Physics*, 1975; V. Keller and J. Hallett, Influence of air velocity on the habit of ice crystal growth from the vapor, *J. Crys. Growth*, 60(11):91–106, 1982; C. Magono and C. Lee, Meteorological classification of natural snow crystals, *J. Fac. Sci.*, Hokkaido Univ., ser. 7, no. 2, pp. 321–335, 1966; U. Nakaya, *Snow Crystals*, 1954.

Snow line

A term generally used to refer to the elevation of the lower edge of a snow field. In mountainous areas, it is not truly a line but rather an irregular, commonly patchy border zone, the position of which in any one sector has been determined by the amount of snowfall and ablation. These factors may vary considerably from one part to another. In regions where valley glaciers descend to relatively low elevations, the summer snow line on intervening rock ridges and peaks is often much higher than the snow line on the glaciers, and in most instances it is more irregular and indefinite. If by the end of summer it has not disappeared completely from the bedrock surfaces, the lowest limit of retained snow is termed the orographical snow line, because it is primarily controlled by local conditions and topography. On glacier surfaces it is sometimes referred to as the glacier snow line or névé line (the outer limit of retained winter snow cover on a glacier).

Year-to-year variation in the position of the orographical snow line is great. The mean position over many decades, however, is important as a factor in the development of nivation hollows and protalus ramparts in deglaciated cirque beds. The average regional level of the orographical snow line in any one year is the regional snow line. Since the regional snow line is controlled entirely by climate and not influenced by local conditions, glaciologists sometimes call it the climatological snow line. The average position of the climatological snow line over a decade or more may be termed the mean climatological snow line. On broad ice sheets and glaciers, the climatological snow line is the same as the névé line. The average position of this, over a decade or more of observation, may be termed the mean névé line. *See* GLACIOLOGY; SNOWFIELD AND NÉVÉ.

Maynard M. Miller

Snow surveying

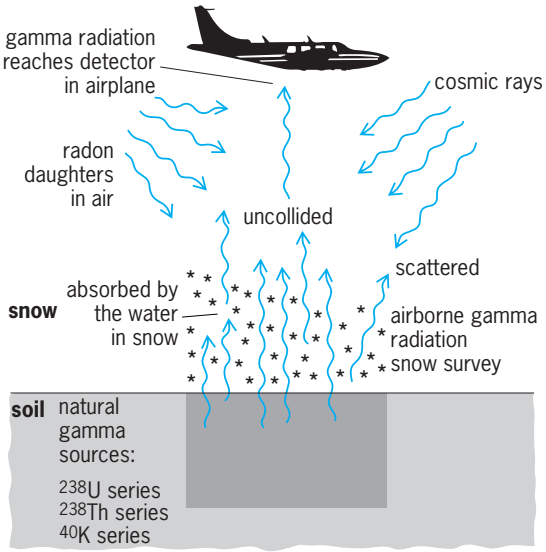
A technique for providing an inventory of the total amount of snow covering a drainage basin or a given region. Most of the usable water in western North America originates as mountain snowfall that accumulates during the winter and spring and appears several months later as streamflow. Snow surveys were established to provide an estimate of the snow water equivalent (that is, the depth of water

produced from melting the snow) for use in predicting the volume of spring runoff. They are also extremely useful for flood forecasting, reservoir regulation, determining hydropower requirements, municipal and irrigation water supplies, agricultural productivity, wildlife survival, and building design, and for assessing transportation and recreation conditions.

Conventional snow surveys are made at designated sites, known as snow courses, at regular intervals each year throughout the winter period. A snow course consists of a series of 5–10 sampling points, spaced 50–100 ft (15–30 m) apart, along a permanently marked and mapped route. A snow sampler is used to measure the snow depth and water equivalent at each point. Average depth and water equivalent are calculated for each snow course. There are about 3000 snow courses operated throughout North America.

Accurate estimates of snow water equivalent over a basin can be obtained only after allowances are made for instrument, observation, or siting biases. Snow samplers in common use overmeasure by up to 10%. Accurate basin snow cover estimates of depth, density, and water equivalent are best obtained from a snow course network or survey sampling scheme specifically designed to represent local vegetation and terrain variations. This type of snow survey could differ from a conventional snow course in the number and spacing of sample points.

A viable alternative snow survey procedure is the airborne gamma-ray technique. Gamma rays are emitted by the natural radioactive decay of uranium, thorium, and potassium in the ground. Since water, in either liquid or solid phase, attenuates gamma radiation, snow pack water equivalent along a flight line can be calculated based on the attenuation of the radiation signal by the snowpack compared to the condition during no snow cover (see **illus**.). Corrections for soil moisture and for any contribution of atmospheric radon gas and cosmic rays to the measured signal are required. Typically, the aircraft flies along a 12-mi (20-km) flight line at 500-ft (150-m) altitude measuring the radiation signal over a 1000-ft-wide (300-m) swath. Each line provides a measured areal average snow water equivalent over a 2.3-mi^2 (6-km^2) area. The airborne gamma method has been used operationally in North America, Scandinavia, and Russia, mainly in shallow snowpack regions.



Snow survey using airborne gamma-radiation method. (*Courtesy of T. R. Carroll*)

Satellite remote sensing and data relay are technologies used to obtain information on snow cover in more remote regions. Satellite imagery is used to map snow-covered areas; in the future, passive microwave sensors may provide information on snow depth and water equivalent. Other communications technologies (satellite data retransmission, meteor burst) permit real-time access of measurements by automatic devices, such as snow pillows or snow depth sensors, reducing the need to travel to more remote sites to conduct snow surveys. *See* GAMMA-RAY DETECTORS; REMOTE SENSING; SURFACE WATER.

Barry E. Goodison

Bibliography. T. R. Carroll, J. E. Glynn, and B. E. Goodison, A comparison of U.S. and Canadian airborne snow water equivalent measurements, *Proceedings of the Western Snow Conference*, 51st Annual Meeting, Vancouver, Washington, pp. 27–37, 1983; B. E. Goodison (ed.), Hydrological applications of remote sensing, *Proceedings of the Hamburg Symposium, August 1983*, IAHS Publ. 145, 1985; D. M. Gray and D. H. Male (eds.), *Handbook of Snow*: *Principles, Processes, Management and Use*, 1981; E. L. Peck, T. R. Carroll, and S. C. Van Demark, Operational aerial snow surveying in the United States, *Hydrol. Sci. Bull.*, 25:51–62, 1980.

Snowfield and névé

The term snowfield is usually applied to mountain and glacial regions to refer to an area of snow-covered terrain with definable geographic margins. Where the connotation is very general and without regard to geographical limits, the term snow cover is more appropriate; but glaciology requires more precise terms with respect to snowfield areas. These terms differentiate according to the physical character and age of the snow cover. Technically, a snowfield can embrace only new or old snow (material from the current accumulation year). Anything older is categorized as firn or ice. The word firn is a derivative of the German adjective *fern*, meaning "of last year," and hence refers to hardened snow not yet metamorphosed to ice which has been retained from the preceding year or years. Thus, by definition, a snowfield composed of firn can be called a firn field. Another term familiar to glaciologists is névé, from the French word for a mass of hardened snow in a mountain or glacier environment. In English, rather than a specific word for the material itself, a descriptive phrase is used such as: "consolidated granular snow not yet changed to glacier ice." Because of

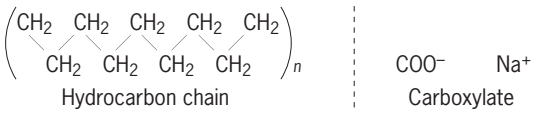
the need for simple terms, however, it is acceptable to use the French term *névé* when specifically referring to a geographical area of snowfields on mountain slopes or glaciers (that is, an area covered with perennial "snow" and embracing the entire zone of annually retained accumulation). For reference to the compacted remnant of the snowpack itself, that is, the material which is retained at the end of an annual melting period on a snowfield or névé, it is appropriate to use the derivative German term firn. *See* GLACIATED TERRAIN; GLACIOLOGY.

Maynard M. Miller

Bibliography. D. I. Benn, *Glaciers & Glaciation*, 1998; S. C. Colbeck, *Dynamics of Snow and Ice Masses*, 1980; W. D. Kingery, *Ice and Snow: Properties, Processes and Applications*, 1963.

Soap

A cleansing agent described chemically as an alkali metal salt of a long-carbon-chain monocarboxylic acid, represented for example by sodium myristate ($NaCOOC_{13}H_{27}$), whose structure is shown below.



The hydrocarbon portion is hydrophobic and the carboxylate portion is hydrophilic.

For detergency purposes, the most useful hydrophobic portion contains 12–18 carbon atoms. Sodium soaps are not useful for soil removal with hydrocarbon chains of less than 12 carbons. When the chain length exceeds 18 carbons, they become insoluble in water.

Manufacturing process. Historically, soap was made from animal fats and wood ashes (which contain the alkali potassium carbonate). The Romans were doing this 2500 years ago, and a similar process was used in colonial America. In 1791, the Leblanc process for preparing sodium hydroxide was developed, making possible a soap-making process based on an alkaline hydrolysis reaction, saponification, according to the reaction below, where R represents the hydrocarbon chain.

$$\underset{\text{Fat}}{C_3H_5(OOCR)_3} + \underset{\text{Caustic soda}}{3NaOH} \rightarrow \underset{\text{Soap}}{3NaOOCR} + \underset{\text{Glycerin}}{C_3H_5(OH)_3}$$

Until the late 1930s, soap was made in huge kettles into which fats, oils, and caustic were piped and heated until boiling vigorously. After cooling for several days, tons of salt were added, causing the mixture to separate into two layers with the neat soap on top and the lye on the bottom. One kettle could produce carload quantities of soap. The product was pumped into giant mixers called crutchers where builders, perfumes, and other desired ingredients were added. (Builders are alkaline compounds which improve the cleaning performance of the detergent.) Finally, the soap was rolled into flakes, cast or milled into bars, or spray-dried into soap powder.

Most soap is now made by a continuous process. Fats and oils may be converted directly to soap by the reaction with caustic, and the neat soap separated by a series of centrifuges or by countercurrent washing. The saponification may be carried out with heat, pressure, and recirculation. Other important modern processes hydrolyze the fats directly with water and catalysts at high temperatures. This permits fractionation of the fatty acids, which are neutralized to soap in a continuous process. Advantages for this process are close control of the soap concentration, the preparation of soaps of certain chain lengths for specific purposes, and easy recovery of the by-product glycerin. After the soap is recovered, it is pumped to the crutcher and treated just as the product from the kettle process. Tallow and coconut oil are the most common fatty materials used for soapmaking.

Characteristics. Because of the marked imbalance of polarity within the molecules of synthetic detergents and soaps, they have unusual surface and solubility characteristics. The feature of their molecular structure which causes these properties is the location of the hydrophilic function at or near the end of a long hydrocarbon (hydrophobic) chain. One part of the molecule is water-seeking while the other portion is oil-seeking. Thus these molecules concentrate and orient themselves at interfaces, such as an oil-solution interface where the hydrophobic portion enters the oil and the hydrophilic portion stays in the water. Consequently, the interfacial tension is lowered and emulsification can result. At an agitated air-solution interface, the excess detergent concentration leads to sudsing.

Considerable research has also been done on the aqueous solutions of soap. Many of the phase characteristics now known to occur in most surfactant-water systems were first discovered in soap systems. These include micelle formation where, at a critical concentration, soap molecules in solution form clusters (micelles). The hydrocarbon chains associate with each other in the interior, and the polar groups are on the outside. Since the interior of the micelle resembles a hydrocarbon, materials such as oil-soluble dyes, which are insoluble in water, will frequently dissolve in a detergent solution containing micelles. At higher concentrations of soap, two different liquid crystalline phases form. Around 50% water–50% soap "middle" phase occurs, and is believed to form when soap molecules coalesce into long chains or rods in a hexagonal array. It is a very sticky, viscous material which must be avoided in the manufacture of soap. At still higher concentrations (30% water–70% soap) neat phase occurs where the soap molecules form extensive bilayer sheets. Although viscous, neat soap can be pumped and is the phase in which soap is recovered after the saponification. *See* MICELLE; SURFACTANT.

There is one serious disadvantage to the general use of soap which has caused its replacement by synthetic detergents for home laundering. The problem is that carboxylate ions of the soap react

with the calcium and magnesium ions in natural hard water to form insoluble materials called lime soap. In areas having a hard water supply, this is the floating curd which appears in the washbowl and as bathtub ring. In laundry products, attempts were made to overcome the problem by adding materials such as sodium carbonate to "build" the action of soap by precipitating calcium and magnesium carbonates. However, both the carbonates and the soap curd deposited on the fabrics being washed, and gradually caused them to become dull, gray, and sometimes stiff.

Uses. Bar soap for personal bathing is the greatest remaining market for soap in the United States, Canada, and western Europe. However, in the developing countries, the detergent industry is still dominated by soap. Some commercial laundries having soft water continue to use soap powders. Metallic soaps are alkaline-earth or heavy-metal long-chain carboxylates which are insoluble in water but soluble in nonaqueous solvents. They are used as additives to lubricating oils, in greases, as rust inhibitors, and in jellied fuels. *See* DETERGENT.

Roy C. Mast

Bibliography. *Kirk-Othmer Encyclopedia of Chemical Technology*, 3d ed., vol. 21, 1983; P. L. Layman, Brisk detergent activity changes picture for chemical suppliers, *Chem. Eng. News.*, 62(4):40, 1984; W. C. Preston, Some correlating principles of detergent action, *J. Phys. Colloid. Chem.*, 52:84, 1948; E. G. Thomssen and J. W. McCutcheon, *Soaps and Detergents*, 1949.

Soapstone

A soft talc-rich rock. Soapstones are rocks composed of serpentine, talc, and carbonates (magnesite, dolomite, or calcite). They represent original peridotites which were altered at low temperatures by hydrothermal solutions containing silicon dioxide, SiO_2; carbon dioxide, CO_2; and other dissolved materials (products of low-grade metasomatism). Among the rock products thus formed are antigorite schists, actinolite-talc schists, and talc-carbonate rocks. To the last belongs the true soapstone, but the whole group of rocks may loosely be referred to as soapstones because of their soft, soapy consistency. Such rocks were selected by prehistoric humans for making primitive vessels and pots, and for making rough carvings for ornamental purposes. *See* TALC.

T. F. W. Barth

Social hierarchy

A fundamental aspect of social organization that is established by fighting or display behavior and results in a ranking of the animals in a group. Social, or dominance, hierarchies are observed in many different animals, including insects, crustaceans, mammals, and birds. In many species, size, age, or sex determines dominance rank. Dominance hierarchies often determine first or best access to food, social interactions, or mating within animal groups.

Dominance and agonistic behavior. When two animals fight, several different behavioral patterns can be observed. Aggressive acts (such as biting, kicking, or stabbing with horns or antlers) and submissive acts (such as a dog rolling on its back and exposing its stomach) are both parts of a fight. Aggression and submission, together, are known as agonistic behavior.

Agonistic encounters often result in a winner, or dominant animal, and a loser. The losing individual may simply retreat or leave the area, so that it is not part of the winner's social group. If the loser stays in the vicinity of the winner, it usually displays submissive behavior when the winner approaches in order to prevent further attacks. The submissive individual is subordinate in the relationship.

An agonistic relationship in which one animal is dominant and the other is submissive is the simplest type of dominance hierarchy. In nature, most hierarchies involve more than two animals and are composed of paired dominant-subordinate relationships. For example, if a young female bighorn sheep enters a herd of other females, she will have agonistic encounters with other females and will establish dominant-subordinate relationships with the other herd members that determine her rank in the group.

The simplest dominance hierarchies are linear and are known as pecking orders. In such a hierarchy the top individual (alpha) dominates all others. The second-ranked individual (beta) is submissive to the dominant alpha but dominates the remaining animals. The third animal (gamma) is submissive to alpha and beta but dominates all others. This pattern is repeated down to the lowest animal in the hierarchy, which cannot dominate any other group member. Linear hierarchies can be observed in chickens.

Other types of hierarchies result from variations in these patterns. If alpha dominates beta, beta dominates gamma, but gamma dominates alpha, a dominance loop is formed. In some species a single individual dominates all members of the social group, but no consistent relationships are formed among the other animals. In newly formed hierarchies, loops or other nonlinear relationships are common, but these are often resolved over time so that a stable linear hierarchy is eventually observed.

Hierarchies in social insects. In some social wasps and bees the queen has agonistic relationships with her workers, and a linear dominance hierarchy is established. In paper wasps (*Polistes* sp.) the queen lunges at her workers and opens her mandibles in a threatening manner; in extreme cases, biting, grappling, and stinging occur. The workers submit by crouching down to the comb. The dominance hierarchy assures that the queen lays most of the eggs in the colony, while the workers build comb and bring food to the queen and her offspring. Males are not a part of the societies of stinging social insects (ants, bees, and wasps) and do not participate in these hierarchies.

In some of the more advanced social insects, such as the honeybee, the queen uses chemical signals to dominate her workers, rather than aggressive behavior. If a honeybee queen dies or is removed from the colony, the workers are able to lay eggs; fights and agonistic behavior over dominance then occur among the workers. *See* CHEMICAL ECOLOGY; SOCIAL INSECTS.

Dominance and mating. Males often fight over access to females and to mating with them. Male dominance hierarchies are seen in many hooved mammals (ungulates). Horns or antlers are usually adaptations used in agonistic encounters among males, and body size is often the determinant of the dominant male. Males may come together in bachelor herds during the nonmating season; when the females are ready to mate, agonistic behavior among the males intensifies and the dominant male is determined. *See* REPRODUCTIVE BEHAVIOR.

Dominance and feeding. Herds of females use dominance hierarchies to determine access to food. A female occupying a favored feeding site will yield to a dominant female but will displace a subordinate female. Agonistic interactions among females are often not as overtly aggressive as those among males, but the effects of the dominance hierarchy can easily be observed by recording which female is able to displace which others. In female dairy cattle, the order of entry into the milking barn is determined by dominance hierarchy, with the alpha female entering first.

Dominance hierarchies occur among very young siblings in many species; dominance determines which nipple is available for suckling. Piglets prefer the foremost nipples, which produce the most milk. Piglets relegated to the hindmost nipples receive less food and grow more slowly. Similar nipple-order dominance appears among domestic kittens.

Genetic factors. Because dominant animals may have advantages in activities such as feeding and mating, they will have more offspring than subordinate animals. If this is the case, then natural selection will favor genes for enhanced fighting ability; a striking example is seen in males with armament used in fighting for access to mates. Large horns or antlers will be be favored by natural selection, as will ferocity in fighting. In some cases, such selection results in seemingly bizarre or disproportionate weapons. However, practicalities, such as ability to move in the environment or to evade predators, may place a limit on such selection. Heightened aggressive behavior may be counterselected by the necessity for amicable social interactions in certain circumstances, such as parent-offspring relationships.

Artificial selection has been used by humans to raise (pit bull terriers, fighting cocks) or to lower (most breeds of dogs, horses, cattle) fighting tendencies and dominance behavior. Domesticated animals are generally easier to handle and manipulate if they are less likely to fight and injure one another or human handlers. They may also be kept under more crowded conditions if fighting tendencies are reduced. Selection for passivity toward humans usually results in relaxation of dominance within animal social groups. *See* BREEDING (ANIMAL).

Complex social hierarchies. Many higher primates live in large groups of mixed sex and exhibit complex social hierarchies. In these groups, intra- and intersexual dominance relationships determine many aspects of group life, including feeding, grooming, sleeping sites, and mating. In rhesus macaques the rank of an individual's mother is very important in determining rank. The cultural and genetic history of a macaque may be more important in determining social rank than its body size. Macaque, baboon, and chimpanzee societies are characterized by cooperative alliances among individuals that are more important than individual fighting ability in maintaining rank.

Significance of hierarchies. Social hierarchies provide a means by which animals can live in groups and exploit resources in an orderly manner. In particular, food can be distributed among group members with little ongoing conflict. Although social hierarchies can be viewed as a group characteristic, they have evolved as a result of the effects of natural selection on individual animals. The dominant animals gain first access to food, mates, or other important environmental features without needing to drive all other members of the species away; the selective benefits to the dominant are obvious. Subordinate animals reduce their risk of injury in combat by submitting, but by doing so they also give up priority of access to resources. However, subordinates are close to resources that may be taken if there is a surplus and, perhaps most important, if the alpha animal dies they may compete for its position.

Another motivation for group living is mutual defense. Even though subordinates receive less food or have fewer opportunities to mate, they may have greatly increased chances of escaping predation. This is particularly striking in herds of ungulates, such as antelopes, in which a few animals are always in a vigilant posture (head up, looking for predators) while the others feed. Thus social hierarchies make group living orderly, allocate resources, and encourage mutual defense; the peaceful coexistence of competing individuals is facilitated by such hierarchies. *See* BEHAVIORAL ECOLOGY; POPULATION ECOLOGY; SOCIAL MAMMALS; TERRITORIALITY. Michael D. Breed

Bibliography. J. Archer, *The Behavioural Biology of Aggression*, 1988; I. Eibl-Eibesfeldt, *Human Ethology*, 1989; J. L. Gould, *Ethology*, 1982; D. H. Morse, *Behavioral Mechanisms in Ecology*, 1990; E. O. Wilson, *Sociobiology: The New Synthesis*, 1990.

Social insects

Insects that share resources and reproduce cooperatively. The shared resources are shelter, defense, and food (collection or production). After a period of population growth, the insects reproduce in several ways.

As social insect groups grow, they evolve more differentiation between members but reintegrate into a

more closely organized system known as eusocial. These are the most advanced societies with individual polymorphism, and they contain insects of various ages, sizes, and shapes. All the eusocial insects are included in the orders Isoptera (termites) and Hymenoptera (wasps, bees, and ants). *See* HYMENOPTERA; INSECTA; ISOPTERA; POLYMORPHISM (GENETICS).

Polymorphism. The social insects have evolved in various patterns.

Termites. The most primitive termites live in small patches of dead wood. After incubation of the fertilized clutch of eggs, the larvae that hatch are fed on glandular food at first and later, after several molts, are given semidigested food from the anus of the adults. In this way they obtain important intestinal protozoa that help them digest wood. Generally, the young do not develop wings and gonads, and most become workers in the nest. A few may molt twice (**Fig. 1**) and form a wingless adult with a large head that defends the others; these are known as soldiers. For some time after, only workers are formed, as the reproductive pair inhibit the formation of their own kind, and so do the soldiers. This control is achieved because development inhibitors, specific for each of the castes inhibited, are secreted into the common food stream. The inhibitors are hormonal in nature and are sex- and soldier-specific.

The most advanced termites have evolved distinct lines of development from egg stage: the reproductive line can be identified from the existence of individuals with wing rudiments (nymphs), but their metamorphosis into sexual adults depends on the season as well as on whether the pair of reproductives withhold the development inhibitors. The asexual line consists of termites of both sexes in various sizes; they can molt into either workers or soldiers that are adult only in that they are developmentally fixed. Usually, soldiers are formed from one sex, which is specific for the species. The ratio of soldiers to workers is also a characteristic of the species; one family of termites no longer produces soldiers at all. Polymorphism in higher termites is thus a complex process that is responsive to both social and environmental conditions.

In the Hymenoptera, the society is composed of only females; males are produced periodically for their sperm. They usually congregate and attract females, or they visit colonies with virgin females and copulate there. In the Hymenoptera, sex is determined largely by whether the individual has one or two sets of chromosomes. Thus the queen has the power to determine the sex of her offspring: if she lets any of her stored sperm reach the egg, a female is produced; if not, a male results. *See* SEX DETERMINATION.

Bees and wasps. In the more primitive bees and wasps, social role (caste) is influenced by interaction with like but not necessarily related individuals. The female that can dominate the others assumes the role of queen, even if only temporarily. Domination is achieved by aggression, real or feigned, or merely by a ritual that is followed by some form of salutation by the subordinates. This inhibits the yolk-stimulating glands and prevents the subordinates from contributing to egg production; if it fails to work, the queen tries to destroy any eggs that are laid. Subordinate females take on more and more of the work of the group for as long as the queen is present and well. At first, all the eggs are fertilized and females develop, with the result that virgin females inhabit the nest for the first batches. They are often undernourished, and this, together with their infertility, reduces their urge to leave the nest and start another one. Such workers are said to be produced by maternal manipulation.

Though many more advanced societies of wasps are often perennial, the common wasps of temperate

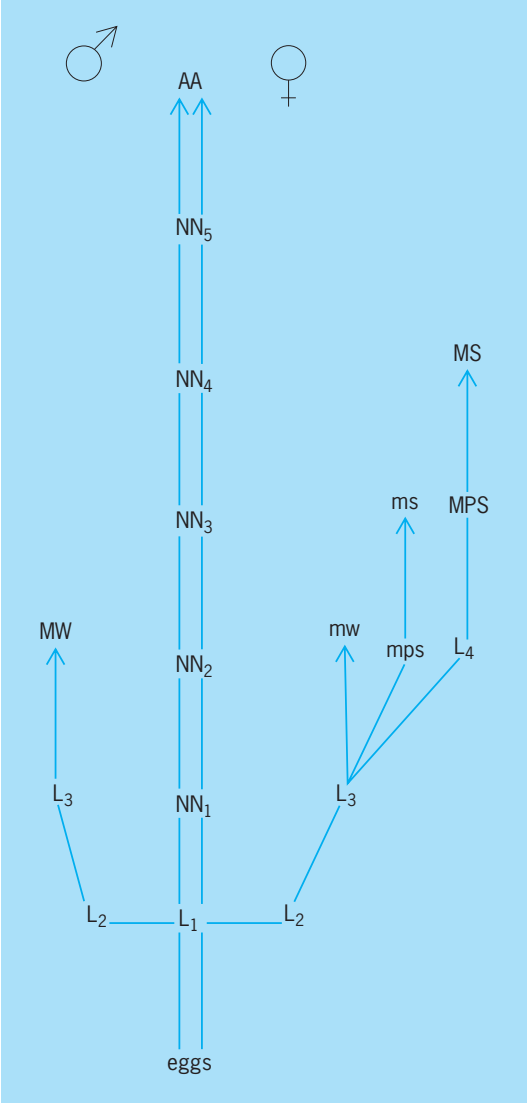


Fig. 1. Scheme of postembryonic development in mature colonies of *Macrotermes*, an advanced African termite that lives in savanna. Larvae (stages L_1–L_4) develop into large (MW) and small (mw) workers, and large (MPS) and small (mps) presoldiers, which become large (MS) and small (ms) soldiers, respectively; soldiers are female only. Sexual nymphs (N_1–N_5) and sexual adults (AA) develop only from larval stage L_1. (*After B. M. Okot-Kotber, Instars and polymorphism of castes Macrotermes michaelseni, Insectes Sociaux, 28(3):233–246, 1981*)

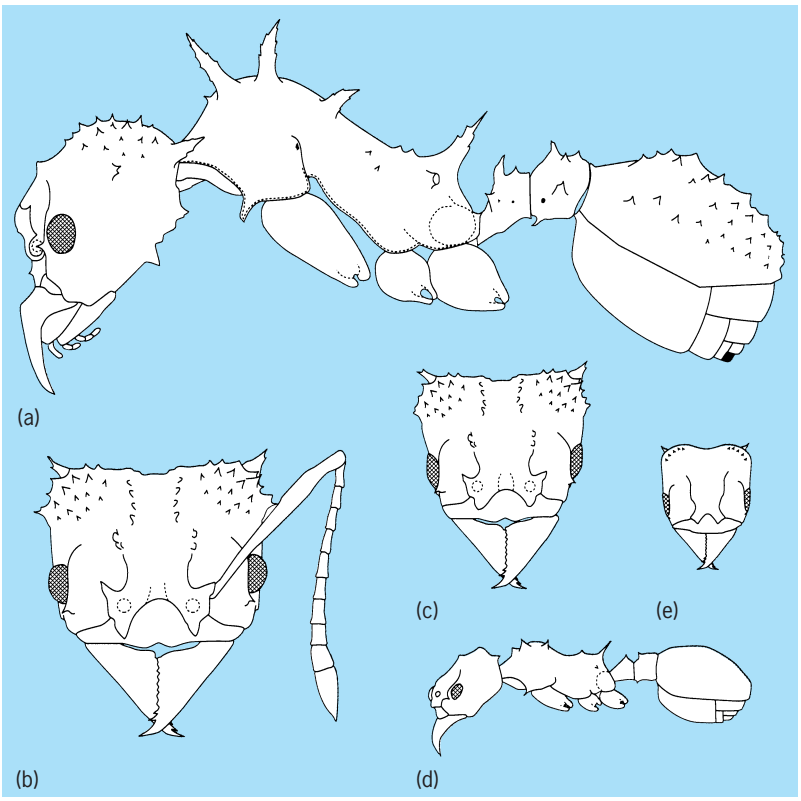


Fig. 2. Heads of the polymorphic *Acromyrmex octospinosus*, an ant that cultures fungus in Trinidad: (*a*) profile; and (*b*) front view of major worker; (*c*) front view of media worker; (*d*) profile; and (*e*) front view of minor worker. (*After W. M. Wheeler, Mosaics and Other Anomalies among Ants, Harvard University Press, 1937*)

regions have a single queen that is distinctly larger than her workers, having been reared in larger cells and given more and better food. If such a queen is removed, the society becomes disorganized, and the workers fight and lay their own unfertilized eggs, giving rise to males. Thus the current queen is not replaced, and the society dies out.

Stingless bees also have queens and workers of unequal size. The queens are fed more than the workers, and when these virgin queens mature they are held in their cells and fed until the current queen fails in some way, probably as an egg layer or a pheromone producer. When this time comes, a virgin is let out onto the comb and assessed in relation to the mother queen. With this type of system, there is little change in the society if the queen is removed. With the honeybee, however, removal of the queen sets up a series of reactions leading to her replacement by a daughter.

This replacement activity is a special case of the normal replacement that takes place in spring. It starts with the formation of queen cells on the lower rims of the central combs in which the mother queen lays fertile eggs. The larva that hatches from this type of egg is fed throughout its life on a jelly from the mouth glands of the workers (worker-cell larvae get jelly for only 3 days). Not only does this special diet yield an adult queen that is larger than the worker, and with a bigger ovary, but she also differs considerably in behavior and lives years rather than months. Queens leave the nest soon after emerging from their cell and search for males; they mate with 10 or more males. After 2 or more years they leave the nest in a swarm of thousands of workers, which amounts to about two-thirds of the total population, to establish a new colony. Workers generally stay in the nest to clean, make cells, and feed larvae until they are strong enough to forage for nectar and pollen. The queen normally stops the reproductive activities of workers by means of pheromones that affect worker behavior and are passed from one to the other by contact, not in the food stream. *See* PHEROMONE.

Ants. Reproductive ants, like termites, engage in a massive nuptial flight, after which the females, replete with sperm, go off to start a new nest. At some stage after the nuptials, the reproductives break off their wings, which have no further use. Workers, however, never have wings because they develop quickly and pass right through the wing-forming stages; their ovaries and genitalia are also reduced. Ant queens can prevent the formation of more queens; as with the honeybee, they do this behaviorally by using pheromones. They also force the workers to feed all larvae the same diet. To this trophogenic caste control is added a blastogenic control; eggs that are laid have a developmental bias toward one caste or another. This is not genetic; bias is affected by the age of the queen and the season: more worker-biased eggs are laid by young queens and by queens in spring. In some ants, workers mature in various sizes (**Fig. 2**). Since they have dis-

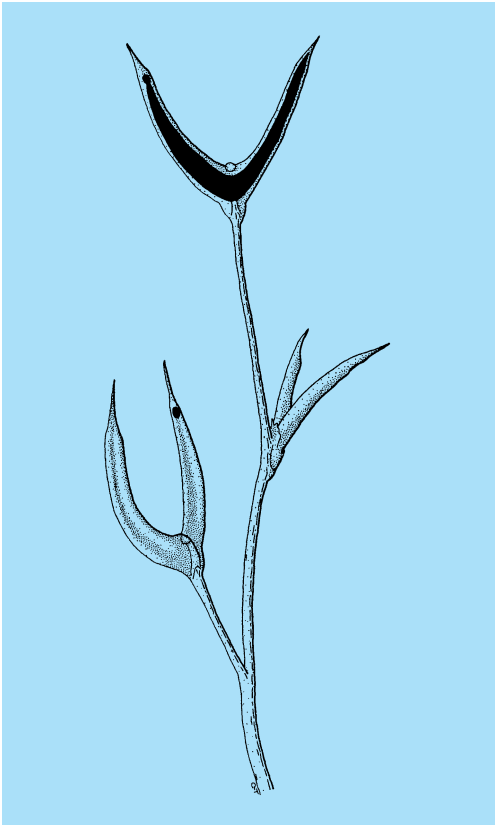


Fig. 3. Thorns of a species of *Acacia*, showing the cavity in which the ant *Pseudomyrmex* nests after entering by a hole near the tip. (*After A. J. Beattie, The Evolutionary Ecology of Ant-Plant Mutualisms, Cambridge University Press, 1985*)

proportionately large heads, the biggest workers are used mainly for defense; they also help with jobs that call for strength, like cutting vegetation or cracking nuts.

Nest construction. Social insects make remarkable nests that protect the brood as well as regulate the microclimate. The simplest nests are cavities dug in soil or soft wood, with walls smoothed and plastered with feces that set hard. Chambers at different levels in the soil are frequently connected by vertical shafts so that the inhabitants can choose the chamber with the best microclimate.

Termites and ants also make many different types of arboreal nests. These nests are usually made of fecal material, but one species of ant (*Oecophylla*) binds leaves together with silk produced in the salivary glands of their larvae that the workers hold in their jaws and spin across leaves. A whole group of ants (for example, *Pseudomyrmex*) inhabit the pith of plants (**Fig. 3**).

Termite mounds such as those of *Macrotermes* (**Fig. 4**) that form conspicuous features in savanna all start as a hole in the ground excavated by a male and a female. They grow to include various types of chambers for adults, juveniles, and the symbiotic fungus, as well as ventilation shafts and cellars that go down to water level. The large mounds of dry leaves made by the wood ant (*Formica*) in temperate forests function to shed water and collect solar radiation. Galleries may go deep into the subsoil and are used by the workers for hibernation.

Each species of wasp cuts a characteristic type of wood and chews it into a soft pulp that it thins out to the consistency of paper and applies gradually to form the nest. Combs of great strength and precise design are constructed in this way, arranged in tiers

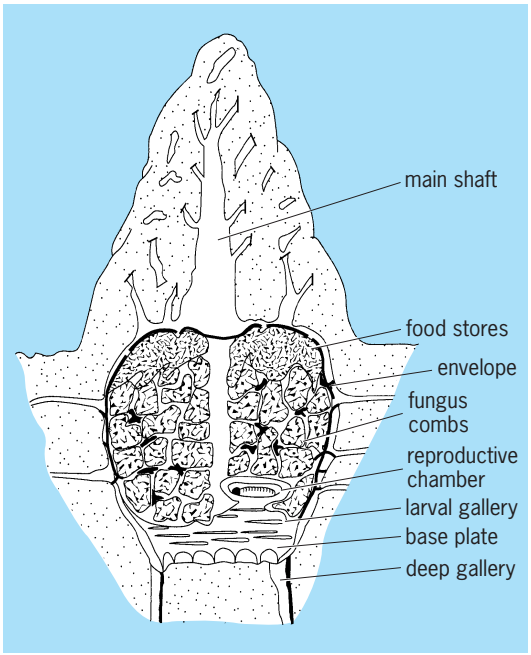


Fig. 4. Section of a *Macrotermes* nest in Nigeria about 18 ft (5.5 m) high and 9–12 ft (2.7–3.7 m) in basal diameter. (*After N. M. Collins, The nests of Macrotermes bellicosus, from Mockwa, Nigeria, Insectes Sociaux, 26(3):240–246, 1979*)

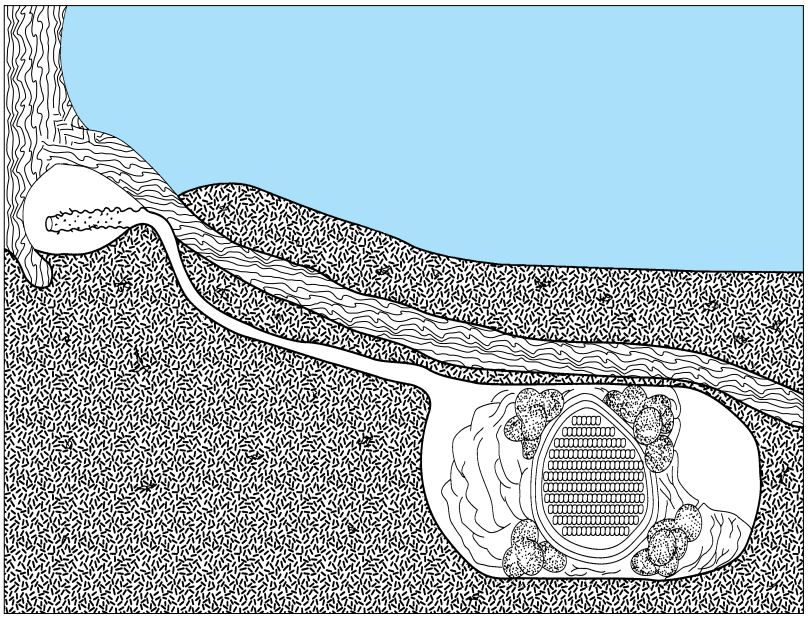


Fig. 5. Section of the nest of a stingless bee (*Trigona dorsalis*) in soil under a tree root. Twelve combs of brood cells are surrounded by a wax envelope, outside of which are many large storage cells supported by struts; the tube to the outside opens under the tree. (*After A. Wille and C. D. Michener, The nest architecture of stingless bees with special reference to those of Costa Rica, Revista de Biologia Tropical, 21(suplemento 1): 1–278, Universidad de Costa Rica, 1973*)

and surrounded by a paper envelope that is heat-refractory as well as impermeable to other insects. The cells are packed tightly in a hexagonal formation and may, in advanced wasps like *Vespula*, be large for reproductives and small for workers. Such a strong, light nest can be suspended from a branch by struts that are smeared with a chemical to repel ants.

Social bees use wax secreted by their cuticular glands and frequently blended with gums from tree exudates for their nest construction. Cells are made cooperatively by a curtain of young bees that scrape wax from their abdomen, chew it with saliva, and mold it into the correct shape; later it is planed and polished. Some primitive species, such as *Bombus*, make combs with spherical cells that are used first for brooding and later for storage of pollen and honey. Stingless bees (for example, *Trigona*) often stratify groups of cells and link cell layers with struts. Such a brood zone may be surrounded with wax envelopes, and these in turn with large globular storage cells (**Fig. 5**). The whole nest is usually in a cavity and is connected to the open air by a long tube. With honeybees the hexagonal comb reaches perfection as a set of back-to-back cells, each sloping slightly upward to prevent honey from running out. The same cells are used repeatedly for brood and for storage; or they may be made a size larger for rearing males. Only the queen cell is pendant, with a circular cross section and an opening below.

Symbiosis with other organisms. The ubiquity and ecological power of social insects depend as much on their ability to evolve mutualistic relations with other organisms as on the coherence of their social organization. Wood and the cellulose it contains is normally available as a source of energy only to

bacteria and fungi. However, it is used as a basic resource by both termites and ants that have evolved a technique of culturing these organisms. Though lower termites have unusual protozoa as intestinal symbionts, higher termites have bacteria in pouches in their hind gut. Many have a fungus that they culture in special chambers in their nests (Fig. 4). The termites feed on woody debris, leaves, and grass cuttings; the fungus digests these materials with the aid of termite feces and produces soft protein-rich bodies that the termites share with their juveniles and reproductives, neither of which are able to feed themselves. Protected from the weather, the fungus can remain active throughout the dry seasons—an inestimable advantage in the subtropics.

The ant–fungus symbiosis is similar but has evolved only in the central American region. The ants carry leaves to the nest to feed their fungus. Leaves from a great many different species are cut, and a distinct preference is shown. In the nest, the waxes are first rubbed off, and the oils, extracted by chewing, are used as a source of energy for the workers. The paste that results is treated with ant feces and enzymes from the fungus to degrade the material and convert it into an easily digested protein that is fed to the larvae and queen. Apart from producing a digestible, energy-rich material for the ant, the fungus extracts and destroys many toxic substances that are unpalatable to ants.

Many ants collect and store seeds that they mash and feed directly to their larvae. Provided they are collected when dry and stored in well-ventilated chambers, these seeds can remain viable and edible for an entire season. The plants benefit because not all the seeds are eaten; some that start to germinate are thrown out with the rubbish of the colony—in effect a way of planting them. Others are left behind by the ants when they change nests, as they often do when they suffer shading by encroaching vegetation. In this way, grass seeds can extend their range into dry areas that they could not reach alone. Some plants produce a hard shiny seed that the ants cannot grip and, to make it attractive, they add an aril, which is a growth that contains attractive foods. The ants take the seed back to their nest and eat the aril without harming the seed, which they later throw away or just leave buried in the soil of their nest. In woodland, most ants choose to live in sunny spots that suit the plant as well.

The dispersal of plant pollen by bees is a well-known symbiosis, and it has led to the evolution of many strange shapes, colors, and scents in flowers. Quite specific flower–bee relationships may exist in which one plant may use very few species of bees for the transfer of pollen. However, it is possible that this relationship may be breaking down, since many species of bee can find the nectar lure directly and cut through the plant tissue without pollinating the plant at all. Also, many bees collect pollen from flowers that do not have nectar; they want the pollen itself, which is a very good food. Highly successful species of stingless bees and honeybees exhibit behavior that attracts other bees of the same nest to newly discovered open flowers. *See* POLLINATION; POPULATION ECOLOGY. M. V. Brian

Bibliography. M. D. Breed (ed.), *The Biology of Social Insects*, 1982; M. V. Brian, *Social Insects: Ecology and Behavioral Biology*, 1983; W. Engels (ed.), *Social Insects*, 1990; B. Holldobler and E. O. Wilson, *The Ants*, 1990; C. D. Michener, *The Social Behaviour of the Bees*, 1990; T. D. Seeley, *Honeybee Ecology: A Study of Adaptation in Social Life*, 1985; E. O. Wilson, *The Insect Societies*, 1990.

Social mammals

Mammals that exhibit social behavior. This may be defined as any behavior stimulated by or acting upon another animal of the same species. In this broad sense, almost any animal which is capable of behavior is to some degree social. Even those animals which are completely sedentary, such as adult sponges and sea squirts, have a tendency to live in colonies and are social to that extent. Social reactions are occasionally given by species other than the animals' own; an example would be the relations between domestic animals and humans.

Mammals are warm-blooded vertebrates which typically live on the land surface of the globe. Several orders, however, have semiaquatic forms, such as muskrats and beavers among rodents, and otters and seals among carnivores. One entire order, the whales, has become completely water-living, and another, the bats, has developed the power of active flight and has taken to the air.

A common characteristic of the entire group is the mammary gland, used for the feeding of the newborn young. Because this is developed only in females, there is a general tendency for a differentiation of labor in which the females care for the young and the males specialize in fighting behavior. In the subclass Marsupialia, the nipples are enclosed in a brood pouch, in which the young spend most of their early life. The early ingestive behavior of mammals, or nursing, thus has a high degree of social significance.

In addition to their great variety of habitats, mammals eat a great variety of foods, and occupy many different ecological niches. There are almost as many mammals which are nocturnal as are active chiefly during the day.

Behavioral capacities. Depending on its habitat, a given species may use one or more of the sense organs more than the others. Some nocturnal rodents such as mice and rats use the tactile and auditory senses a great deal more than the eyes, whereas a diurnal tree-living animal such as a squirrel may chiefly use its eyes. All major sense organs are well developed in each species.

Mammals show a great variety of motor capacities. The limbs can be highly developed for digging as they are in moles and badgers, or used almost entirely for locomotion as they are in the hoofed mammals. An arboreal primate such as the chimpanzee has all four limbs developed for manipulation and prehension.

Social behavior and relationships in mammals

Relationship	Type of social behavior	Typical combinations formed
Care-dependency	Care giving Care soliciting Ingestive (especially nursing) Shelter seeking Eliminative (in some species)	Adult-young (especially female-young)
Sexual	Sexual Investigatory Eliminative (in some species)	Male-female
Dominance-subordination	Agonistic	Male-male (but also in all other combinations)
Leader-follower	Allelomimetic	Female-young (usually weaker in other combinations)
Mutual care	Care giving (grooming)	Adult-adult
Mutual defense	Agonistic	Male-male

Even the tail can be used for prehension by some of the monkeys. At the other extreme, the hindlimbs of whales are externally invisible and the forelimbs are used only as flippers.

Mammals have relatively large and complex brains, with a large development of the cerebral cortex. There is also a considerable capacity for behavioral adaptation through learning. Mammals exhibit relatively few instances of elaborate patterns of behavior which are completely fixed by heredity. In general, mammals are highly adaptable, both from the evolutionary viewpoint and in their capacity to adjust to daily changes in the environment (see **table**).

Social behavior of the American elk. This is one of the largest living species of the deer family, standing as high as a horse. The eyes, ears, and nose are all well developed. The males grow large antlers which are used chiefly in the rutting season and are shed in the early spring. Elk were originally plains-living animals, but are now restricted by hunting to the open areas in the Rocky Mountains. They are highly social animals and live in large herds.

Social life is closely related to the seasons. In the autumn the male herds break up. Each male attempts to round up a herd of females with whom he mates as each comes into estrus. If another male appears, the two begin a furious pushing contest, ending when one is exhausted and driven off. *See* ESTRUS.

As winter comes on, the males and females separate into different herds and migrate to lower altitudes. Hundreds of individuals may move together, producing a migration trail which looks almost as if it were a roadway.

In the spring there is a return migration to the higher altitudes. Each band separates from the large winter herd and goes back to the locality from which it came. Males go higher than females, which stay together and bear their calves. As is true with many members of the deer family, the young calves will "freeze," or stay quiet while the mothers leave to graze or browse. Usually at least one female stays in close proximity to the young animals. The calves at first obtain most of their nourishment by nursing from their mothers. When alarmed, they call to the mother. They grow rapidly during the summer and follow their mothers throughout the first year. The seasonal round of behavior begins again with the rutting season in the autumn.

Types of behavior. The most highly developed type of social behavior is allelomimetic, in which each animal does the same things as those nearby, all responding to each other. This results in the formation and continued existence of a herd. Also of importance in these herbivorous animals is ingestive behavior. Adults spend a great deal of time browsing and grazing. A more highly social sort of ingestive behavior is the nursing of the calves. Related to this is the caregiving or epimeletic behavior of the mothers and the care-soliciting behavior of the young. Agonistic behavior, consisting of fighting and escape behavior, is common in the males during the rutting season. As is true of most herd animals, fighting is individual and groups never combine against one animal. Sexual behavior occurs at the same time, but only for a brief period with each female. Investigative behavior is common, and the animals frequently raise their heads to look around while engaged in grazing or other activities. There is no special pattern of eliminative behavior and relatively little social shelter seeking.

Social relationships. Social behavior results in the formation of social relationships. This is partly based on biological differentiation into three types, males, females, and young, each with its characteristic behavior. There are six possible combinations between them, and each kind of social behavior may result in the formation of a social relationship within a particular combination.

The sexual relationship developed between males and females is a brief one, lasting only as long as the male is able to maintain his dominant position in relation to a herd of females. It has no importance at other times of the year, although it may play some part in holding the male group together.

Both male-female and female-female relationships can involve dominance and subordination. However, as is the case in many grazing animals, whose food is scattered in such a way as to prevent competition, dominance has little importance except in the breeding season. Even here a male may be able to hold a dominant position only for a portion of the season.

In female-young associations the care-dependency relationship is important and long-lasting. As is the case in other herd animals, females will allow only their own young to nurse, and a close association

is built up throughout the nursing period. This relationship lasts into later life and forms the foundation of leadership within the herds. The young born in the same year and within the same herd also form a lasting relationship with each other. This group is particularly important for the males when they separate from the females as adults. In the female herds a leader-follower relationship develops, first between mother and offspring and eventually between older and younger females. This basically allelomimetic relationship is one of the two most important social relationships developed in an elk society.

Unlike the situation among many birds and fishes, the migration trails are apparently learned from one generation to the next and have been altered within historical times. The animals become localized, that is, they become attached to the area in which they were brought up, but there is no defense of territory. Strange herds mingle with each other without fighting, and agonistic behavior is important only in the rutting season.

Other ungulates. The social behavior of other even-toed ungulates follows a pattern similar to that of the elk. The red deer of Scotland are quite similar to them, and even more distantly related animals such as goats, sheep, and bison have the same general characteristics. There is much emphasis on ingestive behavior, because enormous quantities of vegetation must be eaten. A typical grazing animal follows a regular cycle of grazing and stops to chew its cud every few hours. Among the more social species this is always done in concert with others, so that allelomimetic behavior is also highly important. Under natural conditions the result is the enormous herds of ungulates which were once seen on the plains of North America and on the plains of South Africa.

The predominant social relationship of the even-toed ungulates is the mother-young relationship. Its formation has been studied in detail in sheep. The mother sheep will allow only its own lamb to nurse and will drive all others away. The behavior of the mother toward its own lamb is fixed within the first 4 h after birth. The result of this close and definite relationship is that the leadership is developed by females, and social inheritance tends to be matrilineal.

Social behavior of wolves. These carnivorous animals are especially interesting because of their wide distribution and their close association with humans in fact and legend. The first domestic animal, the dog, was derived from them and has almost identical patterns of social behavior.

Under ordinary conditions the food supply of a carnivore is not regularly available in any one area. Consequently the chief activity of wolves is hunting, a form of investigatory behavior. Much of this is done at night. Having spent the day in or near the den, the wolf pack assembles toward evening and hunts far away, usually returning by morning. The hunting range may be 20–50 mi (30–80 km) across. In their hunting activities, they show a great deal of allelomimetic behavior, both in searching for and pulling down game, but there is little evidence of consistent leadership by one individual.

Mating and care of young. Mating behavior takes place in January or February. The entire estrous cycle of a female extends for 6 weeks or more, during which time there is long preliminary courtship as well as actual mating. A peculiarity of sexual behavior is the sexual tie, in which the male and female remain coupled for as long as $^1/_2$ h. It is not known whether the female mates with more than one male, but in related animals which do not run in packs, such as foxes and coyotes, a female typically mates with one male, and the two form a lasting association.

The cubs are born in early spring and are constantly attended by the mother during the first few days of life. As they grow older the mother must leave them to feed and hunt. There is usually a good supply of food at the den, either carried back or vomited by the adults. The cubs begin to eat freshly vomited food at about 3 weeks of age and are completely weaned at 6 or 7 weeks. However, they are still unable to hunt and remain dependent on the adults until approximately 6 months of age.

Thus the litter is in constant association from birth, but parents and other adults leave them for long periods each day. The closest social relationships are made with litter mates, and this is the foundation for the adult pack. Some young wolves are known to remain in the pack in which they were raised, but there is no clear evidence as to how new packs are formed. It is possible that in some cases a group of litter mates may leave an old pack to form one of their own. Most wolf packs consist of five or six adults plus any young animals running with the pack but not old enough to leave it.

Dominance. Agonistic behavior within the pack is organized into a definite dominance order. Because wolves occupy such large ranges, there is no possibility of guarding the boundaries as territories, although some packs regularly hunt over circular runways. Wolves set up scent posts which are regularly visited and at which the animals urinate and defecate. It is possible that in this way a wolf can tell whether a range is occupied and so avoid it. The den area is a definite territory; a strange wolf coming to the den is attacked and driven off.

Wolves are relatively long-lived animals, and social relationships, including the sexual one, are quite persistent. The allelomimetic relationship between litter mates is a most important one, and also important is the dominance-subordination relationship. The relationships between parents and offspring are relatively unimportant.

Social behavior of cat family. By contrast, members of the cat family tend to emphasize the mother-offspring relationship. This is associated with a different method of hunting: lying in wait for the prey and pursuing it for a short distance rather than ranging widely. In addition, cats are less omnivorous and more dependent on meat than members of the dog family. The young develop rapidly and are assisted in

their first hunting by the mother. With the exception of lions and cheetahs, cats tend to be solitary in their hunting activities and show very little allelomimetic behavior.

Social behavior of seals. Alaskan seals show highly organized social behavior upon their breeding grounds. The males take up territories along the shore in the late spring and defend them against males and other mammals. As females arrive, the males guard them as a harem. A few days after arrival the females produce their pups and shortly come into estrus, mating only once. Females leave their pups in groups while they swim off to feed, but males do not feed in this period. Thus two social relationships are prominent on the breeding grounds: the dominance relationship of the harem males, and the mother-offspring relationship. The sexual relationship has little social importance. Females do not breed until 4 or 5 years of age, and males are not able to maintain harems till about 10 years. Large groups of immature males stay outside the territories of the harem males. After the young are mature enough to leave the breeding grounds, the seal herds break up and become almost entirely aquatic until the new breeding season.

Social behavior of carnivores. In general, the social behavior of carnivores is related to their food habits. Investigatory behavior is always prominent. Possibly because of this, carnivores have a strong reputation for intelligence. Some are strongly allelomimetic, but many are almost solitary in their activities. The care of the young is greatly emphasized, because successful hunting depends upon size and experience. Agonistic behavior is also of great importance, as might be expected in a hunting animal. The parent-offspring relationship is important in early life because of the difficult manner of obtaining food. In some species, but not all, the sexual relationship is an important one, and there is a tendency toward the development of lasting bonds between males and females. *See* CARNIVORA.

Social behavior of rodents. Among mammals, rodents are good builders and diggers, and this is reflected in their social life. Perhaps the highest development of rodent society is found in the prairie dog, a large ground squirrel of the western plains in the United States. These animals dig numerous burrows and pile the dirt in mounds, which are the outward signs of a prairie dog town. A town is subdivided into territories, each inhabited by a coterie of prairie dogs. A coterie includes one or two male and two or three female adults, which produce numerous young and guard the territory of perhaps half an acre. Mating takes place in February and March, at which time there is a good deal of fighting between males and some reorganization of territories. Each female in the territory retires to a particular burrow and guards it as long as the young are belowground. Males must stay in other holes. When the young come out in April the females relax and all is again peaceful within the coterie.

The young learn the territorial boundaries by being challenged as they cross them. There is some social organization throughout the entire town, and any prairie dog that sees danger gives a warning signal which alerts all the rest. In June and July the adults leave the territory to construct new burrows on the outskirts of the colony. Thus there is a regular cultural inheritance. The young learn the territory from the parents and inherit the burrows. This system provides for colonization and also means that the experienced adult animals live on the outskirts of the colony, the part most exposed to danger.

In behavior typical of most other rodents, the principal type of social behavior is the care of the young in the nest, a relationship which tends to break up as the young leave. The male-female relationship is more important than in ungulates but is relatively unstable. The young are bound together chiefly by the territorial system. There is almost no allelomimetic behavior and consequently very little coordinated group activity. In the less social rodents such as mice, an entire society may consist of a mother and her young. *See* RODENTIA.

Social behavior of primates. Primates form a large and variable order consisting of nearly 200 species. With the exception of humans and a few species that have penetrated into mild temperate regions, primates live in tropical climates, perhaps because of their highly developed fingers, which become stiff and unusable in cold weather. However, they have evolved in widely different ways as they have become adapted to different habitats. The plains-living baboons with their highly developed teeth and habits of running on all fours are similar in many ways to dogs, while the tree-living galagos have a general appearance similar to that of domestic cats. Gorillas have become entirely herbivorous and in their large size and huge paunches resemble the larger ungulates. There is even a slothlike primate, the slow loris, which lives entirely in the tops of trees and moves deliberately, although not hanging upside down continually as do true sloths.

Unlike most other mammals, primates have a highly developed sense of color vision, and some species, such as mandrills, have evolved great differences in color and form between the sexes and between adults and young, similar to the plumage differences in certain birds. Again as in birds, hearing is well developed and olfaction is relatively unimportant.

Primates have the best-developed capacities for manipulation and prehension of any group of mammals. An opposable thumb is found in many species, and the hindlimbs are often well developed for manipulation. In some arboreal forms the tail is an important grasping organ. The characteristic mode of reproduction involves the production of a single infant. Unlike the young of the ungulates, those of the primates are usually born in a relatively immature state and require constant care and considerable maternal manipulation over long periods.

Baboons. These plains-living primates are in many aspects of their social life more similar to humans

than are the more closely related anthropoid apes. A savannah baboon society consists of several males of various ages, a larger group of females, and assorted young of both sexes. The whole group occupies an area of approximately 2 mi^2 (5 km^2), wandering over it in search of food in the daytime and returning to certain trees for sleeping at night. The males are organized into a definite dominance order. The most dominant male usually stays in the center of the group with other males spaced out around him. This results in the younger and presumably more active and alert males being on the edge of the group. When a baboon sights a predator, it gives an alarm call, and all the males quickly gather on the side of the troop toward the predator and threaten or attack it. The combined force is usually enough to scare off most predators except lions.

The females and their infants are likely to be found close to the dominant male, not because he is attractive, but because this is the safest place in the group. Females in estrus may mate with the dominant male or may prefer another. There is no evidence of permanent consortship, and females freely move from one male to another. There is no fighting between males over females, and almost all displays of agonistic behavior from whatever cause are confined to threats.

The dominance order is developed largely through playful activities of the young. When an infant is hurt by another baboon or when a young animal playfully attacks an older one, the aggressor is chased and threatened by one of the older males. In this way the young baboon soon learns his place in the hierarchy. There is no special relationship with the biological father and his offspring, but successive offspring of the same female retain a special relationship with her and each other. This is usually shown by contact and mutual grooming. Baboons thus develop a cohesive and cooperative society which maintains itself over long periods under favorable conditions.

Related primates such as rhesus monkeys and other macaques develop a similar social organization although there is considerable variation with respect to the amount of overt agonistic behavior and the amount of direct contact between individuals. Cultural transmission of information has been observed in the Japanese macaque. One troop was fed grain on a sandy shore. After one member learned to separate sand and grain by flotation, the habit spread through the group, the oldest members adopting it last.

Langurs. The Indian langur monkey is much more arboreal, but a troop will frequently descend to the ground to feed. They live chiefly on leaves. The adults have a striking appearance, with dark faces surrounded by white hair rising to a crest on top. As in most other primates, there are almost always more males than females in the usual troop. In addition, small all-male groups are common. Compared to macaques, langur groups are more relaxed and show less aggressive behavior. The adult males in a troop are organized into a definite dominance order and are always dominant over females. Dominance relationships among females are less definite. Males determine the direction of movement of the group. When an infant is born, all adult females are intensely interested and the infant is passed from hand to hand.

Group activities are integrated by whooping calls from males, which draw the group together in forested areas and also communicate the troop's location to nearby troops. Interactions of an aggressive nature are common between groups. Attacks are usually initiated by the low-ranking males, and adjacent troops show indications of a dominance order with respect to each other. As in many other primates, there is a core area in which the troop spends most of its time and a larger area over which the troop wanders. Ranges overlap, and there is no defense of territorial boundaries as such. Intertroop conflicts increase as the home ranges are compressed by population pressure.

Howler monkeys. These New World monkeys are entirely arboreal. They exhibit very little agonistic behavior of any kind, except that when troops draw near to each other the males roar loudly, and the two groups thus avoid closer contact. All feeding is done in the tops of trees. As a troop moves from place to place, males usually take the lead, although there is no single leader of the group. As with baboons, females in estrus move from male to male without conflict. The care-dependency relationship between mothers and offspring is the most prominent.

Anthropoids. Of the three major genera, orangutans are virtually unknown with respect to social organization in the wild, and much information remains to be gathered about the other two. Gorillas are the largest anthropoids. They live in small groups consisting of one or two males plus a few females and their offspring, and perhaps a few solitary males. The sex organs are inconspicuous, and sexual behavior is rarely observed. Little fighting takes place within the group. The animals respond to potentially dangerous predators by beating their chests, by threatening, and occasionally by attacking. Mature males are organized into a dominance order, but there is no hierarchy among females. Home ranges overlap, and when two troops meet they may exhibit a few threats, or mingle together and travel peacefully for a time. Both gorillas and chimpanzees live in dense forests and frequently build temporary nests in trees.

Social organization among chimpanzees is much less definite than in many other primates. The animals in a given area rarely form large groups but associate with each other in small, temporary bands consisting of females with young, all-male groups, or groups of males and females without young. There are few indications of dominance. Fighting is rare but there is some threatening. Much of the social behavior is composed of mutual grooming, as in other primates. Most food is fruit, but chimpanzees occasionally kill small animals and eat them. They also use crude tools in the form of twigs and rolled-up

leaves to extract termites from their nests, and these habits are apparently learned from mothers by infants, indicating the existence of some degree of culture.

Thus while chimpanzees are, as individuals, more similar to humans than any other primates, their unstable social organization is quite dissimilar. The size and composition of primate societies vary from species to species. In the gibbons, in which the two sexes are much alike in physique and both are highly aggressive, the typical group includes only one adult male and female. The forest-living great apes, such as chimpanzees and gorillas, live in small groups of eight or nine, mostly young, with one or two each of adult males and females. In contrast, the macaques may form groups as large as 150. Thus the ground-living primates form larger groups than those which live in trees. However, there is a great variety of social organization and behavior in primates, and no one species can be taken as an exact model for primitive human behavior. *See* APES.

Development of social organization. The postnatal development of each species is closely related to the social organization typical of the adults. Every highly social animal has a short period early in life when it readily forms attachments to any animal with which it has prolonged contact. This critical period for primary socialization can be demonstrated by removing young from their natural parents and rearing them by hand. Experience before this period has no effect; during the period the young animal transfers all its native social behavior to its human contact; thereafter this process becomes increasingly difficult, although captivity and hand feeding have some effect even on adults.

Process of socialization. The process of socialization begins almost immediately after birth in ungulates like the sheep, and the primary relationship is formed with the mother. In dogs and wolves the process does not begin until about 3 weeks of age, at a time when the mother is beginning to leave the pups. Consequently the strongest relationships are formed with litter mates, thus forming the foundation of a pack. Many rodents stay in the nest long after birth; primary relationships are therefore formed with nest mates. Young primates are typically surrounded by a group of their own kind, but because they are carried for long periods the first strong relationship tends to be with the mother.

Mammalian society. In summary, mammals may develop all types of social behavior to a high degree, but not necessarily in every species. The above examples include four highly developed societies from representative orders; in other species certain types of social behavior may be completely absent. Mammals have great capacities for learning and adaptation, which means that social relationships are often highly developed on the basis of learning and habit formation as well as on the basis of heredity and biological differences. The resulting societies tend to be malleable and variable within the same species and to show considerable evidence of cultural inheritance from one generation to the next. Mammalian societies have been completely described in relatively few forms, and new discoveries will probably reveal the existence of a greater variety of social organization.

Basic human social organization and behavior obviously differs from that of all other primates, although relatedness does exist. At the same time, the range of variability of human societies as seen in the nuclear family does not approach that in mammals as a whole. Human societies are characterized by the presence of all fundamental types of social behavior and social relationships rather than by extreme specialization. *See* ETHOLOGY; HUMAN ECOLOGY; REPRODUCTIVE BEHAVIOR; SOCIOBIOLOGY. John P. Scott

Bibliography. W. C. Allee, *The Social Life of Animals*, 1938, reprint 1976; S. A. Altmann (ed.), *Social Communication among Primates*, 1967, reprint 1982; C. A. Bramblett, *Patterns of Primate Behavior*, 2d ed., 1993; C. R. Carpenter, *Behavioral Regulators of Behavior in Primates*, 1974; W. E. Etkin, *Social Behavior and Organization among Vertebrates*, 1964; J. P. Scott, *The Evolution of Social Systems*, vol. 3 of *Monographs in Psychology*, 1989; N. Tinbergen, *Social Behaviour in Animals*, 2d ed., 1990.

Sociobiology

A scientific discipline that applies principles of evolutionary biology to the study of animal and human social behavior. It is a synthesis of ethology, ecology, and evolutionary theory in which social behavior is viewed as a product of natural selection and other biological processes. *See* ECOLOGY; ETHOLOGY; ORGANIC EVOLUTION.

Underlying principles. Sociobiology predicts that individuals will behave in ways that maximize their fitness (their success at projecting copies of their genetic material into succeeding generations) and argues that such behaviors can arise through the same evolutionary processes that operate on other trait systems. The central principle underlying sociobiology is that an individual's behavior is shaped, in part, by its genes, and thus is heritable and subject to natural selection. Natural selection is simply the result of the differential survival and reproduction of individuals who show heritable variation in a trait. The variants of a trait that allow individuals to survive better or reproduce better—the variants that convey greater fitness—will increase in frequency in a population over time. *See* BEHAVIORAL ECOLOGY.

As an example of natural selection, consider a population made up of asexually reproducing, haploid individuals carrying one of two alternative forms of a gene, allele *A* or allele *a*, initially at equal frequency. In each generation, individuals carrying one of these alleles have more offspring than individuals carrying the other. Over time, the frequency of the allele associated with greater fertility will increase in the

population, while that associated with lower fertility will decrease. Applied to behavioral traits, the logic is similar. Individuals carrying allele *A* are more likely to perform behavior *X*, whereas individuals carrying allele *a* are more likely to perform alternative behavior *x*. If the different behaviors, *X* and *x*, result in differential survival or reproduction, those individuals performing the more successful behavioral option will pass on more copies of the corresponding allele to the next generation. Eventually, the population will tend to be composed of individuals carrying the more successful allele and performing the more fit behavior. For diploid, sexually reproducing animals, this process is somewhat more complex (because of the reassortment of alleles at meiosis), but the outcome is fundamentally equivalent. *See* POPULATION GENETICS.

The distinctive sociobiological perspective on behavior views behaviors as strategies that have evolved through natural selection to maximize an individual's fitness. Sociobiologists employ this perspective in an inductive way, predicting that, all else being equal, animals should behave in ways that lead to greater fitness. Note that under this view there is absolutely no need to assume that maximizing fitness is in any way a conscious goal underlying an animal's behavior—successful behavior is simply an emergent result of the evolutionary process and not the force driving that process.

Some critics of the sociobiological perspective argue strongly against the notion of any genetic determinism for behavior, instead viewing environment and experience as the most important causal variables shaping behavior. This resurrects the ancient debate over nature versus nurture. Environment and experience are undoubtedly important, and the stance of most sociobiologists is that behavior—like all other aspects of an organism's phenotype—is the result of an interaction between genes and environment. Neither nature (genes) nor nurture (environment and experience) alone determines behavior. Rather, all behavior is influenced by both factors to varying degrees. It is the interaction between genotypes and environments that determines the direction and degree to which behavioral variants are associated with differential reproduction and survival—thus, the nature versus nurture debate is largely misguided. *See* BEHAVIOR GENETICS.

Altruism. Animals sometimes behave in ways that seem to reduce their own personal fitness while increasing that of other individuals. For example, in many species of social mammals and in numerous species of birds, individuals commonly give alarm vocalizations that alert others to the presence of a predator while seemingly rendering themselves more conspicuous to attack. In many species of social insects, large numbers of nonreproductive workers often labor on behalf of a very small number of reproductive individuals. Evolutionary biologists since Charles Darwin have noted that these seemingly altruistic acts pose a challenge to the notion of an animal's behavior being a strategy to maximize individual fitness. *See* SOCIAL INSECTS; SOCIAL MAMMALS.

Inclusive fitness. One resolution to this paradox comes from taking a "gene's eye view" of evolution, which recognizes that Darwinian fitness (an individual's success in producing offspring) is simply a specific case of the more general concept of inclusive fitness. Because an individual shares genes with its relatives, there are actually two different routes by which it can pass on copies of those genes to the next generation: first, through personal or direct reproduction; and second, by helping relatives to reproduce. Inclusive fitness is a composite measure of an individual's genetic contribution to the next generation that considers both of these routes.

Kin selection. Animals who behave altruistically toward relatives may, in fact, be behaving selfishly in the genetic sense if their behavior sufficiently enhances their inclusive fitness through its effects on the survival and reproduction of relatives. This phenomenon is referred to as kin selection. Because close relatives share more genes, on average, with an individual than do distant relatives, such seemingly altruistic behavior tends to be directed toward closer kin.

Numerous studies have demonstrated the importance of kin selection on structuring individual social behavior in a wide array of animal taxa. For example, female ground squirrels are especially likely to give alarm calls when their relatives are nearby, but they call far less often when close relatives are absent. Similarly, in cooperatively breeding birds and callitrichid primates (marmosets and tamarins), individuals that assist a breeding pair in rearing young are typically older offspring of the pair and thus are helping to rear their own younger siblings. Kin selection has also been implicated in the evolution of complex caste societies in some species of social insects. Because of their peculiar genetic system, workers in many species of social wasps and bees are more closely related to their sisters (other workers) than they would be to their own offspring if they were to reproduce. By helping the queen (their mother) to reproduce, they actually do more for their own inclusive fitness than if they were to attempt to produce offspring on their own. *See* SOCIAL HIERARCHY.

Reciprocity. Another solution to the altruism paradox—one that can even operate among nonrelatives under certain conditions—is reciprocity. If the recipient of some altruistic behavior is likely to repay a donor in the future, it may be beneficial for the donor to perform the behavior even at some immediate cost to its own fitness. This is especially true if the cost is low and the anticipated benefit is high. Examples include food sharing among unrelated vampire bats and alliance behavior in some primates.

Under some conditions, it is possible for altruism to result from natural selection operating at the level of groups, rather than at the level of individuals or genes (for example, if groups containing altruists were at a sufficient selective advantage compared with groups containing only selfish individuals). However, within such groups, selection acting

on individuals would still tend to favor selfishness over altruism. In any event, much of the cogency of sociobiology has resulted from the recognition that selection seems to be most powerful when it acts at lower levels—notably, genes and individuals rather than groups or species.

Other social strategies. Over the past three decades, the sociobiological view of behavior as an evolved strategy for maximizing inclusive fitness has proven to be a powerful explanatory paradigm for investigating many other aspects of animal social behavior. The phenomenon of sociality, for example, has been extensively explored from this perspective. Living in a social group is a behavior that involves numerous potential benefits (including improved access to food, more effective deterrence of or defense against predators, and enhanced opportunities for nepotism) as well as numerous potential costs (such as having to share food resources and competition over mates). Each species—and individuals within each species—can be investigated as to the relative costs and benefits associated with the decision of whether or not to be social.

Similar considerations apply to many other strategic (though typically unconscious) decisions that animals make during their lives, such as whether to reproduce sexually, when in life to begin reproducing (with regard to season and age at maturation), how many partners to mate with and who those partners should be, whether to compete with other individuals over access to partners, and whether to bestow parental care and the level of care to give. For example, consider the case of mating and parenting strategies for males versus females. Biologists define males as the sex that produces a large number of small, relatively mobile gametes (sperm), and females as the sex that produces fewer, larger, and relatively immobile gametes (eggs). In many species, this initial inequality between the sexes in parental investment is compounded by additional forms of care provided primarily by females (for example, gestation and lactation in mammals). Fundamental differences between males and females in the level of investment that each provides result in marked differences between the sexes in maximum potential lifetime fitness. In general, individual males of most species are potentially able to produce many more offspring than even the most fertile females since the male contribution to an offspring can be as little as a single sperm. Male reproductive success is thus limited, in many cases, by access to fertile females and is often enhanced by achieving a maximum number of successful copulations. For a female, fitness is much more limited by access to resources than by access to mates because of her much lower lifetime reproductive potential and relatively greater investment in each offspring. In general, the sex providing less parental investment (typically males) is expected to compete among themselves for access to the sex providing greater investment (typically females). Males, accordingly, are likely to maximize their fitness by being aggressive and competitive with one another and are expected to be less discriminating about mate choices, whereas females are likely to be more selective in their mate choices. Significantly, species in which the pattern of parental investment is reversed (as among certain species of birds, fish, and frogs in which males provide substantially more prenatal or postnatal care than females) also show a reversal in the pattern of competitive behavior, with females tending to be larger, showier, and more aggressive than males. *See* MATERNAL BEHAVIOR; REPRODUCTIVE BEHAVIOR; SEXUAL DIMORPHISM.

The sociobiological concept of behaviors as fitness-enhancing evolutionary strategies has led to many other insights concerning the expression of animal social behaviors, including predictions regarding parent-offspring conflict, sibling rivalry, birth order effects, development of dominance hierarchies, alliance and coalition formation, intentional deception, status striving, and cooperative hunting. Sometimes, distinct behavioral strategies for dealing with the same set of environmental challenges may exist in a population simultaneously—for example, some individuals might be likely to escalate fights while others back away. Sociobiologists sometimes interpret this as an example of an evolutionarily stable strategy set in which the frequency of one behavioral type may depend on the frequency of others maintained in the population.

Human behavior. Although most of the research in sociobiology has focused on understanding the behavior of nonhumans, sociobiological explanations have been used to interpret patterns of human behavior as well. This has generated much controversy within both the scientific and lay communities. Many people have objected to what they see as the political and ethical implications of applying concepts from evolutionary biology to the human species. Much of this concern is well founded, given the past abuses of Social Darwinism and various other racist and eugenic programs. However, an important distinction must be made between trying to understand the biological underpinnings of human behavior and morally condoning or condemning that behavior. Sociobiology may well shed considerable light on why humans do the things they do—for example, why men generally tend to be more aggressive and competitive than women, or why people tend to favor their relatives over nonkin—but it does not render judgment. In fact, a sociobiological perspective on human behavior emphasizes the existence of cross-cultural universals—features of behavior that all human beings share by virtue of their biology that transcend notions of race, class, and ethnicity. As such, sociobiological studies could constitute a potent antidote to racist ideologies.

David P. Barash; Anthony Di Fiore

Bibliography. D. P. Barash, *Sociobiology and Behavior*, 2d ed., Elsevier, 1982; L. L. Betzig (ed.), *Human Nature: A Critical Reader*, Oxford University Press, 1996; L. Cronk, N. Chagnon, and W. Irons (eds.), *Adaptation and Human Behavior: An Anthropological Perspective*, Aldine De Gruyter, 2000; R. Dawkins, *The Selfish Gene*, Oxford University Press, 1990; J. R. Krebs and N. B. Davies (eds.),

Behavioural Ecology: An Evolutionary Approach, 4th ed., Blackwell Science, 1997; V. Standen and R. A. Foley (eds.), *Comparative Socioecology: The Behavioural Ecology of Humans and Other Mammals*, Blackwell Science, 1989; E. O. Wilson, *Sociobiology: The New Synthesis*, 25th Anniversary Edition, Harvard University Press, 2000.

Soda niter

A nitrate mineral having chemical composition $NaNO_3$ (sodium nitrate). Soda niter, otherwise known as nitratite, is by far the most abundant of the nitrate minerals. It crystallizes in the rhombohedral division of the hexagonal system. It sometimes occurs as simple rhombohedral crystals but is usually massive granular. The mineral has a perfect rhombohedral cleavage, conchoidal fracture, and is rather sectile. Its hardness is 1.5 to 2 on Mohs scale, and its specific gravity is 2.266. It has a vitreous luster and is transparent. It is colorless to white, but when tinted by impurities, it is reddish brown, gray, or lemon yellow.

Soda niter is a water-soluble salt found principally as a surface efflorescence in arid regions, or in sheltered places in wetter climates. It is usually associated with niter, nitrocalcite, gypsum, epsomite, mirabilite, and halite.

The only large-scale commercial deposits of soda niter in the world occur in a belt roughly 450 mi (725 km) long and 10–50 mi (16–80 km) wide along the eastern slope of the coast ranges in the Atacama, Tarapaca, and Antofagasta deserts of northern Chile. The deposits consist of a thin bed of nitrates and associated minerals, varying from a few inches to a few feet in thickness, overlain by a shallow overburden of sand and gravel. The crude soda niter, known as caliche, is about one-fourth sodium nitrate, admixed with other salts, notably bloedite, anhydrite, gypsum, polyhalite, halite, glauberite, and darapskite, together with minor amounts of various iodates, chromates, and borates. Because of the presence of the iodate minerals lautarite and dietzeite, these deposits yield most of the world's supply of iodine.

The origin of these deposits is controversial. It is generally agreed that the nitrates were transported by ground water and deposited by evaporation. The source of the nitrate has been attributed to (1) guano; (2) nitrogen fixation by electrical storms; (3) the bacterial fixation of nitrogen from vegetable matter; and (4) a volcanic source in nearby Triassic and Cretaceous rocks. The last source seems the most probable. *See* NITROGEN CYCLE.

Chilean nitrate had a monopoly of the world's fertilizer market for many years, but now occupies a subordinate position owing to the development of synthetic processes for nitrogen fixation which permit the production of nitrogen from the air. This led to the commercial production of artificial nitrates and reduced the need to import Chilean nitrates for use in the manufacture of fertilizers and explosives. *See* CALICHE; EXPLOSIVE; FERTILIZER.

Small deposits of soda niter similar to those in Chile are found in Bolivia, Peru, North Africa, Egypt, Russia, India, and the western United States. *See* NITER; NITRATE MINERALS.

George Switzer

Sodalite

A mineral of the feldspathoid group, crystallizing in the isometric system, with chemical composition $Na_4Al_3Si_3O_{12}Cl$. Its tektosilicate framework, in which half the tetrahedrally coordinated Si atoms are replaced by Al, is arranged in four- and six-membered rings. The rings define sets of channels which intersect to form cavities. Na atoms reside in the channels and Cl in the cavities. Crystals are rare, usually dodecahedrons. Sodalite is usually massive or granular with poor cleavage. The Mohs hardness is 5.5–6.0, and the density 2.3. The luster is vitreous, and the color is usually blue but may also be white, gray, or green.

Sodalite is common in nephelite syenites and associated rock types. It also occurs in metasomatized carbonates at contacts with alkaline igneous rocks. Notable occurrences are at Mount Vesuvius, Italy; at Bancroft, Ontario, Canada; and on the Kola Peninsula, Russia. In carbonaceous chondrite meteorites, it occurs as a gas-phase alteration product of refractory condensate inclusions. *See* FELDSPATHOID; SILICATE MINERALS.

Lawrence Grossman

Sodium

A chemical element, Na, atomic number 11, and atomic weight 22.9898. Sodium is between lithium and potassium in the periodic table. The element is a soft, reactive, low-melting metal with a specific gravity of 0.97 at 20°C (68°F). Sodium is commercially the most important alkali metal. The physical properties of metallic sodium are summarized in the **table** below. *See* PERIODIC TABLE.

Sodium ranks sixth in abundance among all the elements in the Earth's crust, which contains 2.83% sodium in combined form. Only oxygen, silicon,

1																	18
1 H	2											13	14	15	16	17	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	3	4	5	6	7	8	9	10	11	12	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112	113					

lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb

actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

Physical properties of sodium metal

Property	Temperature °C	Temperature °F	Metric (scientific) units	British (engineering) units
Density	0	32	0.972 g/cm^3	60.8 lb/ft^3
	100	212	0.928 g/cm^3	58.0 lb/ft^3
	800	1472	0.757 g/cm^3	47.3 lb/ft^3
Melting point	97.5	207.5		
Boiling point	883	1621		
Heat of fusion	97.5	207.5	27.2 cal/g	48.96 Btu/lb
Heat of vaporization	883	1621	1005 cal/g	1809 Btu/lb
Viscosity	250	482	3.81 millipoises	4.3 kinetic units
	400	752	2.69 millipoises	3.1 kinetic units
Vapor pressure	440	824	1 mm	0.019 lb/in.^{2}
	815	1499	400 mm	7.75 lb/in.^{2}
Thermal conductivity	21.2	70.2	0.317 cal/(s)(cm)(°C)	76 Btu/(h)(ft)(°F)
	200	392	0.193 cal/(s)(cm)(°C)	46.7 Btu(h)(ft)(°F)
Heat capacity	20	68	0.30 cal/(g)(°C)	0.30 Btu/(lb)(°F)
	200	392	0.32 cal/(g)(°C)	0.32 Btu/(lb)(°F)
Electrical resistivity	100	212	965 microhm-cm	
Surface tension	100	212	206.4 dynes/cm	
	250	482	199.5 dynes/cm	

aluminum, iron, and calcium are more abundant. Sodium is, after chlorine, the second most abundant element in solution in seawater. The important sodium salts found in nature include sodium chloride (rock salt), sodium carbonate (soda and trona), sodium borate (borax), sodium nitrate (Chile saltpeter), and sodium sulfate. Sodium salts are found in seawater, salt lakes, alkaline lakes, and mineral springs. *See* ALKALI METALS.

Sodium reacts rapidly with water, and even with snow and ice, to give sodium hydroxide and hydrogen. The reaction liberates sufficient heat to melt the sodium and ignite the hydrogen. When exposed to air, freshly cut sodium metal loses its silvery appearance and becomes dull gray because of the formation of a coating of sodium oxide.

Sodium does not react with nitrogen. Sodium and hydrogen react above about 200°C (390°F) to form sodium hydride. Sodium reacts with ammonia, forming sodium amide. Sodium also reacts with ammonia in the presence of coke to form sodium cyanide.

Sodium does not react with paraffin hydrocarbons but does form addition compounds with naphthalene and other polycyclic aromatic compounds and with arylated alkenes. The reaction of sodium with alcohols is similar to, but less rapid than, the reaction of sodium with water. Sodium reacts with organic halides in two general ways. One of these involves condensation of two organic, halogen-bearing compounds by removal of the halogen, allowing the two organic radicals to join directly. The second type of reaction involves replacement of the halogen by sodium, giving an organosodium compound. *See* ORGANOMETALLIC COMPOUND.

Sodium chloride, or common salt, NaCl, is not only the form in which sodium is found in nature but (in purified form) is the most important sodium compound in commerce as well. Sodium hydroxide, NaOH, is also commonly known as caustic soda. Sodium carbonate, Na_2CO_3, is best known under the name soda ash.

A major use of sodium is in the reduction of animal and vegetable oils to long-chain fatty alcohols; these alcohols are raw materials for detergent manufacture. Sodium is used to reduce titanium and zirconium halides to their respective metals. Sodium chloride is used in curing fish, meat packing, curing hides, making freezing mixtures, and food preparation (including canning and preserving). Sodium hydroxide is used in the manufacture of chemicals, cellulose film, rayon soap pulp, and paper. Sodium carbonate is used in the glass industry and in the manufacture of soap, detergents, various cleansers, paper and textiles, nonferrous metals, and petroleum products. Sodium sulfate (salt cake) is used in the pulp industry and in the manufacture of flat glass. *See* GLASS; HALITE; PAPER. Marshall Settig

The sodium ion (Na^+) is the main positive ion present in extracellular fluids and is essential for maintenance of the osmotic pressure and of the water and electrolyte balances of body fluids. Hydrolysis of adenosine triphosphate (ATP) is mediated by the membrane-bound enzyme Na^+, K^+-ATPase (this enzyme is also called sodium pump). The potential difference associated with the transmembrane sodium and potassium ion gradients is important for nerve transmission and muscle contraction. Sodium ion gradients are also responsible for various sodium ion–dependent transport processes, including sodium-proton exchange in the heart, sugar transport in the intestine, and sodium-lithium exchange and amino acid transport in red blood cells. *See* POTASSIUM. Duarte Mota de Freitas

Bibliography. F. A. Cotton et al., *Advanced Inorganic Chemistry*, 6th ed., Wiley-Interscience, 1999; D. R. Lide, *CRC Handbook Chemistry and Physics*, 85th ed., CRC Press, 2004; C. A. Pasternak (ed.), *Monovalent Cations in Biological Systems*, 1990.

Sodium-vapor lamp

An arc discharge lamp with sodium vapor as the emitting species. Low-pressure and high-pressure types are in common commercial usage.

Low-pressure lamps. The low-pressure sodium lamp has remarkably high luminous efficiency, or efficacy, producing as much as 200 lumens per watt of input power. The radiation is nearly monochromatic yellow in color, and is used where color rendition is unimportant such as on piers, in harbors, for low-energy-cost security lighting, on bridges, and on roadways. The very high luminous efficacy is due to very stringent control of heat losses from the arc. The key to the heat conservation is the indium oxide film (**Fig. 1**) used to reflect infrared energy back to the arc tube and the evacuated outer jacket that minimizes the conducted thermal losses. Lamps at 18 and 10 W of power have been developed for very low-energy-cost uses. The 10-W lamp produces light at 100 lm/W. *See* ILLUMINATION; LUMINOUS EFFICACY.

The low-pressure sodium lamp has an operating sodium pressure of about 10^{-5} atm (1 pascal), and 10^{-3} to 10^{-2} atm (100 to 1000 pascals) of inert gas (typically neon) for arc starting and operating purposes. The arc tube wall temperature is about 520°F (270°C).

The chemistry problems of lamps containing elemental sodium are formidable. Ordinary glasses react at temperatures near 480°F (250°C), and low-pressure sodium lamps are specially coated with a sodium-resistant glass for protection to 520°F (270°C).

High-pressure lamps. When the sodium pressure is increased by a higher liquid-sodium-pool temperature, the yellow sodium D-line resonance radiation reaches a condition where pressure broadening and self-absorption occur, producing a lamp no longer monochromatic yellow but golden-white in color. The high-pressure sodium lamp has a maximum arc tube temperature typically above 2000°F (1100°C) with a sodium amalgam reservoir temperature about 1300°F (700°C). The term high-pressure is used merely to distinguish this light source from the low-pressure sodium lamp. The plasma arc column of the high-pressure sodium lamp has a total pressure of sodium, mercury, and inert gas of typically slightly less than 1 atm (10^5 Pa).

The severity of reactions involving elemental sodium increases with the higher operating temperature and pressure of the high-pressure sodium lamp, and special materials must be used for the arc containment vessel and for the electrical feedthrough seals. Indeed, the success of the high-pressure sodium lamp is due to the invention of an arc containment material resistant to sodium vapor at high temperature. This material is a very high-purity aluminum oxide ceramic (**Fig. 2**) with the ability to pass light with little attenuation. (Previously all discharge lamps used either glass or fused quartz as the containment material.) In order to conserve losses, the high-pressure sodium lamp uses low-thermal-

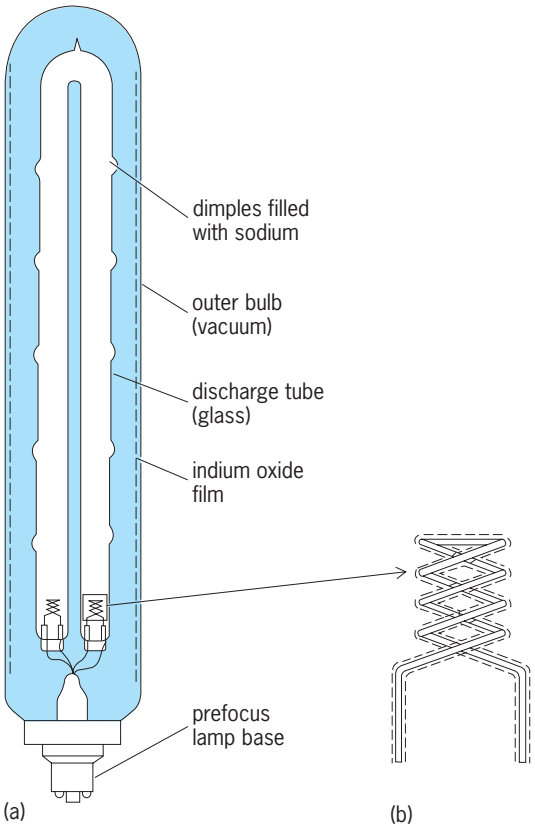


Fig. 1. Low-pressure sodium lamp. (*a*) Lamp structure. (*b*) Structure of electrode consisting of coiled tungsten wire coated with oxides of barium, strontium, and calcium.

conductivity buffers of mercury and xenon in the sodium discharge. The mercury and xenon make almost no contribution to the lamp spectra.

Properties and uses. In comparison with other arc discharge lamps such as high-pressure mercury, metal halide, and fluorescent, the sodium-vapor lamps produce very low levels of ultraviolet radiation: about 0.5% of the input lamp power results in ultraviolet with the major amount at 360 nanometers. The ultraviolet radiation in the erythema-sensitive regions near 300 nm is virtually nonexistent. Thus tanning of the skin is not possible with sodium-vapor lamps. *See* FLUORESCENT LAMP; MERCURY-VAPOR LAMP; METAL HALIDE LAMP; ULTRAVIOLET LAMP; ULTRAVIOLET RADIATION.

Low-pressure sodium lamps are used commercially in lamp powers from 10 to 180 W at efficacies from 100 to 200 lm/W. High-pressure sodium lamps are offered in powers ranging from 35 to 1000 W and from 60 to 150 lm/W. The maximum efficacies of sodium lamps are considerably greater than those of other light sources such as fluorescent lamps (approximately 90 lm/W), high-pressure mercury lamps (60 lm/W), tungsten halogen lamps (25 lm/W), and incandescent lamps (15 lm/W). Both types of sodium lamps convert 30 to 35% of input power to useful visible light.

Because of the more compact source size, the light from a high-pressure sodium lamp can be

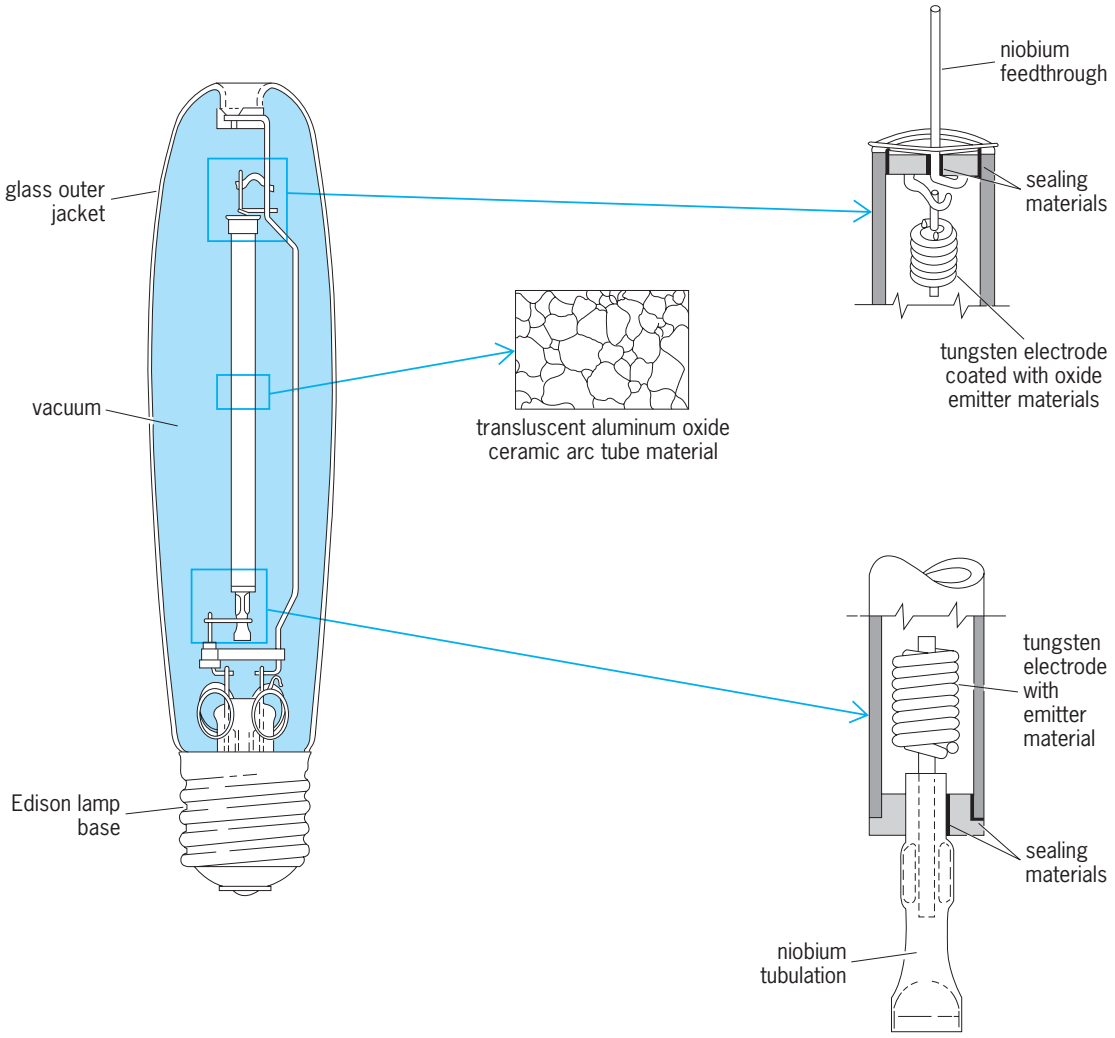


Fig. 2. High-pressure sodium lamp and components.

more effectively directed by reflectors and refractors than that from the low-pressure sodium lamp. Therefore the higher efficacy of the low-pressure sodium lamp is offset by its reduced luminous efficiency. Sodium-vapor lamps are replacing high-pressure mercury lamps in many floodlight and roadway applications. More sodium-vapor lamps are manufactured annually than the high-pressure mercury and metal halide lamps combined. *See* LAMP; VAPOR LAMP.

Charles I. McVey

Bibliography. M. A. Cayless and A. M. Marsden, *Lamps and Lighting*, 3d ed., 1983; J. J. DeGroot and A. J. Van Vliet (eds.), *The High Pressure Sodium Lamp*, 1986; Special issue on light source technology, *IEE Proceedings*, vol. 127A, no. 3, 1980.

Sofar

The acronym for sound fixing and ranging. The concept is based on finding the geographic location of sound that is created by an explosive or other impulsive event, such as a missile impact, by using the time of arrival of the signal at three or more receiving stations. This is known as hyperbolic fixing. This concept was used to locate aircraft that crashed at sea. Although other techniques have replaced this application, the term sofar channel has remained in use, referring to what is also known as the deep sound channel. *See* HYPERBOLIC NAVIGATION SYSTEM.

Sofar channel. Sound signals that are created in the deep oceans are known to travel very long distances. This is due to two important properties of sound in seawater. First, the speed of sound, which is approximately 1500 meters per second (5000 feet per second), is dependent on temperature and pressure, increasing with both as they increase. Second, the absorption of sound is very small, especially at lower frequencies. For most deep seas the water temperature is highest at or near the surface and decreases with depth down to a few hundred meters, where it reaches a nearly constant value, which in most oceans is about 4°C (39°F). Pressure, of course, increases with depth. Sound speed will, therefore, decrease with depth due to temperature, and increase with depth due to pressure. Typically this leads to a minimum value of sound speed at a depth of around

1000 m (3300 ft). This depth is usually referred to as the channel axis.

The sound radiating from an explosive source, for example, can be depicted as a set of rays that are emitted in all directions. Each ray can be traced using a simple rule that is based on Snell's law. From Snell's law it can be shown that a ray will always bend in a direction away from a region of higher sound speed. Thus there will be a set of rays that starts upward, and then bends toward the axis without reaching the sea surface. They will pass through the channel axis and then, since the sound speed increases with depth below the axis, they will then bend upward toward the axis where they will then repeat the same excursion, and so on. There will be another set that starts downward and then bends upward toward the channel axis. Thus part of the sound is trapped in the channel and, due to the very low loss due to absorption, intensity is essentially decreased only due to the cylindrical spreading of the rays. Without the channel sound would be spread spherically and thereby lose much more energy. *See* REFRACTION OF WAVES; UNDERWATER SOUND.

Shape of received signal. Another important feature of the sofar channel is its effect on the shape of the received signal. As an example, consider a signal that originates on the channel axis and is received at a range that is also on the axis. Each ray, having its own vertical angle of emission, will follow its own serpentine path. Using Snell's law, each path can be traced and, knowing the sound speed as a function of depth, the time of arrival can be determined. At each point along the channel axis there will be a unique set of rays that arrive. However, it will be found that they do not have the same arrival time. Rays that make the widest paths from the axis will arrive the earliest, with the ray that stays on the axis arriving last. The signal will then be the sum of a family of individual signals. It is the arrival time of the last of these that most accurately determines range.

Ralph R. Goodman

Software

A set of instructions that cause a computer to perform one or more tasks. The set of instructions is often called a program or, if the set is particularly large and complex, a system. Computers cannot do any useful work without instructions from software; thus a combination of software and hardware (the computer) is necessary to do any computerized work. A program must tell the computer each of a set of minuscule tasks to perform, in a framework of logic, such that the computer knows exactly what to do and when to do it. *See* COMPUTER PROGRAMMING.

Programming languages. Programs are written in programming languages especially designed to facilitate the creation of software. In the 1950s, programming languages were numerical languages easily understood by computer hardware; often, programmers said they were writing such programs in machine language. The computers understood certain number sequences to mean that a particular computer operation was to be performed, and programmers would write down such operations, one instruction per line, telling the computer both the number of the operation and the numerical address of one or more pieces of data on which the operation was to be performed (**Fig. 1**). [Data were stored in the computer at a location identified by a numerical address.]

Example of relationship between assembler (mnemonic) and machine-languge code*

Mnemonic	Code	Instruction†
A	21	Add
AM	11	Add (I)
B	49	Branch
BB	42	Branch Back
BD	43	Branch on Digit
BI	46	Branch No Indicator
BNF	44	Branch No Flag
BNI	47	Branch on Indicator
BNR	45	Branch No Record Mark
BT	27	Branch and Transmit
BTM	17	Branch and Transmit (I)
C	24	Compare
CF	33	Clear Flag
CM	14	Compare (I)
DN	35	Dump Numerically
H	48	Halt
K	34	Control
M	23	Multiply
MM	13	Multiply (I)
NOP	41	No OPeration
RA	37	Read Alphamerically
RN	36	Read Numerically
S	22	Subtract
SF	32	Set Flag
SM	12	Subtract (I)
TD	25	Transmit Digit
TDM	15	Transmit Digit (I)
TF	26	Transmit Field
TFM	16	Transmit Field (I)
TR	31	Transmit Record
WA	39	Write Alphamerically
WN	38	Write Numerically

*From R. C. Houghton, Growing up with software tools, in R. L. Glass, *In the Beginning: Recollections of Software Pioneers*, pp. 272–291, IEEE Computer Society Press, 1998; © 1998 IEEE.
†I means Immediate.

Machine language was cumbersome, error-prone, and hard to change. In the latter 1950s, assembler (or assembly) language was invented. Assembler language was nearly the same as machine language, except that symbolic (instead of numerical) operations (see **table**) and symbolic addresses were used, making the code considerably easier to change, although it remained cumbersome and error-prone. Some assembler code is still written, usually for applications where extremely rapid code performance is needed, or where the code must interact at a fairly intimate level with the computer hardware.

The programmable aspects of computer hardware have not changed much since the 1950s. Computers still have numerical operations, and numerical addresses by which data may be accessed. However, programming languages have changed considerably. In the late 1950s, they began to move away from the machine view and toward a view of the problem being solved. Now programmers use

high-level languages, which look much more like English than a string of numbers or operation codes (**Fig. 2**).

Well-known programming languages include Basic, Java, and C. Basic is a fairly simple language, easily taught yet general in application. It has been modified into Visual Basic, a language useful for writing the portion of a program that the user "talks to." (This portion of a program is called the user interface. Since modern interfaces tend to be pictorial or "graphical" for ease of use, most such program portions are now called graphical user interfaces or GUIs.) Java is especially useful for creating software that runs on a network of computers, and is designed so that a program created to be used on one computer can be easily moved to a different kind of computer and run there with minimal change. (This so-called porting of software from one machine to another is normally extremely complicated and error-prone.) C and C^{++} are powerful but complex languages for writing such software as systems software and games. *See* HUMAN-COMPUTER INTERACTION; LOCAL-AREA NETWORKS; WIDE-AREA NETWORKS.

COBOL (Common Business-Oriented Language) and Fortran (for formula translation) are heavily used in spite of their origins in the 1950s. They are useful for solving, respectively, business problems or scientific and engineering problems involving heavy use of mathematics. Both languages benefit from ongoing modernization programs; although their modern versions resemble those of the 1950s, the languages have been updated to reflect some (but not all) of the best of modern-day thinking regarding programming languages. *See* PROGRAMMING LANGUAGES.

Programs have never been written with 1's and 0's (binary numbers). They are represented in binary inside the computer, but the programmer never sees that representation. *See* NUMBERING SYSTEMS; NUMERICAL REPRESENTATION (COMPUTERS).

Types of software. There have been enormous changes in the software field. Perhaps the change of greatest significance to the general public is the development of already-built software that may be purchased. In the early days of software, when none had been created before, all software was built from scratch. A large amount of such custom-built software is still constructed for the specific needs of its users, since new tasks continue to be invented for computers and software, and since the intricate details of software often require that a new variant be custom-built to perform a task similar to one done before.

Some portions of custom-built software may be reused from previous products, although the ability to reuse components eludes the plug-together-software vision of those who advocate massive-scale reuse. Frequently, software can be modified from older versions rather than built from scratch to perform a task similar to those done before. This task is known as software maintenance. (Traditionally, in other fields, the word "maintenance" means to repair faults. Fault repair in software is also known as maintenance, but that activity is less common than enhancing old software to do new tasks.) Old software that continues to be modified to do new things is often referred to as legacy software.

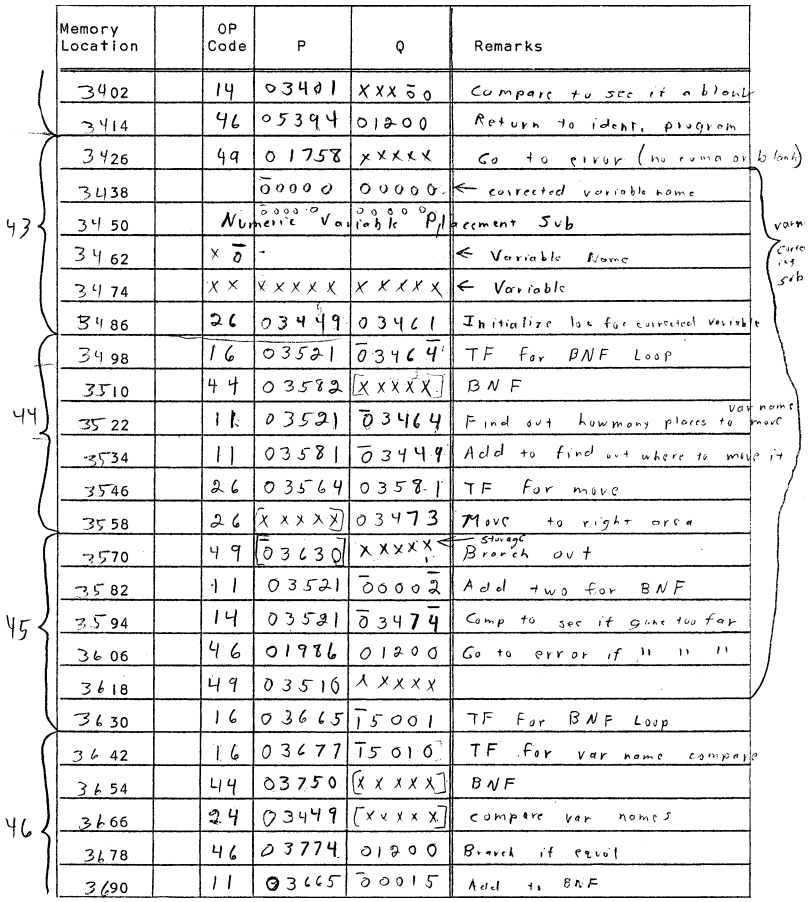


Fig. 1. Example of machine-language code. (*From R. C. Houghton, Growing up with software tools, in R. L. Glass, In the Beginning: Recollections of Software Pioneers, pp. 272–291, IEEE Computer Society Press, 1998; © 1998 IEEE*)

Packaged software such as word processors, spreadsheets, graphics and drawing tools, email systems, and games are widely available and used. Some software packages are enormous; for example, enterprise resource planning (ERP) software can be used by companies to perform almost all of their so-called backoffice software work. Several vendors make ERP backoffice software so generalized that a company can use it for the majority of its accounting, marketing, personnel, manufacturing, and other information needs. *See* COMPUTER GRAPHICS; ELECTRONIC MAIL; VIDEO GAMES; WORD PROCESSING.

Because software is such a versatile product, it can be used for an enormously varied set of tasks. The number and variety of those tasks are so huge that it is difficult to even categorize them. One generally accepted categorization, however, identifies the following major areas.

Systems software. This is software necessary to support the running of an application program. Operating systems are needed to link the machine-dependent needs of a program with the capabilities of the machine on which it runs. Compilers translate programs from high-level languages into machine

```
procedure RecordErrorMessage
  (
     ErrorCode: ErrorCode_t;
     var Status: Status_t
     );
{ This routine outputs an error message based on an error code
  supplied by the calling routine. The way it outputs the message
  depends on the current processing state, which it retrieves
  on its own. It returns a variable indicating success or failure . }

var
     ProcessingMethod: ProcessingMethod_t;
     ErrorMessage:     Message_t;
     FileStatus:       Status_t;

begin
    { set the default status }
    Status := Failure;

     { look up the message based on the error code }
     LookupErrorMessage( ErrorCode, ErrorMessage );

     { if the error code is valid }
     if ( ErrorMessage.ValidCode ) then begin

          { determine the processing method }
          ProcessingMethod := CurrentProcessingMethod;

          { if doing interactive  processing }
          if ( ProcessingMethod = Interactive ) then begin

               { print the error message interactively and declare success }
               PrintInteractiveMessage( ErrorMessage. Text );
               Status := Success
          end

          { else doing batch processing }
          else if ( ProcessingMethod = Batch ) then begin

               { if the batch message file opens properly }
               FileStatus := OpenMessageFile;
               if ( FileStatus = Success ) then begin

                    { log the error message to the batch file, close the file,
                      and declare success }
                    LogBatchMessage( ErrorMessage.Text );
                    CloseMessageFile;
                    Status := Success
               end { if }
          end { else }
     end

     { else the message code is not valid }
     else begin

          { notify the user that an internal error has been detected }
          PrintInteractiveMessage( 'Internal Error: Invalid error code',
               'in RecordErrorMessage()' )
     end
```

Fig. 2. Example of high-level (Pascal) code. (*From S. McConnell, Code Complete: A Practical Handbook of Software Construction, Microsoft Press, 1993; reproduced by permission of Microsoft Press; all rights reserved*)

languages. Database programs keep track of where and how data are stored on the various storage facilities of a typical computer, and simplify the task of entering data into those facilities or retrieving the data. Networking software provides the support necessary for computers to interact with each other, and with data storage facilities, in a situation where multiple computers are necessary to perform a task, or when software is running on a network of computers (such as the Internet or the World Wide Web). *See* DATABASE MANAGEMENT SYSTEM; INTERNET; OPERATING SYSTEM; TELEPROCESSING; WORLD WIDE WEB.

Business applications software. This software processes transactions, produces paychecks, and does the myriad of other tasks that are essential to running any business. ERP systems are cutting down on (but not eliminating) the need to build customized business systems. Roughly two-thirds of software applications are in the business area.

Scientific and engineering software. This software satisfies the needs of a scientific or engineering user to perform enterprise-specific tasks. For example, a rocket propulsion company might develop software to evaluate the thrust that a particular combination of propellants will produce when they are mixed in a propulsion chamber. Such software is often highly mathematical in nature. Because scientific and engineering tasks tend to be very enterprise-specific, there has been no generalization of this application area analogous to that of the ERP for backoffice business systems. The scientific-engineering application usually is considered to be in second place only to business software in terms of number of software products built.

Edutainment software. This software instructs (educates) or plays games with (entertains) the user. Such software often employs elaborate graphics and complex logic. This is one of the most rapidly growing software application areas, and includes software to produce special effects for movies and television programs.

Real-time software. This software operates in a time-compressed, real-world environment. Although most software is in some sense real-time, since the users of modern software are usually interacting with it via a GUI, real-time software typically has much shorter time constraints. For example, software that controls a nuclear reactor, or a bomb dropping from an aircraft, must make decisions and react to its environment in minuscule fractions of a second, not the 1–5-second response time typically needed by a human being. Software with such tight timing requirements often must be produced to even tighter constraints than other software; performance analysis of such software is essential to make sure that the software can perform its tasks in the available time.

The preceding areas are often known as application areas or application domains, since they involve applying the computer to a certain kind of problem's solution.

Internal functioning. A look at how software functions internally reveals several matters of interest.

Program listing. Software is an ethereal concept since it is a set of instructions, not a physical entity, and it has no weight and occupies no space. To the programmer, a program is embodied in its listing, the textual representation of the set of instructions, written in the language of the programmer's choosing. A listing may be printed on a ream of paper (the typical program contains many thousands, perhaps many millions, of programming language statements) or scrolled across a computer screen. To change, or add to, a program, the programmer uses a compiler to manipulate the statements of the programming language in much the same fashion as a word processor manipulates written text (with the big difference that, when the program is complete, the compiler translates the program into machine code).

Client-server and three-tier systems. The average program has at least three major parts. As software becomes more complex, those parts have come to represent three separate program portions.

One portion is the GUI. Expectations of users for their interaction with a program are far higher than in the past, and the GUI has, therefore, increased enormously in complexity. For significant software systems, it is not unusual for the GUI not only to be a separate piece of software but also to run on a separate computer. This GUI computer is often known as the client computer, because the client or user sits at that computer when using the system.

Another portion of the program interacts with the environment in which the system runs. Often, especially for business applications, that will involve manipulating data in a database. Because database access (or any other kind of environment interaction) is a very different problem from, for example, interacting with the user, this portion of the program also often runs on its own computer. This computer is called the server computer, since it serves up information from the environment to the remainder of the program. Often, especially for business applications, the software on the server is a database package especially built for convenient access to masses of data.

The third portion of a typical contemporary program is called the application portion. It contains those portions of the program that provide the solution to the user's specific problem. A decision must be made as to where the application portion should run. Running it on the client computer along with the GUI results in a so-called fat client (fat because the client computer contains more software than it would have without that decision). Running it on the server computer is sometimes called a fat server approach. But more often, the application program runs in its own computer. A three-computer solution is known as a three-tier system, and the application computer then becomes known as the middle-tier computer. For programs and systems of any complexity, the three-tier approach is becoming more common. *See* CLIENT-SERVER SYSTEM; DISTRIBUTED SYSTEMS (COMPUTERS).

Computing environment. The idea of a two- or three-tiered approach to computing hardware, and the resulting two- or three-way distribution of linked pieces of software, is relatively recent. Before such client-server or three-tiered solutions were employed, most software ran on a single computer. If the program in question was quite large, most often it would run on a mainframe computer. Mainframe computers are still used for the majority of legacy software and for some new custom software development. Mainframes are typically very large, powerful, and costly. Because mainframes have been around for many years, they provide software support often not present in the more recent computing approaches. For example, system security software, to prevent unauthorized users from accessing material intended to be secure, has matured in the mainframe environment, and is still considerably weaker in the more recent types of systems. *See* COMPUTER SECURITY.

Other kinds of computers for which application software is commonly written are minicomputers (mid-level, mid-speed, mid-price machines, typically used for mid-sized applications or as one element of a multitiered system) and microcomputers (now more commonly called personal computers or PCs), which have come to be ubiquitous and range in power and speed from modest to quite high. *See* MICROCOMPUTER.

Choice of programming language. With the advent of multiple program portions, software development has become considerably more complicated. The choice of a programming language, for example, is much more difficult. Whereas it was formerly considered sensible to develop all of a software system in the same programming language, now the different portions are often developed in entirely different languages.

The relatively complex GUI, for example, can most conveniently be developed in one of the so-called visual languages, since those languages (such as Visual Basic) contain powerful facilities for creating it. Such languages are said to be event-driven, helping the programmer write software that reacts to user-initiated events on the client computer screen.

The server software, on the other hand, will likely be built using a database package and the database language SQL (a Structured Query Language, for inquiring into the contents of a database). If the server software is also responsible for interacting with a network such as the Internet, it may also be coded in a network-support language such as Java. An object-oriented approach (which focuses on the objects to be manipulated by the program, rather than the functions to be performed or the data on which functions are to be performed) may be adopted in its development, since the software will need to manipulate objects on the Internet.

The best choice for the middle tier will probably be a language that facilitates writing software for the application domain in question. Even though COBOL and Fortran are quite old, their application focus makes them reasonable choices in this environment, if the application falls clearly into their domain. Alternatively, C and C^{++} are often used for writing the

middle tier of software. *See* COMPUTER PROGRAMMING; OBJECT-ORIENTED PROGRAMMING; SOFTWARE ENGINEERING. Robert L. Glass

Bibliography. F. P. Brooks, Jr., *The Mythical Man-Month: Essays on Software Engineering*, Addison-Wesley, 1995; T. DeMarco and T. Lister, *Peopleware: Productive Projects and Teams*, Dorset House, 1999; W. S. Humphrey, *Managing Technical People: Innovation, Teamwork, and the Software Process*, Addison-Wesley, 1997; S. McConnell, *Code Complete: A Practical Handbook of Software Construction*, Microsoft Press, 1993; M. C. Paulk et al., *The Capability Maturity Model: Guidelines for Improving the Software Process*, Addison-Wesley, 1995.

Software engineering

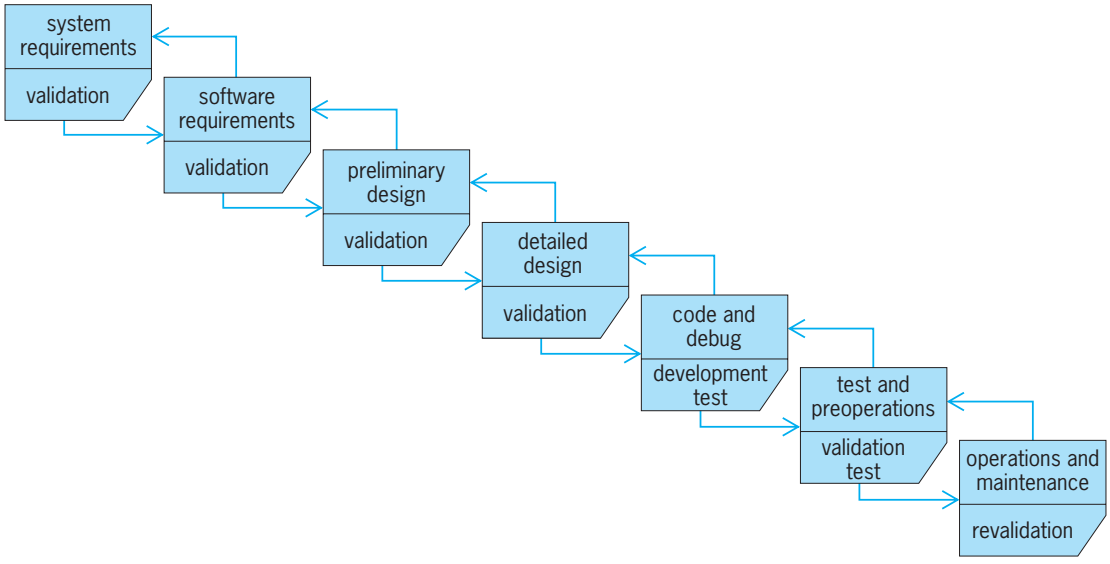
The process of manufacturing software systems. A software system consists of executable computer code and the supporting documents needed to manufacture, use, and maintain the code. For example, a word processing system consists of an executable program (the word processor), user manuals, and the documents, such as requirements and designs, needed to produce the executable program and manuals.

Software engineering is ever more important as larger, more complex, and life-critical software systems proliferate. The rapid decline in the costs of computer hardware means that the software in a typical system often costs more than the hardware it runs on. Large software systems may be the most complex things ever built. This places great demands on the software engineering process, which must be disciplined and controlled.

To meet this challenge, software engineers have adapted many techniques from older engineering fields, as well as developing new ones. For example, divide and conquer, a well-known technique for handling complex problems, is used in many ways in software engineering. The software engineering process itself, for example, is usually divided into phases. The definition of these phases, their ordering, and the interactions between the phases specify a software life-cycle model. The best-known life-cycle model is the waterfall model (see **illus.**) consisting of a requirements definition phase, a design phase, a coding phase, a testing phase, and a maintenance phase. The output of each phase serves as the input to the next. *See* SYSTEMS ENGINEERING.

Requirements. The purpose of the requirements phase is to define what a system should do and the constraints under which it must operate. This information is recorded in a requirements document. A typical requirements document might include a product overview; a specification of the development, operating, and maintenance environment for the product; a high-level conceptual model of the system; a specification of the user interface; specification of functional requirements; specification of nonfunctional requirements; specification of interfaces to systems outside the system under development; specification of how errors will be handled; a listing of possible changes and enhancements to the system; and other auxiliary information such as a glossary and index. Each requirement, usually numbered for reference, must be testable. The requirements definition process is difficult for several reasons.

Developing requirements requires that the customers of a system be identified and that their needs be understood. There may well be communication problems between customers who know their business lines but may be unfamiliar with computer technology, and software engineers who know computer technology but do not understand customers' business lines. Since there is a time lag between the requirements definition and the delivery of the system, the requirements of the customer may change while



Waterfall software life-cycle model. (*After B. W. Boehm, Software engineering, IEEE Trans. Comp., C-25:1226–1241, 1975; © IEEE*)

the system is being built. This problem is known as requirements volatility.

Another problem stems from the language that is used to express the requirements. Requirements stated in a natural language, such as English, are expressive and understandable by users of the system but may be imprecise. Formal languages such as predicate calculus, set-theoretic notations, and finite-state machines are precise but may lack expressiveness and may be difficult or impossible for end users of the system to understand. In practice, most requirements documents are written in natural language, but formal languages are sometimes used for parts of the system. Finite-state machines, for example, are often used to specify telecommunication system activities such as call handling.

The traceability problem refers to the difficulty of cleanly mapping a given requirement to its implementation in design and code. This makes it difficult to determine which parts of a system implement a given requirement and whether a given requirement has been completely and accurately implemented.

To deal with these problems, many approaches to requirements definition have been tried. One of the more successful is the JAD (joint application design) method, which uses a focused multiday meeting of customers and developers to produce a requirements document. JAD uses rapid prototyping to help specify user interfaces. Data collected from JAD sessions indicate that significant improvements in requirements productivity and quality can be achieved using JAD.

A prototype is a partial implementation of a software system. Throwaway prototypes are built to clarify or validate requirements or to explore feasibility and then are discarded. A breadboard prototype is built as the initial part of a system and is not discarded. A key to prototyping is to use a language tool that allows rapid prototype creation. For example, interface generators allow user interfaces to be drawn, and then the system generates the code for the interface.

Design. In the design phase, a plan is developed for how the system will implement the requirements. The plan is expressed using a design method and notation. Many methods and notations for software design have been developed. Each method focuses on certain aspects of a system and ignores or minimizes others. This is similar to viewing a building with an architectural drawing, a plumbing diagram, an electrical wiring diagram, and so forth.

Principles. Principles of software design have also emerged. Abstraction, the suppression of details, is used throughout the design process to identify key design objects and operations. Separation of concerns guides the decomposition of a system into smaller parts that can be solved separately. Reusable patterns and components are used to improve system quality and reliability. Design tools support design tasks, though they primarily support the drawing of diagrams using a particular design notation and database functions, rather than guiding the design process per se.

Researchers have identified several common architectural styles for software systems. These include the pipeline architecture where the system is viewed in terms of data flowing through a pipe and transformed by data filters, data abstraction where each subsystem hides a type representation, a repository architecture where computational processes write to and read from a common data store, and a layered architecture where each architectural layer provides certain services and hides lower levels of abstraction.

A key method for handling the complexity of large software systems is modularity. A system is modular if it is composed of simple independent parts that communicate through well-defined interfaces. The conceptual simplicity of the modular system parts allows them to be more easily documented and tested and to be better maintained.

Another key method for handling design complexity is information hiding, a decomposition method where each module hides internal processing details from other modules. Each module thus has two parts; a public part with information needed to use the component, and private information such as implementation details that a user does not need to know.

Two other important design principles are coupling and cohesion. Cohesion refers to the degree to which parts of a program unit, for example functions in a module, are related. A highly cohesive module performs a single task and can usually be described by a specific verb and object like "print a file" or "sort a list." The best type of cohesion is among elements that jointly implement an abstract data type. More cohesive modules tend to be more modular and to be easier to understand and document. Coupling refers to the strength of the linkage between two modules, based on the amount and kinds of information they exchange. Minimal coupling is a design goal. The worst type of coupling is among modules that share global data (data that are defined outside the scope of individual modules). The best is coupling where modules manipulate the same data type.

Design methods. The most common design methods are structured design, data structure design, dataflow design, and object-oriented design. Each of these methods has a defined process to guide design and a set of notations for recording the design. One of the earliest methods of design focused on the flow of control in a software system based on program branching and functional activities. This information was recorded in flowcharts which provided graphical symbols for procedures, decisions, starting and stopping, and connecting pages of the diagram. This method has fallen into disfavor because flow of control is thought to be a low-level implementation detail that should be specified at the end, rather than the beginning, of the design process.

Structured design, sometimes called top-down design, begins with a high-level functional view of a system and progressively decomposes it into lower and lower levels of detail. Structure charts are one notation for recording top-down designs. This

notation provides symbols for system units at various levels of decomposition, and connectors that show how data are passed among units. Structured design is widely used and is fairly easy to understand, but fails to provide detailed guidance for system decomposition.

Dataflow design focuses on the data that flows through the system and on the processes needed to transform it. Flow of control issues are deferred until the end of the design process. Dataflow notation provides symbols for data flowing through the system, data transforms, logical alternation, and the beginning and end of data processes. Data dictionaries provide precise definitions of the data in the system, and pseudo-code is used in the final steps of the design process to record flow of control sequences. Dataflow design provides a complete method, and dataflow notation is useful in communicating system development issues to end users. The method is less appropriate for systems that are algorithmically intense, such as mathematical software. The method also provides little opportunity to incorporate abstract data types. *See* DATAFLOW SYSTEMS.

A data type is a set of data values and allowed operations on those values. For example, an integer data type has integers as its data values and the operations addition, subtraction, multiplication, and division. An abstract data type is one with implementation details abstracted away, providing an implementation-neutral mathematical abstraction. Formally defining an abstract data type involves developing a name for it, defining the sets needed for its description, defining the syntax of its operations, and defining the semantics of the operations. *See* ABSTRACT DATA TYPE; DATA STRUCTURE.

Data abstraction is a key idea in object-oriented design, which views a software system as a collection of objects which communicate via message passing. An object is an encapsulated collection of data and operations, called methods. A message consists of the name of an object, the name of one of its methods, and the parameters needed to use the method. A class is a specification of the shared features of related objects. Objects belonging to a class are said to be instances of the class. Inheritance is the process by which one class acquires the variables and operations of another. *See* OBJECT-ORIENTED PROGRAMMING.

Coding. The coding phase of the software lifecycle is concerned with the development of code that will implement the design. This code is written is a formal language called a programming language. Programming languages have evolved over time from sequences of ones and zeros directly interpretable by a computer, through symbolic machine code, assembly languages, and finally to higher-level languages that are more understandable to humans. *See* PROGRAMMING LANGUAGES.

Most coding today is done in one of the higher-level languages. When code is written in a higher-level language, it is translated into assembly code, and eventually machine code, by a compiler. Many higher-level languages have been developed, and they can be categorized as functional languages, declarative languages, and imperative languages.

Functional languages are modeled on mathematical function theory. Computations in these languages, such as Lisp and ML, are specified using the application of functions.

Declarative languages are modeled on formal logic. Computations in these languages, such as Prolog, which is based on a subset of predicate calculus, are specified by stating facts and using logical implication.

Imperative (algorithmic) languages, such as Fortran, Cobol, and C, use a precise series of instructions to specify computations. Object-oriented languages, such as C^{++} and Java, are a special kind of imperative language that adds constructs to support object-oriented programming. These constructs include class declarations and inheritance. Almost all engineering of large software systems is done using an imperative language.

Following the principle of modularity, code on large systems is separated into modules, and the modules are assigned to individual programmers. A programmer typically writes the code using a text editor. Sometimes a syntax-directed editor that "knows" about a given programming language and can provide programming templates and check code for syntax errors is used. Various other tools may be used by a programmer, including a debugger that helps find errors in the code, a profiler that shows which parts of a module spend most time executing, and optimizers that make the code run faster.

Testing. Testing is the process of examining a software product to find errors. This is necessary not just for code but for all life-cycle products and all documents in support of the software such as user manuals.

The software testing process is often divided into phases. The first phase is unit testing of software developed by a single programmer. The second phase is integration testing where units are combined and tested as a group. System testing is done on the entire system, usually with test cases developed from the system requirements. Acceptance testing of the system is done by its intended users.

Testing can be done either top-down of bottom-up. In top-down testing, high-level routines are implemented and tested and then used as a testing environment, called a test harness, for lower-level routines. Bottom-up testing proceeds by first developing low-level routines, testing them, and then progressively combining and testing them as parts of larger and larger program units. In practice, both methods are used together in a process called sandwich testing.

The basic unit of testing is the test case. A test case consists of a test case type, which is the aspect of the system that the test case is supposed to exercise; test conditions, which consist of the input values for the test; the environmental state of the system to be used in the test; and the expected behavior of the system given the inputs and environmental factors.

Test cases can be either black-box or white-box.

Black-box test cases are developed without reference to the internal structure of the code under test—that is, the code is treated as a black box. In white-box testing, the internals of the system are examined to help develop test cases.

Ideally, one would like to test all possible input-output-environment combinations, but this is impossible because such a set is typically very large, or even infinite. To address this problem, a method called equivalence partitioning is used. In equivalence partitioning, the test input and environment space are divided into classes based on common attributes. A few members of each class are then selected as representative test cases for the class.

Since it is impossible to test software exhaustively, various heuristics are used to allocate test cases to different portions of the system. One heuristic is to focus testing on the parts of the system that are most crucial. Another is to allocate testing resources on estimates of where the most errors are likely to be found, based on the size or complexity of system parts or on the parts that consume the most processing time.

Another practical problem is knowing when enough testing has been done and the product is ready to be released. This can be addressed using a cumulative fault plot, a bivariate graph of the cumulative faults found via testing versus time. Such a plotted curve typically displays an S shape since at first few faults are found as the testing process gears up, then the curve rises steeply as most of the errors are found, and finally the curve flattens again since the remaining errors are more difficult to detect. A system is considered ready to ship when the final part of the curve remains consistently flat.

In white-box testing, the aim is to ensure that all of the code has been tested. This measurement can be in terms of lines of code, branch points, or execution paths through the program. Tools to support white-box testing instrument (add tracking code to) assembly or source code in order to keep track of which parts of the code have been exercised by test cases and which have not.

The coverage analysis process uses this information to design test cases that cumulatively test the code to the desired percentage. The process is to first examine the code to see which parts have not been exercised, design test cases to exercise the untested code, run the test cases, and then run a coverage analysis tool. If the desired level of test coverage has been reached, the process ends; otherwise, the process is repeated.

When software is changed to fix a bug or add an enhancement, a serious error is often introduced. To ensure that this does not happen, all test cases must be rerun after each change. The process of rerunning test cases to ensure that no error has been introduced is called regression testing. Various tools, including Unix shell scripts, can be used to support regression testing. *See* SOFTWARE TESTING AND INSPECTION.

Walkthroughs and inspections. Walkthroughs and inspections are used to improve the quality of the software development process. Consequently, the software products created by the process are improved. A quality system is a collection of techniques whose application results in continuous improvement in the quality of the development process. Elements of the quality system include reviews, inspections, and process audits.

A quality system should address two different aspects of quality: process and product quality. Process quality is the fitness of the development process for producing a product, while product quality is the fitness for use of the product resulting from some process. Quality systems should answer two basic questions about process and product quality: Are we building the right product (assessing product quality)? Are we building the product right (assessing process quality)?

Many of the things that must be tested in the software engineering process are not amenable to the kinds of techniques discussed for software. For the many textual documents that are typically produced for a software product and for software engineering processes, the best available testing methods are walkthroughs and inspections. Since code is also a kind of document, it can be tested with walkthroughs and inspections. Several types of walkthroughs and inspections are used.

A software review is an objective evaluation of a software engineering product or process. Though informal reviews, lacking systematic procedures and structures, may be adequate for small simple projects, larger projects will benefit from formal reviews. According to data collected on many projects, inspections can detect errors earlier in the life-cycle. This is important because the cost to repair errors rises rapidly as the software life-cycle progresses. Earlier detection via inspections can give up to 25% improvements in productivity. Other data indicate that defects found by inspections can reduce costs tenfold.

Code inspections are group readings of code by its designer, programmer, and tester. A moderator manages the inspection. During inspection, the programmer reads the code aloud and describes the intent of each statement. Errors are documented and categorized, but no attempt is made to fix the error during the inspection; this is done later. Average inspection rates for code are about 90 noncommentary source lines per hour. The inspection process can be guided by an error checklist of kinds of errors that might be encountered such as data declaration errors, computation errors, and control flow errors.

Another kind of inspection is a process audit whose purpose is to improve quality and productivity by improving the software production process. The best-known process evaluation for software is the SEI (Software Engineering Institute) process capability maturity model (CMM). This model is based on Philip Crosby's model for general process quality improvement. Crosby's model is based, in turn, on one of the basic principles of modern quality theory, which is that the best way to assure the quality of a product is to control the process which produces it. *See* QUALITY CONTROL.

The CMM is an evaluation model that classifies a process as being at one of five levels: (5) Optimizing — constant improvement based on measurement and feedback. (4) Managed—measured process and quality and productivity. (3) Defined—process defined and institutionalized. (2) Repeatable—process dependent on individuals, but basic project controls established. (1) Initial—no formal procedures or management control. The highest level, 5, includes continual process improvement, another hallmark of modern quality theory. Experience has shown that most software organizations today are at level 1, and only a very few are at level 5.

Maintenance. Large software systems are not static; rather, they change frequently both during development and after deployment. Maintenance is the phase of the software life-cycle after deployment. The maintenance phase may cost more than all of the others combined and is thus of primary concern to software organizations. The Y2K problem was, for example, a maintenance problem.

Activities. Maintenance consists of three activities: adaptation, correction, and enhancement. Enhancement is the process of adding new functionality to a system. This is usually done at the request of system users. This activity requires a full life-cycle of its own. That is, enhancements demand requirements, design, implementation, and test. Studies have shown that about half of maintenance effort is spent on enhancements.

Adaptive maintenance is the process of changing a system to adapt it to a new operating environment, for example, moving a system from the Windows operating system to the Linux operating system. Adaptive maintenance has been found to account for about a quarter of total maintenance effort. Corrective maintenance is the process of fixing errors in a system after release. Corrective maintenance takes about 20% of maintenance effort.

Software configuration management. Since software systems change frequently over time, an important activity is software configuration management. This consists of tracking versions of life-cycle objects, controlling changes to them, and monitoring relationships among them. Configuration management activities include version control, which involves keeping track of versions of life-cycle objects; change control, an orderly process of handling change requests to a system; and build control, the tracking of which versions of work products go together to form a given version of a software product.

Version control is supported by version control tools such as the Revision Control System (RCS) and Source Code Control System (SCCS). These tools support the storage and retrieval of versions of text files. To use them, a file is put into the change control system. After a version of a file is retrieved and modified, it is resubmitted to the system, which calculates changes between the old and new versions and stores only the changes, rather than both complete versions. This greatly reduces storage overhead. The system assigns unique numbers to versions and variations, allowing an orderly tracking process. A version replaces the original file, and a variation is an alternative to the original file.

Change control must be done in an orderly way on a large project. When a change is requested to some life-cycle work product, a form, sometimes called a modification request (MR), is created and logged into a change control database. A review board decides the severity of the problem described in the modification request and how the modification request should be handled, and then assigns personnel to address the handling of it.

Build control is concerned with how to put various versions of life-cycle work products together to build a system or part of a system. The make tool on UNIX, for example, allows the specification of which versions of software modules are needed to build an executable program, and the commands to actually do the build. In the C language, for example, an executable file depends on the object and library files needed to generate it. These, in turn, depend on source code files written in the C language, which are used by the compiler to create the object and library files.

Software reuse. The methods described so far, while effective in many cases, may not be adequate for building systems of the size and complexity sometimes required. One method that has been proposed to address this problem is software reuse, that is, the use of knowledge or artifacts from previous systems to build new ones. While software reuse has always been done in some way, recent research has attempted to get large productivity and quality gains by focusing on reuse that is systematic and domain-based.

A key idea in systematic domain-based reuse is that organizations seldom build completely new software systems. Rather, they build similar systems with variants to address different customer needs. These similar systems are said to be in a domain, which may be thought of informally as a business line. For example, a company's domain may be word processing software. The company will produce new versions with added functionality over time and may well provide variants for different operating systems.

Part of the impetus for more software reuse has come from considering other engineering disciplines. In electrical engineering, for example, systems are constructed from standardized components such as integrated circuits, resistors, and capacitors. A study of the history of hardware manufacturing has shown that that field made the transition from craft-based manufacturing to manufacturing based on interchangeable parts around 1840. Software production is in this transition process now.

Software reuse can be either horizontal or vertical. Horizontal reuse is reuse across domains. Much reuse is horizontal. For example, standard function libraries that support activities like searching and sorting are of this kind. Vertical reuse is reuse within a domain. For example, a component library might be

developed specifically for word processing systems that includes modules to handle text buffers, font management, and so on.

Actions that can produce more reuse fall into four categories: economic, legal, technical, and managerial. Clearly reuse improvement is undertaken to produce some benefit. The bottom-line benefit for a company will be greater profits. More immediate benefits, presumed to improve profits, include higher productivity, better quality, faster time to market, and improved bid and late life-cycle estimation. There is evidence that reuse can indeed produce these results under the right circumstances. These circumstances will be different for different organizations, requiring the proper blend of managerial, technical, legal, and economic factors.

Managerial factors include a rewards program for producers and consumers of reusable assets. Making an asset reusable will require more effort than making an equivalent one-use asset. This additional cost must be amortized across the multiple uses of the component. Incentives for producers of components must be established. Some companies, for example, provide cash bonuses for reusable assets accepted by the company reuse library. Since most software engineers are not currently trained to do systematic reuse and domain engineering, management must also provide education and motivation.

Legal factors may become the greatest barrier to systematic reuse. Current laws regarding software copyrights, patents, and liabilities are unclear. For example, if a company sells an asset and it is integrated into an application that fails because of the asset, causing loss of life or property, does the purchasing organization have the right to sue for damages? Another problem is that many software algorithms are now being patented each year. This may, in the future, provide a disincentive to reuse.

Technical factors include the issues of how to create reusable assets, how to design new applications to take advantage of them, and what kind of software engineering environment is needed to support reuse. Other technical factors concern what kind of software development process is needed and how to create and maintain a reuse library.

Economic factors include how to do a cost benefit analysis of reuse, performance measurement of a reuse program, methods for system costing, product pricing criteria and strategies to take reuse into account, and how to finance support costs for a reuse program.

A reuse library consists of a database for storing reusable components and a search interface that allows users to find and retrieve items in the database. Components for the database can be purchased, built, or reengineered from existing software. Before being added to the library, the components must be certified to assure that they meet quality standards, and must be classified.

The most common method in classifying reusable components is enumerated classification, where the subject area is broken into mutually exclusive, usually hierarchical classes. Other classification methods are faceted classification where the subject area is analyzed into basic terms which are organized as attribute-value pairs, and free-text keyword where index terms are extracted directly for the reusable component description. Reuse library tools should ideally be integrated with tools for configuration management, since the reusable assets in the database will change as error corrections and enhancements are made.

Domain engineering. Domain engineering is the process of analyzing a domain, recording important information about the domain in domain models, creating reusable assets, and using the domain models and assets to create new systems in the domain. It has two phases: domain analysis and domain implementation.

Domain analysis is the process of finding common and variable parts of systems in a domain and constructing domain models. Code, documents, and expert information are used as inputs to the domain analysis process. A domain vocabulary is developed which is used to describe systems in the domain. A key model that must be developed is a generic architecture that describes all of the systems in the domain. Reusable components will also often be identified in the domain analysis process.

These models are used to develop tools to speed the process of developing systems in the domain. One tool is domain-specific components that can be integrated into systems in the domain. Domain-specific programming languages can also be developed. These languages can be either large complete languages that include domain-specific operators and operands, or little languages that are small domain-specific languages developed for a specific purpose. Another approach, if sufficient formal information about the domain is known, is to develop application generators.

Application generators are programs that produce all or most of a software system when given a high-level specification. Application generators can be particularly powerful tools for reuse, giving great quality and productivity improvements. To develop an application generator, a language must be developed that allows the desired system to be specified. A specification in this language is then passed to the generator that produces the code for the application in the desired target language. Lex, for example, is an application generator for lexical analyzers. The specifications for a desired lexical analyzer are written in the lex language. The lex generator then produces the C code to implement the lexical analyzer. *See* SOFTWARE.

William B. Frakes

Bibliography. B. I. Blum, *Software Engineering: A Holistic Approach*, Oxford University Press, 1992; E. J. Braude, *Software Engineering: An Object-Oriented Perspective*, Wiley, 2000; M. Dorfman and R. H. Thayer (eds.), *Software Engineering*, IEEE Computer Society, 1996; R. S. Pressman, *Software Engineering: A Beginner's Guide*, McGraw-Hill, 1988; R. S. Pressman, *Software Engineering: A*

Practitioner's Approach, 5th ed., McGraw-Hill, 2000; I. Sommerville, *Software Engineering*, 6th ed., Addison-Wesley, 2000.

Software metric

A rule for quantifying some characteristic or attribute of a computer software entity. For example, a simple one is the FileSize metric, which is the total number of characters in the source files of a program. The FileSize metric can be used to determine the measure of a particular program, such as 3K bytes. It provides a concrete measure of the abstract attribute of program size. Of course, there are better size metrics. The FileSize metric is used for source programs; other metrics can be used for software entities such as requirements documents, design object models, or database structure models. Metrics for requirements and design documents can be used to guide decisions about development and as a basis for predictions, such as for cost and effort. Metrics for programs can be used to support decisions about testing and maintenance and as a basis for comparing different versions of programs. Ideally, metrics for the development cost of software and for the quality of the resultant program are desirable; but, lacking those, metrics for size and complexity are commonly used.

The fundamental property that must be satisfied by all software metrics is the representation condition. It requires that if one entity is less than another entity in terms of a selected attribute, then any metric for that attribute must associate a smaller number to the first entity than it associates with the second entity. For example, the FileSize metric is not truly a metric since it does not meet the representation condition. It is possible to take a "small" program and replace all variables with very long names, so that the measure of FileSize of this "small" program is larger than that of some "larger" program that uses short variable names.

Size metrics. The earliest size metric for programs was Lines of Code (LOC). Unfortunately, LOC, although commonly used, is not a good predictor of cost of development, likelihood of errors, and so forth. Additionally, there is no standard way to measure LOC; variations include counting or not counting blank lines, comments, declarations, and extra lines used for formatting.

Several metrics that better define program size have been introduced. The number of unique operators, η_1 (operators are symbols such as + and function and procedure names), and unique operands, η_2 (generally operands are the constants and identifiers in a program), can be used to predict software size according to the equation below. Similar metrics

$$N = \eta_1 \log_2 \eta_1 + \eta_2 \log_2 \eta_2$$

that relate to complexity have been defined, but they have not been widely used.

Complexity metrics. These metrics are intended to provide a quantitative measure of the abstract attribute of complexity. Intuitively, if one software entity is more complex than another software entity, then it follows that the more complex entity may have more errors, may be harder to understand, or may be more difficult to maintain. Unfortunately, complexity metrics provide only an imperfect measurement of the attribute of complexity.

A commonly used complexity metric for programs is based on the cyclomatic number used in graph theory. The cyclomatic number on the control flow graph of a program or design pseudocode has been shown to be equal to one plus the number of decisions in the control graph. A good rule of thumb is that program modules with cyclomatic number greater than ten should be avoided; that is, these programs need to be rewritten.

Whereas the Cyclomatic Number metric is intended to measure complexity of single program units (such as a function, procedure, or method), a number of metrics have been introduced to measure the complexity of object models and programs. In object-oriented software, objects can be inherited from other objects, and one metric is the length of these inheritance lists. Objects can also call functions in other objects, and the number of these calls can be measured. Since functions within an object use data local to the object, lack of cohesion can be measured by noting how many pairs of functions do not use any common data. *See* OBJECT-ORIENTED PROGRAMMING.

Software requirements metric. The Function Points (FP) metric has been developed for application to software requirements. It is used to predict the size and cost of development of a proposed software entity. The developer identifies and counts all inputs, outputs, inquiries, files, and interfaces and rates each as simple, average, or complex. Additional characteristics of the project are ranked on a scale of 1 to 5.

Types of scales. The area of measure theory specifies five possible types of measurement scales: nominal, ordinal, interval, rational, and absolute. Every software metric must be one of the five types of scales. The type of scale determines the types of statistics that can be applied to the metric. A metric must be at least an interval scale to take averages and must be at least a rational scale to consider ratios between values. For example, the cyclomatic number is an interval scale, and thus it would be correct to calculate the average cyclomatic number of a set of modules. However, it could not be correctly said that one module has twice the cyclomatic number of another module. For example, if a new module was created from two copies of a module, the new module would not have twice the cyclomatic number of the original module. *See* MEASURE THEORY.

Goal-Question metric (GQM). This is a method for developing metrics that will be more acceptable to developers. The first step is to select goals that should be maximized or minimized. For example, a goal

might be customer satisfaction. Next, questions are developed that are related to the goals—for example, "Is the customer finding too many errors?" Finally, metrics are proposed that quantify answers to the questions. For example, a metric could be the number of customer-reported errors. Since the metrics are closely related to the goals, use of the metrics should contribute to achievement of the goal. *See* COMPUTER PROGRAMMING; SOFTWARE ENGINEERING.

David A. Gustafson; William Hankley

Bibliography. N. Fenton, *Software Metrics*, Chapman and Hall, 1991; A. Melton (ed.), *Software Measurement*, International Thomson Computer Press, 1995; H. Zuse, *Software Complexity Measures and Methods*, W. de Gruyter, 1991.

Software testing and inspection

When software does not operate as it is intended to do, a software failure is said to occur. Software failures are caused by one or more sections of the software program being incorrect. Each of these incorrect sections is called a software fault. The fault could be as simple as a wrong value. A fault could also be the omission of a decision in the program. Faults have many causes, including misunderstanding of requirements, overlooking special cases, using the wrong variable, misunderstanding of the algorithm, and even typing mistakes. Software that can cause serious problems if it fails is called safety-critical software. Many applications in aircraft, medicine, nuclear power plants, and transportation involve such software.

Software testing is the execution of the software with the purpose of detecting faults. Software inspection is a manual process of analyzing the source code to detect faults. Many of the same techniques are used in both procedures. Other techniques can also be used to minimize the possibility of faults in the software. These techniques include the use of formal specifications, formal proofs of correctness, and model checking. However, even with the use of these techniques, it is still important to execute software with test cases to detect possible faults.

Software testing involves selecting test cases, determining the correct output of the software, executing the software with each test case, and comparing the actual output with the expected output. More testing is better, but costs time and effort. The value of additional testing must be balanced against the additional cost of effort and the delay in delivering the software. Another consideration is the potential cost of failure of the software. Safety-critical software is usually tested much more thoroughly than any other software.

The number of potential test cases is huge. For example, in the case of a simple program that multiplies two integer numbers, if each integer is internally represented as a 32-bit number (a common size for the internal representation), then there are 2^{32} possible values for each number. Thus, the total number of possible input combinations is 2^{64}, which is more than 10^{19}. If a test case can be done each microsecond (10^{-6} second), then it will take thousands of years to try all of the possible test cases. Trying all possible test cases is called exhaustive testing and is usually not a reasonable approach because of the size of the task, as this example indicates.

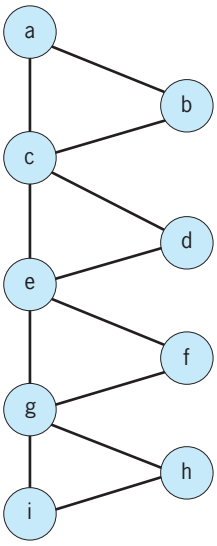
A test case should include the values of the inputs and the expected values of all outputs. In the following example, a test case includes three numbers that represent the inputs and a string that represents the value of the output. For example, (3 3 3 "equilateral") is a test case where the inputs are 3, 3, and 3, and the expected output is "equilateral."

The triangle problem, introduced by Glenford Myers, will be used to illustrate software testing. The problem is a simple one: Given the lengths of three sides of a triangle, tell whether the triangle is scalene, isosceles, or equilateral. The C++ code for one solution is as follows:

```
float a,b,c;
cin ≪ a ≪ b ≪ c;
char* type;
type = "scalene";
if (a==b || a==c || b==c)
{
        type = "isosceles";
}
if (a==b && a==c)
{
        type = "equilateral";
}
if (a>=b+c || b>=a+c || c>=a+b)
{
        type = "not a triangle";
}
if (a<=0 || b<=0 || c<=0)
{
        type = "bad inputs";
}
cout<<type;
```

In this solution, the three triangle types are possible outputs, but there are also two error conditions: The inputs may fail to be positive numbers, and the values may fail to be values of the sides of a legitimate triangle. The code has the property that the last "if" statement that is true for a particular test case is the one that determines the answer. For example, all of the "if" statements are true for the inputs 0 0 0, therefore the answer is "bad inputs."

Most of the testing criteria are easier to visualize using the control flow graph shown in the **illustration**. Each circle on the graph represents a block of code in the program. Each of these blocks is a section of code without any decisions inside the block. Thus, node *a* represents the code up to the logical expression of the first "if" statement. Node *b* represents the assignment of "equilateral" to the variable type. The arcs between the nodes represent the



Control flow diagram for the triangle problem.

possible flow of control. The arrow from node *a* to node *b* represents the flow of control if the first "if" statement is true. The arrow from node *a* to node *c* represents the flow of control if the first "if" statement is false.

Structural testing. One approach to software testing is to find test cases so that all statements in a program are executed. For example, the test case (3 4 5 "scalene") will execute all the statements except the assignment of "isosceles," "equilateral," "not a triangle," and "bad inputs" (nodes *a*, *c*, *e*, *g*, and *i* are covered). Additional test cases will be required to execute those assignments. This approach is a type of structural testing and is called "every statement coverage." Structural test methods use the code to select test cases. In this case the criterion is to execute all statements. "Every statement coverage" is often considered the minimal acceptable coverage for software testing. This coverage does not guarantee that every fault that exists in the program will be detected.

A more extensive criterion for test selection is "every branch coverage." This means that each branch coming out of every decision is tested. On the control flow graph "every branch coverage" is satisfied by executing every arc. A minimal test set that achieves this coverage would be (0 0 0 "bad inputs") and (3 4 5 "scalene").

Instead of just requiring the whole decision to be true or false, the "multiple condition coverage" criterion requires all combinations of truth values for each simple comparison in a decision to be covered. For example, the first decision statement is true if *a* equals *b* or *a* equals *c* or *b* equals *c*. The test case (4 3 3 "isosceles") makes the first comparison true (TFF). Because of how compilers handle multiple conditions, if a condition of a set of "or-ed" conditions is true, the conditions to the right will not be checked. Thus, this test case covers TFF, TTF, TFT, and TTT. Test case (3 4 3 "isosceles") covers FTF and FTT; (4 3 3 "isosceles") covers FFT; and (3 4 5 "scalene") covers FFF. (If two of the comparisons are true, then the third comparison is also true.)

Another approach is called dataflow coverage. The basis for coverage is the execution paths between the statement where a variable is assigned a value (a def or definition) and a statement where that value is used. These paths must be free of other definitions of the variable of interest. The basic criterion is to cover all def-use pairs. There are a number of variations. In the triangle problem, the variable of interest is type. There are five definitions of the variable type and one use of the variable type (the cout statement). Thus, test cases must be found for the paths *acegi*, *bcegi*, *degi*, *fgi*, and *hi*.

Functional testing. Testing compares the actual behavior of the software with the expected behavior. That expected behavior is usually described in a specification. In fact, software without a specification of what it should do could be considered correct because it does what it does. In functional testing, the test case selection is based on information found in the specification. In the triangle example, the specification says that software is supposed to identify the type of triangle. A very simple but effective approach is to ensure that all possible outputs are generated by the test set. This would require that the test set include an equilateral, an isosceles, and a scalene test case.

More involved functional test case selection involves analyzing the conditions inherent in the task. For the triangle problem, these conditions would be the definition of equilateral, isosceles, scalene, and being a triangle. For example, being a triangle requires that each side is less than the sum of the other two sides. Each valid combination of all these conditions would be considered to be a subdomain that should be tested. The test criterion would be to include at least one test case from each subdomain.

Boundary testing. Another approach to test selection concentrates on the boundaries between the subdomains. This approach recognizes that many faults are related to the boundary conditions. Test cases are chosen to check whether the boundary is correct. Test cases on the boundary and test cases just off the boundary are chosen. An example in the triangle problem is not-a-triangle decision. One boundary is $a = b + c$. Test cases such as (6 4 2 "not a triangle") and (100 50 50 "not a triangle") might be chosen on the boundary. An off-the-boundary test case might be (51 25 25 "isosceles").

Prospects. None of these approaches guarantees that faults will be detected. Current research is aimed at measuring the effectiveness of these testing approaches and investigating ways to improve the fault detection ability. Work is also being done on quantifying the effectiveness of testing in terms of predicting the reliability of software that has been successfully tested with a specific technique. *See* SOFTWARE; SOFTWARE ENGINEERING. David A. Gustafson

Bibliography. B. Beizer, *Software Testing Techniques*, 2d ed., Van Nostrand Reinhold, 1990; D. Gustafson, *Schaum's Outline of Theory and*

Problems of Software Engineering, McGraw-Hill, 2001; B. Marick, *The Craft of Software Testing*, Prentice Hall, 1995; G. Myers, *The Art of Software Testing*, Wiley, 1979.

Soil

Finely divided rock-derived material containing an admixture of organic matter and capable of supporting vegetation. Soils are independent natural bodies, each with a unique morphology resulting from a particular combination of climate, living plants and animals, parent rock materials, relief, the ground waters, and age. Soils support plants, occupy large portions of the Earth's surface, and have shape, area, breadth, width, and depth. (However, the term "soil" as used by engineers means unconsolidated rock material.) *See* PEDOLOGY.

Origin and Classification

Soil covers most of the land surface as a continuum. Each soil grades into the rock material below and into other soils at its margins, where changes occur in relief, ground water, vegetation, kinds of rock, or other factors which influence the development of soils. Soils have horizons, or layers, more or less parallel to the surface and differing from those above and below in one or more properties, such as color, texture, structure, consistency, porosity, and reaction (**Fig. 1**). The horizons may be thick or thin. They may be prominent or so weak that they can be detected only in the laboratory. The succession of horizons is called the soil profile. In general, the boundary of soils with the underlying rock or rock material occurs at depths ranging from 1 to 6 ft (0.3 to 1.8 m), though the extremes lie outside this range.

Origin. Soil formation proceeds in stages, but these stages may grade indistinctly from one into another. The first stage is the accumulation of unconsolidated rock fragments, the parent material. Parent material may be accumulated by deposition of rock fragments moved by glaciers, wind, gravity, or water, or it may accumulate more or less in place from physical and chemical weathering of hard rocks. *See* WEATHERING PROCESSES.

The second stage is the formation of horizons. This stage may follow or go on simultaneously with the accumulation of parent material. Soil horizons are a result of dominance of one or more processes over others, producing a layer which differs from the layers above and below.

Major processes. The major processes in soils which promote horizon differentiation are gains, losses, transfers, and transformations of organic matter, soluble salts, carbonates, silicate clay minerals, sesquioxides, and silica. Gains consist normally of additions of organic matter, and of oxygen and water through oxidation and hydration, but in some sites slow continuous additions of new mineral materials take place at the surface or soluble materials are deposited from ground water. Losses are chiefly of materials dissolved or suspended in water percolating through the profile or running off the surface. Transfers of both mineral and organic materials are common in soils. Water moving through the soil picks up materials in solution or suspension. These materials may be deposited in another horizon if the water is withdrawn by plant roots or evaporation, or if the materials are precipitated as a result of differences in pH (degree of acidity), salt concentration, or other conditions in deeper horizons.

Other processes tend to offset those that promote horizon differentiation. Mixing of the soil occurs as the result of burrowing by rodents and earthworms, overturning of trees, churning of the soil by frost, or shrinking and swelling. On steep slopes the soil may creep or slide downhill with attendant mixing. Plants may withdraw calcium or other ions from deep horizons and return them to the surface in the leaf litter.

Saturation of a horizon with water for long periods makes the iron oxides soluble by reduction from ferric to ferrous forms. The soluble iron can move by diffusion to form hard concretions or splotches of red or brown in a gray matrix. Or if the iron remains, the soil will have shades of blue or green. This process is called gleying, and can be superimposed on any of the others. *See* DIFFUSION.

The kinds of horizons present and the degree of their differentiation, both in composition and in structure, depend on the relative strengths of the processes. In turn, these relative strengths are determined by the way humans use the soil as well as by the natural factors of climate, plants

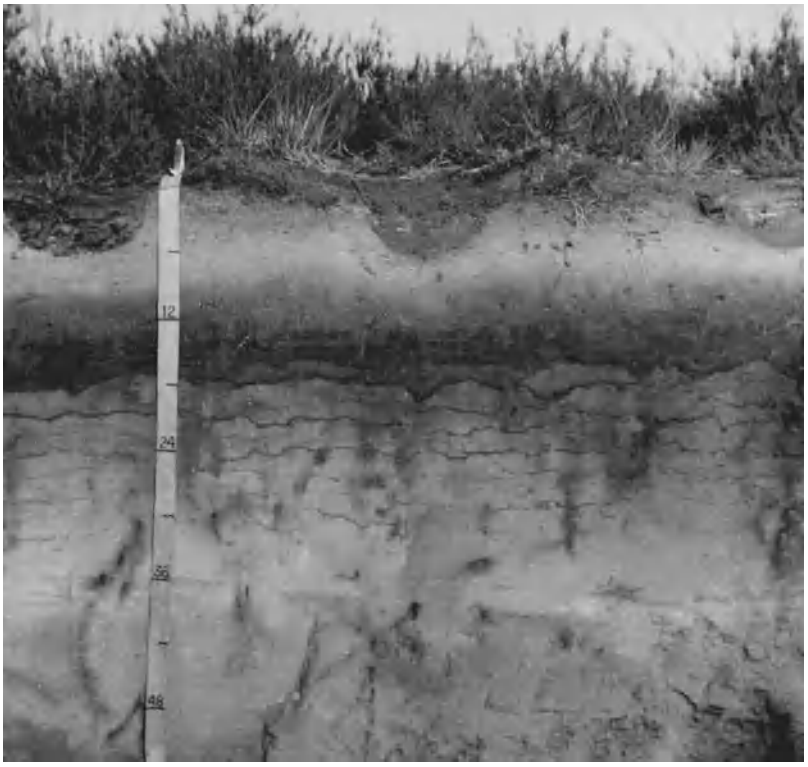


Fig. 1. Photograph of a soil profile showing horizons. The dark crescent-shaped spots at the soil surface are the result of plowing. The dark horizon lying 9–18 in. (23–45 cm) below the surface is the principal horizon of accumulation of organic matter that has been washed down from the surface. The thin wavy lines were formed in the same manner.

and animals, relief and ground water, and the period of time during which the processes have been operating.

Composition. In the drier climates where precipitation is appreciably less than the potential for evaporation and transpiration, horizons of soluble salts, including calcium carbonate and gypsum, are often found at the average depth of water penetration.

In humid climates, some materials normally considered insoluble may be gradually removed from the soil or at least from the surface horizons. A part of the removal may be in suspension. The movement of silicate clay minerals is an example. The movement of iron oxides is accelerated by the formation of chelates with the soil organic matter. Silica is removed in appreciable amounts in solution or suspension, though quartz sand is relatively unaffected. In warm humid climates, free iron and aluminum oxides and low-activity silicate clays accumulate in soils, apparently because of low solubility relative to other minerals. *See* CHELATION.

In cool humid climates, solution losses are evident in such minerals as feldspars. Free sesquioxides tend to be removed from the surface horizons and to accumulate in a lower horizon, but mixing by animals and falling trees may counterbalance the downward movement.

Structure. Concurrently with the other processes, distinctive structures are formed in the different horizons. In the surface horizons, where there is a maximum of biotic activity, small animals, roots, and frost action keep mixing the soil material. Aggregates of varying sizes are formed and bound by organic matter, microorganisms, and colloidal material. The aggregates in the immediate surface tend to be loosely packed with many large pores among them. Below this horizon of high biotic activity, the structure is formed chiefly by volume changes due to wetting, drying, freezing, thawing, or shaking of the soil by roots of trees swaying with the wind. Consequently, the sides of any one aggregate, or ped, conform in shape to the sides of adjacent peds.

Water moving through the soil usually follows root channels, wormholes, and ped surfaces. Accordingly, materials that are deposited in a horizon commonly coat the peds. In the horizons that have received clay from an overlying horizon, the peds usually have a coating or varnish of clay making the exterior unlike the interior in appearance. Peds formed by moisture or temperature changes normally have the shapes of plates, prisms, or blocks.

Horizons. Pedologists have developed sets of symbols to identify the various kinds of horizons commonly found in soils. The nomenclature originated in Russia, where the letters A, B, and C were applied to the main horizons of the black soils of the steppes. The letter A designated the dark surface horizon of maximum organic matter accumulation, C the unaltered parent material, and B the intermediate horizon. The usage of the letters A, B, and C spread to western Europe, where the intermediate or B horizon was a horizon of accumulation of free sesquioxides or silicate clays or both. Thus the idea developed that a B horizon is a horizon of accumulation. Some, however, define a B horizon by position between A and C. Subdivisions of the major horizons have been shown by either numbers or letters, for example, Bt or B2. No internationally accepted set of horizon symbols has been developed. In the United States the designations (**Fig. 2**) have been widely used since about 1935, with minor modifications made in 1962. Lowercase letters were added to numbers in B horizons to indicate the nature of the material that had accumulated. Generally, "h" is used to indicate translocated humus, "t" for translocated clay, and "ir" for translocated iron oxides. Thus, B2t indicates the main horizon of clay accumulation.

Classification. Systems of soil classification are influenced by concepts prevalent at the time a system is developed. Since ancient times, soil has been considered as the natural medium for plant growth. Under this concept, the earliest classifications were based on relative suitability for different crops, such as rice soils, wheat soils, and vineyard soils.

Early American agriculturists thought of soil chiefly as disintegrated rock, and the first comprehensive American classification was based primarily on the nature of the underlying rock.

In the latter part of the nineteenth century, some Russian students noted relations between the steppe and black soils and the forest and gray soils. They developed the concept of soils as independent natural bodies formed by the influence of environmental factors operating on parent materials over time. The early Russian classifications grouped soils at the highest level, according to the degree to which they reflected the climate and vegetation. They had classes of Normal, Abnormal, and Transitional soils, which later became known as Zonal, Intrazonal, and Azonal. Within the Normal or Zonal soils, the Russians distinguished climatic and vegetative zones in which the soils had distinctive colors and other properties in common. These formed classes that were called soil types. Because some soils with similar colors had very different properties that were associated with differences in the vegetation, the nature of the vegetation was sometimes considered in addition to the color to form the soil type name, for example, Gray Forest soil and Gray Desert soil. The Russian concepts of soil types were accepted in other countries as quickly as they became known. In the United States, however, the term "soil type" had been used for some decades to indicate differences in soil texture, chiefly texture of the surface horizons; so the Russian soil type was called a Great Soil Group.

Many systems of classification have been attempted but none has been found markedly superior; most systems have been modifications of those used in Russia. Two bases for classification have been tried. One basis has been the presumed genesis of the soil; climate and native vegetation were given major emphasis. The other basis has been the observable or measurable properties of the soil. To a

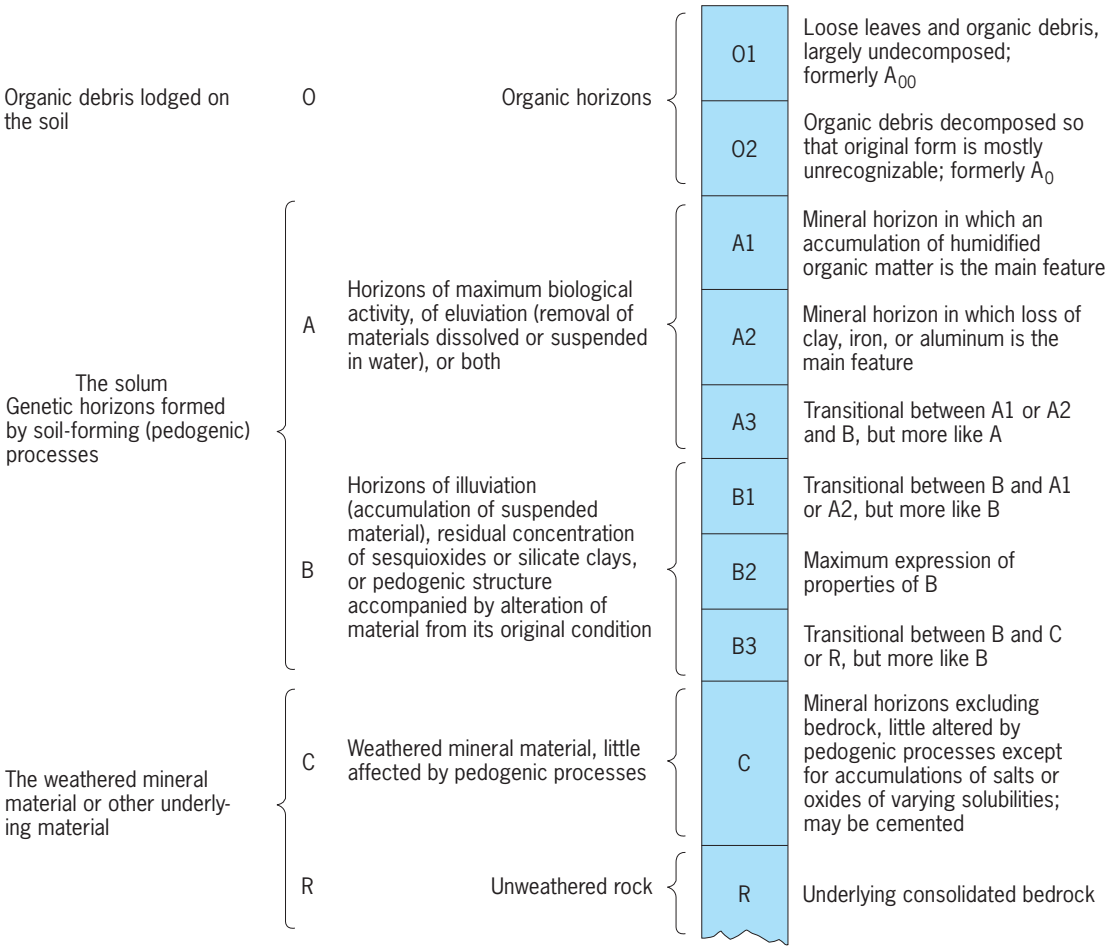


Fig. 2. Hypothetical soil profile having all principal horizons. Other symbols are used to indicate features subordinate to those indicated by capital letters and numbers. The more important of these are as follows: ca, as in Cca, accumulations of carbonates; cs, accumulations of calcium sulfate; cn, concretions; g, strong gleying (reduction of iron in presence of ground water); h, illuvial humus; ir, illuvial iron; m, strong cementation; p, plowing; sa, accumulations of very soluble salts; si, cementation by silica; t, illuvial clay; x, fragipan (a compact zone which is impenetrable by roots).

considerable extent, of course, these are used in the genetic system to define the great soil groups. The morphologic systems, however, have not used soil genesis as such, but have attempted to use properties that are acquired through soil development.

The principal problem in the morphologic systems has been the selection of the properties to be used. Grouping by color, tried in the earliest systems, produces soil groups of unlike genesis.

The Soil Survey staff of the U.S. Department of Agriculture (USDA) and the land-grant colleges adopted a different classification scheme in 1965. The system differs from earlier systems in that it may be applied to either cultivated or virgin soils. Previous systems have been based on virgin profiles, and cultivated soils were classified on the presumed characteristics or genesis of the virgin soils. The system has six categories, based on both physical and chemical properties. These categories are the order, suborder, great group, subgroup, family, and series, in decreasing rank.

Nomenclature. The names of the taxa or classes in each category are derived from the classic languages in such a manner that the name itself indicates the place of the taxa in the system and usually indicates something of the differentiating properties. The names of the highest category, the order, end in the suffix "sol," preceded by formative elements that suggest the nature of the order. Thus, Aridisol is the name of an order of soils that is characterized by being dry (Latin *arudys*, dry, plus *sol*, soil). A formative element is taken from each order name as the final syllable in the names of all taxa of suborders, great groups, and subgroups in the order. This is the syllable beginning with the vowel that precedes the connecting vowel with "sol." Thus, for Aridisols, the names of the taxa of lower classes end with the syllable "id," as in Argid and Orthid (**Table 1**).

Suborder names have two syllables, the first suggesting something of the nature of the suborder and the last identifying the order. The formative element "arg" in Argid (Latin *argillus*, clay) suggests the horizon of accumulation of clay that defines the suborder.

Great group names have one or more syllables to suggest the nature of the horizons and have the suborder name as an ending. Thus great group names

TABLE 1. Soil orders

Order	Formative element in name	General nature
Alfisols	alf	Soils with gray to brown surface horizons, medium to high base supply, with horizons of clay accumulation; usually moist, but may be dry during summer
Aridisols	id	Soils with pedogenic horizons, low in organic matter, and usually dry
Entisols	ent	Soils without pedogenic horizons
Histosols	ist	Organic soils (peats and mucks)
Inceptisols	ept	Soils that are usually moist, with pedogenic horizons of alteration of parent materials but not of illuviation
Mollisols	oll	Soils with nearly black, organic-rich surface horizons and high base supply
Oxisols	ox	Soils with residual accumulations of inactive clays, free oxides, kaolin, and quartz; mostly tropical
Spodosols	od	Soils with accumulations of amorphous materials in subsurface horizons
Ultisols	ult	Soils that are usually moist, with horizons of clay accumulation and a low supply of bases
Vertisols	ert	Soils with high content of swelling clays and wide deep cracks during some seasons

have three or more syllables but can be distinguished from order names because they do not end in "sol." Among the Argids, great groups are Natrargids (Latin *natrium*, sodium) for soils that have high contents of sodium, and Durargids (Latin *durus*, hard) for Argids with a hardpan cemented by silica and called a duripan.

Subgroup names are binomial. The great group name is preceded by an adjective such as "typic," which suggests the type or central concept of the great group, or the name of another great group, suborder, or order converted to an adjective to suggest that the soils are transitional between the two taxa.

Family names consist of several adjectives that describe the texture (sandy, silty, clayey, and so on), the mineralogy (siliceous, carbonatic, and so on), the temperature regime of the soil (thermic, mesic, frigid, and so on), and occasional other properties that are relevant to the use of the soil.

Series names are abstract names, taken from towns or places near where the soil was first identified. Cecil, Tama, and Walla Walla are names of soil series.

Order. In the highest category, 10 orders are recognized. These are distinguished chiefly by differences in kinds and amount of organic matter in the surface horizons, kinds of B horizons resulting from the dominance of various specific processes, evidences of churning through shrinking and swelling, base saturation, and lengths of periods during which the soil is without available moisture. The properties selected to distinguish the orders are reflections of the degree of horizon development and the kinds of horizons present.

Suborder. This category narrows the ranges in soil moisture and temperature regimes, kinds of horizons, and composition, according to which of these is most important. Moisture or temperature or soil properties associated with them are used to define suborders of Alfisols, Mollisols, Oxisols, Ultisols, and Vertisols. Kinds of horizons are used for Aridisols, compositions for Histosols and Spodosols, and combinations for Entisols and Inceptisols.

Great group. The taxa (classes) in this category group soils that have the same kinds of horizons in the same sequence and have similar moisture and temperature regimes. Exceptions to horizon sequences are made for horizons so near the surface that they are apt to be mixed by plowing or lost rapidly by erosion if plowed.

Subgroup. The great groups are subdivided into subgroups that show the central properties of the great group, intergrade subgroups that show properties of more than one great group, and other subgroups for soils with atypical properties that are not characteristic of any great group.

Family. The families are defined largely on the basis of physical and mineralogic properties of importance to plant growth.

Series. The soil series is a group of soils having horizons similar in differentiating characteristics and arrangement in the soil profile, except for texture of the surface portion, and developed in a particular type of parent material.

Type. This category of earlier systems of classification has been dropped but is mentioned here because it was used for almost 70 years. The soil types within a series differed primarily in the texture of the plow layer or equivalent horizons in unplowed soils. Cecil clay and Cecil fine sandy loam were types within the Cecil series. The texture of the plow layer is still indicated in published soil surveys if it is relevant to the use of the soil, but it is now considered as one kind of soil phase.

Classifications of soils have been developed in several countries based on other differentia. The principal classifications have been those of Russia, Germany, France, Canada, Australia, New Zealand, and the United States. Other countries have modified one or the other of these to fit their own conditions. Soil classifications have usually been developed to fit the needs of a government that is concerned with the use of its soils. In this respect soil classification has differed from classifications of other natural objects, such as plants and animals, and there is no international agreement on the subject.

Many practical classifications have been developed on the basis of interpretations of the usefulness of soils for specific purposes. An example is the capability classification, which groups soils according to the number of safe alternative uses, risks of damage, and kinds of problems that are encountered under use.

Surveys. Soil surveys include those researches necessary (1) to determine the important characteristics of soils, (2) to classify them into defined series and other units, (3) to establish and map the boundaries between kinds of soil, and (4) to correlate and predict adaptability of soils to various crops, grasses, and trees; behavior and productivity of soils under different management systems; and yields of adapted crops on soils under defined sets of management practices. Although the primary purpose of soil surveys has been to aid in agricultural interpretations, many other purposes have become important, ranging from suburban planning, rural zoning, and highway location, to tax assessment and location of pipelines and radio transmitters. This has happened because the soil properties important to the growth of plants are also important to its engineering uses.

Soil surveys were first used in the United States in 1898. Over the years the scale of soil maps has been increased from $^1/_2$ or 1 in. to the mile (8 or 16 mm to the kilometer) to 3 or 4 in. to the mile (47 to 63 mm to the kilometer) for mapping humid farming regions, and up to 8 in. to the mile (126 mm to the kilometer) for maps in irrigated areas. After the advent of aerial photography, planimetric maps were largely discontinued in favor of aerial photographic mosaics. The United States system has been used, with modifications, in many other countries. *See* AERIAL PHOTOGRAPH.

Two kinds of soil maps are made. The common map is a detailed soil map, on which soil boundaries are plotted from direct observations throughout the surveyed area. Reconnaissance soil maps are made by plotting soil boundaries from observations which are made at intervals. The maps show soil and other differences that are of significance for present or foreseeable uses.

The units shown on soil maps usually are phases of soil series. The phase is not a category of the classification system. It may be a subdivision of any class of the system according to some feature that is of significance for use and management of the soil, but not in relation to the natural landscape. The presence of loose boulders on the surface of the soil makes little difference in the growth of a forest, but is highly

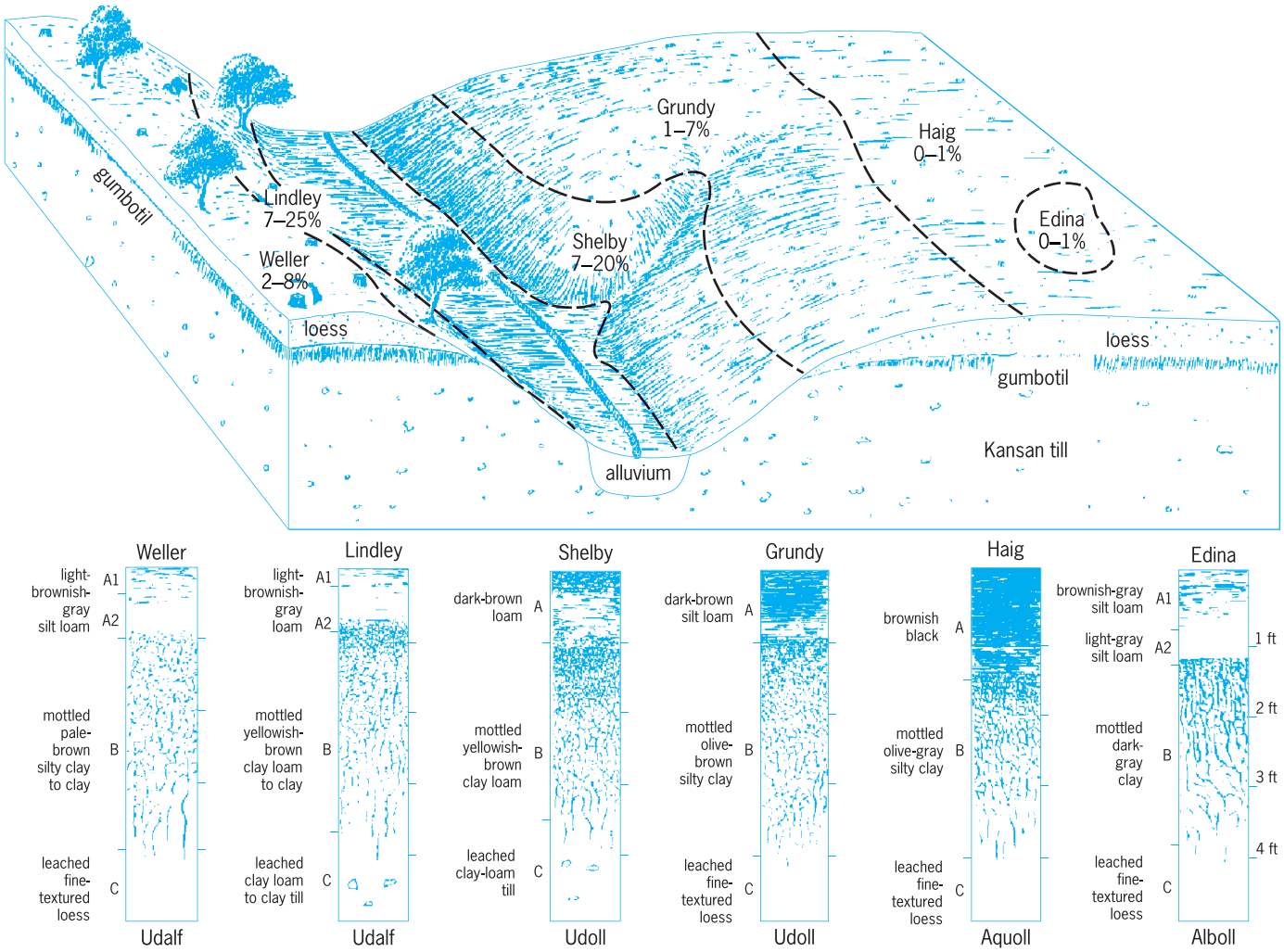


Fig. 3. Relation of the soil pattern to relief, parent material, and native vegetation on a farm in south-central Iowa. The soil slope gradient is expressed as a percentage. 1 ft = 30 cm. (*After R. W. Simonson et al., Understanding Iowa Soils, Brown, 1952*)

significant if the soil is to be plowed. Phases are most commonly based on slope, erosion, presence of stone or rock, or differences in the rock material below the soil itself. If a legend identifies a phase of a soil series, the soils so designated on a soil map are presumed to lie within the defined range of that phase in the major part of the area involved. Thus, the inclusion of lesser areas of soils having other characteristics is tolerated in the mapping if their presence does not appreciably affect the use of the soil. If there are other soils that do affect the use, inclusions up to 15% of the area are tolerated without being indicated in the name of the soil.

If the pattern of occurrence of two or more series is so intricate that it is impossible to show them separately, a soil complex is mapped, and the legend includes the word "complex," or the names of the series are connected by a hyphen and followed by a textural class name. Thus the phrase Fayette-Dubuque silt loam indicates that the two series occur in one area and that each represents more than 15% of the total area.

In places the significance of the difference between series is so slight that the expense of separating them is unwarranted. In such a case the names of the series are connected by a conjunction, for example, Fayette and Downs silt loam. In this kind of mapping unit, the soils may or may not be associated geographically.

It is possible to make accurate soil maps only because the nature of the soil changes with alterations in climatic and biotic factors, in relief, and in ground waters, all acting on parent materials over long periods of time. Boundaries between kinds of soil are made where such changes become apparent. On a given farm the kinds of soil usually form a repeating pattern related to the relief (**Fig. 3**).

Because concepts of soil have changed over the years, maps made 50 or more years ago may use the same soil type names as modern maps, but with different meanings.

Guy D. Smith

Soil Suborders

Soil suborders are broad classes at one level in the soil classification system adopted in the United States in 1965. A total of 47 suborders form the full set of classes in the second highest category (each category is a set of classes of parallel rank) in the system.

The number of local kinds of soils in a large country is also large. For example, 11,500 soil series have been recognized in soil surveys made in the United States through 1979. On the average, a series consists of six phases, which are the local kinds of soils. This means that approximately 66,000 local kinds have been defined up to the present time in a single large country, though all parts of that country have not been studied.

Despite the myriads of local kinds over the land surface of the Earth, all soils share some characteristics. They can all be related to one another in some way. The relationships are close for some pairs of local kinds and distant for others. The similarities and differences among the thousands of local kinds permit their grouping into sets of progressively broader classes in order to show degrees of kinship.

Though it is impossible for a single mind to retain concepts of 11,500 series or 66,000 phases, the salient features of a few dozen broad classes can be remembered. Consequently, the nature of the 47 suborders is described in this article. The purpose is to provide a general picture of the kinds of soils in the United States and the world. The 47 suborders grouped into the 10 orders are:

- Alfisols
  - Aqualfs
  - Boralfs
  - Udalfs
  - Ustalfs
  - Xeralfs
- Aridisols
  - Argids
  - Orthids
- Entisols
  - Aquents
  - Arents
  - Fluvents
  - Orthents
  - Psamments
- Histosols
  - Fibrists
  - Folists
  - Hemists
  - Saprists
- Inceptisols
  - Andepts
  - Aquepts
  - Ochrepts
  - Plaggepts
  - Tropepts
  - Umbrepts
- Mollisols
  - Albolls
  - Aquolls
  - Borolls
  - Rendolls
  - Udolls
  - Ustolls
  - Xerolls
- Oxisols
  - Aquox
  - Humox
  - Orthox
  - Torrox
  - Ustox
- Spodosols
  - Aquods
  - Ferrods
  - Humods
  - Orthods
- Ultisols
  - Aquults
  - Humults
  - Udults
  - Ustults
  - Xerults
- Vertisols
  - Torrerts
  - Uderts
  - Usterts
  - Xererts

The brief individual descriptions of the suborders are arranged in the same sequence as the list.

Broad regional distribution of soils for the world has been recognized (**Fig. 4**). Each region outlined on the map is identified by the name of the most extensive suborder among the component soils. In every region, suborders other than the most extensive one are important.

Alfisols. These soils have A and E horizons that are mostly pale in color and that have lost silicate clay, sesquioxides, and bases such as calcium and magnesium.

The soils have B horizons with accumulations of silicate clay and with moderate to high levels of exchangeable calcium and magnesium. The C horizons are usually lighter in color and lower in clay than the B horizons.

Alfisols are most extensive in humid, temperate regions but range from the edges of the tundra and the desert into the tropics. Mostly, the soils were under forest or savanna vegetation, though some were under prairie. All have been formed on land surfaces that are neither old nor yet among the youngest in the world.

Occurring as they do in many parts of the world, these soils are used for a wide variety of crops. Some remain in forest, and those under drier climates are used chiefly for grazing.

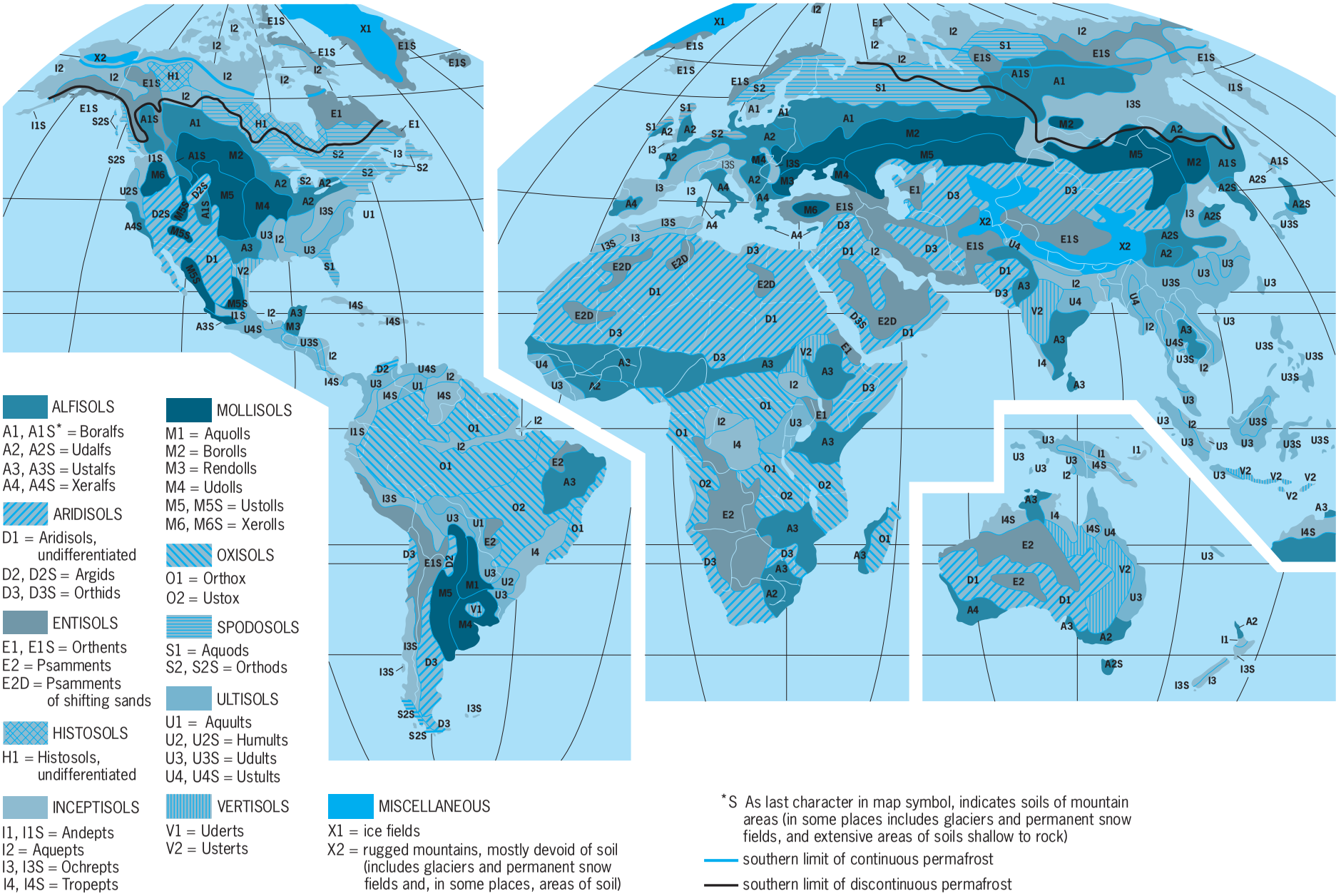


Fig. 4. General soil map of the world. Each region is identified by the name of the most extensive suborder. Other suborders are present in every region and are important in most of them. (*Soil Conservation Service, USDA*)

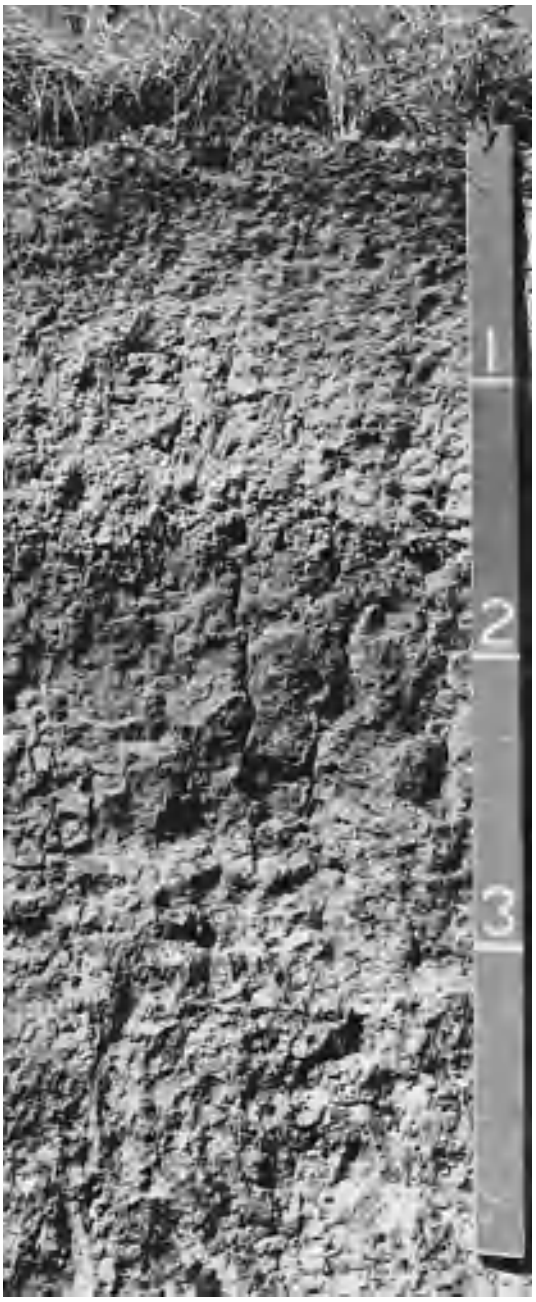


Fig. 5. Profile of an Aqualf with pale A and E horizons about 18 in. (45 cm) thick resting on B horizon high in clay which grades into C horizon at a depth of about 4 ft (1.2 m); numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

Aqualfs. These are the seasonally wet Alfisols. They generally occur in depressions or on rather wide flats in local landscapes. In addition to the general morphology and composition shared with other soils of the order, Aqualfs are marked by gray or mottled colors reflecting their wetness (**Fig. 5**).

Boralfs. These are the well-drained Alfisols of cool or cold regions, such as west-central Canada and Russia. The soils occur either at high altitudes or in high latitudes, including some frigid zones.

In their morphology and composition, the soils are much like the Udalfs, though colors are more dull on the whole and the surplus of calcium and magnesium a little higher.

Udalfs. These are the well-drained Alfisols of humid, temperate climates. The soils are important in the north-central part of the United States, in western Europe, and in eastern Asia. Udalfs differ from Aqualfs in that B horizons are characteristically brown or yellowish brown and lack marks of wetness. These have higher mean annual temperatures than do the Boralfs and are moist for higher proportions of the year than the Ustalfs and Xeralfs (**Fig. 6**).

Ustalfs. These are well-drained Alfisols occurring in somewhat drier and mostly warmer regions than Udalfs. On the whole, the soils have more reddish B horizons and are a little higher in calcium and magnesium than Udalfs. These soils are intermittently dry during the growing season.

Xeralfs. These are well-drained Alfisols found in regions with rainy winters and dry summers, in what are called mediterranean climates. Like the Ustalfs in nature of B horizons, the soils have A horizons that tend to become massive and hard during the dry season. Some of the soils have duripans [cemented

Fig. 6. Profile of a Udalf with A and E horizons 12 in. (30 cm) thick over darker B horizon with blocky structure grading into C horizon at a depth of about 4 ft (1.2 m); larger numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

layers at depth of 2 or 3 ft (0.6 or 0.9 m)] that interfere with root growth.

Aridisols. These are major soils of the world's deserts, which form about one-fourth of the land surface. Soils of other orders, especially the Entisols, are also present but less extensive in the deserts.

Formed under low rainfall, Aridisols have been leached little and are therefore high in calcium, magnesium, and other more soluble elements. The low rainfall has also limited growth of plants, mostly shrubs and similar species, so that the soils are low in organic matter and nitrogen. The combined A and B horizons are rarely more and usually less than 1 ft (30 cm) thick. The A horizons are light-colored and usually calcareous. All horizons are neutral or mildly alkaline in reaction.

Most Aridisols in use provide some grazing for nomadic herds. On the other hand, if water and other resources, including adequate skills, are available and climate is favorable, some Aridisols will support a large variety and produce high yields of crops.

Argids. These well-drained Aridisols have B horizons of silicate clay accumulation. The B horizons are characteristically brown or reddish in color and grade into lighter colored C horizons marked by carbonate accumulation. On the whole, these soils occupy the older land surfaces in desert regions (**Fig. 7**).

Orthids. These Aridisols lack B horizons of clay accumulation. Many are free of carbonates in the A and upper B horizons; most are well drained. Common colors are gray or brownish gray with little change from top to bottom of the profile. A few Orthids are fairly high in soluble salts such as sodium sulfate and sodium chloride, whereas others are high in calcium carbonate throughout. More extensive than the Argids, generally, Orthids occupy younger but not the youngest land surfaces in deserts.

Entisols. These soils have few and faint horizons, with reasons for the limited horizonation differing among suborders. Reasons for the practical absence of horizons are indicated for individual suborders.

Entisols occur in all parts of the world and may be found under a wide variety of vegetation. Most, though not all, are on young land surfaces, distributed from the tundra through the tropics and from the deserts to the rainiest climates. Entisols have a wide range in usefulness. Some are highly productive and others are not.

Aquents. These Entisols have been under water until very recent times at the margins of oceans, lakes, or seas. The wetness is reflected in the bluish-gray or greenish-gray colors. Examples are the soils in recently reclaimed polders of the Netherlands. The total extent of Aquents in the world is very small.

Arents. These are Entisols because of severe disturbance of soils formerly classifiable in other orders. The sequence of horizons has been disrupted completely, and remnants of those horizons can be found randomly distributed in the profiles of Arents.

Fluvents. These well-drained Entisols are in recently deposited alluvium. They occur along streams or in fans where the rate of sediment deposition is high. Marks of sedimentation are still evident, and identifiable horizons are lacking, except for slightly darkened surface layers or A horizons. Small bodies of these soils are scattered over all parts of the world.

Fig. 7. Profile of an Argid with pale silty A and E horizons, darker B horizon higher in clay, and calcareous C horizon; profile shown in 20 in. (50 cm) deep; numbers on scale indicate inches. 1 in. = 2.5 cm. (*Photograph by R. W. Simonson*)

Orthents. These well-drained Entisols are of medium or fine texture, mostly on strong slopes. The grass strips serve to remove sediment from overland flow. The soils may have A horizons or slightly darkened surface layers an inch or so thick but otherwise lack evidence of horizonation. Many of the soils are shallow to bedrock.

Psamments. These Entisols are of sandy texture. Like the Orthents, these may have thin A horizons, which grade into thick C horizons. The sandiness is the distinctive character of the suborder (**Fig. 8**).

Histosols. These are wet soils consisting mostly of organic matter, popularly known as peats and mucks. Most have restricted drainage and are saturated with

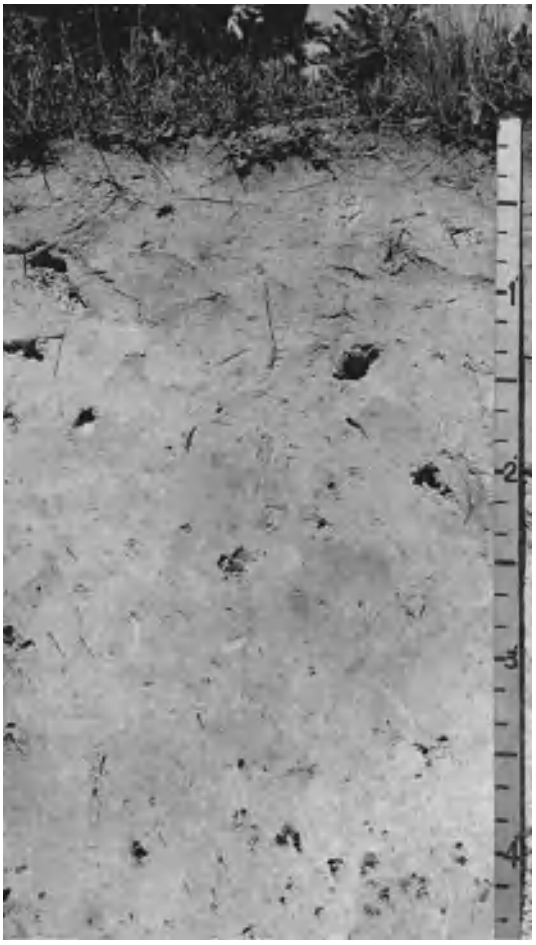


Fig. 8. Profile of a Psamment lacking evident horizons and consisting of sand throughout; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

water much of the time. A few are wet but not fully saturated. Widely distributed over the world, these soils may occur in small or large bodies, with the latter occurring chiefly at high latitudes. A large proportion of the total area is idle. Where the climate is favorable, some of the soils have been drained and are producing vegetables and other crops. *See* BOG; PEAT.

Fibrists. These Histosols consist mainly of recognizable plant residues or sphagnum moss. They are saturated with water most of the year unless drained (**Fig. 9**).

Folists. These Histosols consist of forest litter resting on rock or rubble. Drainage is not restricted, but a combination of rainfall, fog, and low temperatures keeps the litter wet.

Hemists. These Histosols consist of partially decayed plant residues. Plant structures have largely been destroyed but an appreciable share of the mass remains as fibers when rubbed vigorously. The soils are saturated with water much of the time unless drained.

Saprists. These Histosols consist of residues in which plant structures have been largely obliterated by decay. A very small part of the mass remains as fibers after vigorous rubbing. The soils are saturated with water much of the time unless drained. Most Saprists in the United States are known as muck.

Inceptisols. These soils have faint to moderate horizonation but lack horizons of accumulation of translocated substances other than carbonates and silica. Two of the suborders have distinct dark A horizons, and most have B horizons formed by losses and transformations without corresponding gains in substances. Thus, the Inceptisols are in some ways intermediate in horizonation between the Entisols and Vertisols on the one hand, and the Alfisols, Mollisols, Spodosols, and Ultisols on the other.

Inceptisols are widely distributed, ranging from the arctic through the tropics and from the margins of the desert into regions of heavy rainfall. They may consequently be found under a wide variety of vegetation. Usefulness of the soils has as wide a range as does their distribution. Some are highly productive and others are of little or no value.

Andepts. These Inceptisols are formed chiefly in volcanic ash or in regoliths with high components of ash. Mostly, the soils tend to be fluffy. They have thick dark A horizons, rather high levels of acidity, and poorly crystalline clay minerals. The soils are

Fig. 9. Profile of a Fibrist with little or no change from the surface to a depth of 5 ft (1.5 m); soil consists of partly decayed plant residues; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

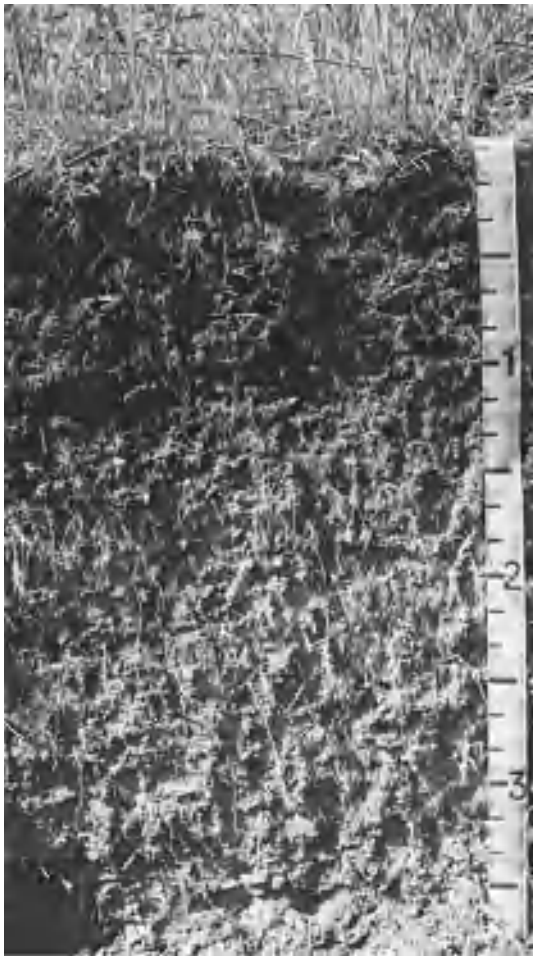


Fig. 10. Profile of an Andept with thick, dark A horizon, faint B horizon, and lighter C horizon; fine plant roots are numerous; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

widely distributed but seem to be restricted to regions of fairly recent volcanic activity (**Fig. 10**).

Aquepts. These Inceptisols are wet or have been drained. Like the Aqualfs, the soils have gray or mottled B and C horizons, but they lack silicate clay accumulation in their profiles. The A horizons may be dark and fairly thick or they may be thin, as they are in many of the soils (**Fig. 11**).

Ochrepts. These Inceptisols have pale A horizons, darker B horizons, and lighter colored C horizons. The B horizons lack accumulations of translocated clay, sesquioxides, or humus. The soils are widely distributed, occurring from the margins of the tundra region through the temperate zone but not in the tropics. Ochrepts also occur in the fairly dry regions though not in deserts (**Fig. 12**).

Plaggepts. These Inceptisols have very thick surface horizons of mixed mineral and organic materials added as manure or as human wastes over long periods of time. For the world as a whole, such soils are of negligible extent, but they are conspicuous where found.

Tropepts. These Inceptisols have moderately dark A horizons with modest additions of organic matter, B horizons with brown or reddish colors, and slightly paler C horizons. The soils are less strongly weathered than the geographically associated Ultisols and Oxisols. In general appearance, the profiles are much like that of the Orthox (**Fig. 13**). Tropepts are restricted to tropical regions, largely to those of moderate and high rainfall.

Umbrepts. These Inceptisols have dark A horizons more than 10 in. (25 cm) thick, brown B horizons, and slightly paler C horizons. The soils are strongly acid, and the silicate clay minerals are crystalline rather than amorphous as in the Andepts. The Umbrepts occur under cool or temperate climates, are widely distributed, and are of modest extent.

Mollisols. These soils have dark or very dark, friable, thick A horizons high in humus and bases such as calcium and magnesium. Most have lighter colored or browner B horizons that are less friable and about as thick as the A horizons. All but a few have paler C horizons, many of which are calcareous.

Major areas of Mollisols occur in subhumid or semiarid cool and temperate regions. They meet the desert along their drier margins and meet soils such

Fig. 11. Profile of an Aquept with thin, dark A horizons, fairly thick, light-gray B horizon, and stone in C horizon beside tape; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

Fig. 12. Profile of an Ochrept with litter on the surface, dark A horizon 4 in. (10 cm) thick, thin B horizon, and pale C horizon; deeper profile is marked by plant roots and traces of former roots; numbers on scale indicate inches. 1 in. = 2.5 cm. (*Photograph by R. W. Simonson*)

as the Alfisols at their more humid margins. Mollisols were formed under vegetation consisting chiefly of grasses and are thus the major ones of former prairies and steppes. The soils occupy rather young land surfaces.

Though they produce a variety of crops, Mollisols are largely used for cereals. These soils now produce a major share of the world's output of corn and wheat. Topography is generally favorable for the operation of large machinery, and many Mollisols are therefore in large farms. Yields have a wide range, depending on climatic conditions. Wide fluctuations in yield with wet and dry years are normal for the Mollisols marginal to arid regions. On the other hand, yields are consistently high for those under more humid climates.

Albolls. These are Mollisols with dark A horizons, pale E horizons, distinct B horizons marked by clay accumulation, and paler C horizons. The soils are set, especially in the upper part, for some part of the year. Mostly, the soils occur on upland flats and in shallow depressions.

Aquolls. These are wet Mollisols unless they have been drained. Because they were formed under wet conditions, the soils have thick or very thick, nearly black A horizons over gray or mottled B and C horizons. If they have not been drained, the soils may be under water for part of the year, but they are seasonally rather than continually wet (**Fig. 14**).

Borolls. These are Mollisols of cool and cold regions. Most areas are in moderately high latitudes or at high altitudes. The soils have fairly thick, nearly black A horizons, dark grayish-brown B horizons, and paler C horizons that are commonly calcareous. The B horizons of some soils have accumulations of clay. These soils are extensive in western Canada and Russia.

Rendolls. These are the Mollisols formed in highly calcareous parent materials, regoliths with more than 40% calcium carbonate. The soils may be calcareous to the surface and must have high levels of carbonates within a depth of 20 in. (50 cm). Rendolls do not have horizons of carbonate accumulation. The profiles consist of dark or very dark A horizons grading into pale C horizons. For the most part, Rendolls are restricted to humid, temperate regions.

Udolls. These are Mollisols of humid, temperate and warm regions where maximum rainfall comes during the growing season. The soils have thick, very dark A horizons, brown B horizons, and paler C horizons.

Fig. 13. Profile of an Orthox with slightly darkened A horizon about 1 ft (30 cm) thick, little further change with depth, and deep penetration by fine plant roots; scale in feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

Fig. 14. Profile of an Aquoll with very thick, dark A horizon, signs of mixing and burrowing by animals at depths between 3 and 3.5 ft (0.9 and 1.1 m), and lighter C horizon at the bottom; the numbers on the scale to the left indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

Throughout the profile these soils are browner than the Borolls and are not as cold. Udolls lack horizons of accumulation of powdery carbonates. Some of the soils have B horizons of clay accumulation and others do not. These soils are major ones of the Corn Belt of the United States (**Fig. 15**).

Ustolls. These are the Mollisols of temperate and warm climates with lower rainfall than the Udolls. The soils are therefore dry for an appreciable part of each year, usually more than 90 cumulative days. Horizons and their sequence are much the same as for Udolls except that many Ustolls have accumulations of powdery carbonates at depths of 40 in. (100 cm) or less.

Xerolls. These are Mollisols of regions with rainy winters and dry summers. The nature and sequence of horizons are much like those of the Ustolls. The soils are completely dry for a long period during the summer of each year.

Oxisols. These soils have faint horizonation, though formed in strongly weathered regoliths. The surface layers or A horizons are usually darkened and moderately thick, but there is little evidence of change in the remainder of the profile. Because of the intense or long weathering, the soils consist of resistant minerals such as kaolinite, forms of sesquioxides, and quartz. Weatherable minerals such as feldspars have largely disappeared. Moreover, the clay fraction has limited capacity to retain bases such as calcium and magnesium.

The soils are porous and readily penetrated by water and plant roots. A distinctive feature of Oxisols is the common occurrence of tubular pores about the diameter of ordinary pins extending to depths of 6 ft (1.8 m) or more.

Oxisols are largely restricted to low altitudes in humid portions of the tropics. Any occurring elsewhere seem to be relicts of earlier geologic ages. All occupy old land surfaces. Most were formed under forest, with some under savanna vegetation. Regions with Oxisols as major soils are extensive, ranking second in total area only to the Aridisols of deserts.

Most Oxisols remain in forest or savanna and produce little food and fiber. Many regions are sparsely inhabited, with natives depending on shifting cultivation for much of their food. A small proportion of the total area is cultivated with modern technology and is highly productive. Even so, management to

Fig. 15. Profile of a Udoll with thick, dark A horizon, B horizon gradational in color, and rather pale C horizon; filled former animal burrows in B horizon; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

ensure sustained high yields is still to be developed for the more strongly weathered Oxisols.

Aquox. These are seasonally wet Oxisols found chiefly in shallow depressions. Because of their wetness, deeper profiles are dominantly gray, with or without mottles and nodules or sheets of iron and aluminum oxides. Total extent is extremely small.

Humox. These are well-drained Oxisols high in organic matter and moist all or nearly all year. Profiles have dark A horizons 1 ft (0.3 m) or so thick over generally reddish B and C horizons. The high amounts of organic matter distinguish Humox from other suborders of Oxisols, and the soils are also moist much more of the year than are Torrox and Ustox. Humox are believed to be of limited extent, restricted to relatively cool climates and high altitudes for Oxisols.

Orthox. These are well-drained Oxisols moderate to low in organic matter and moist all or nearly all year. Orthox are much like Humox in general appearance, but their profiles are lower in organic matter. They are moist more of each year than are Torrox and Ustox. Although good data on extent are lacking, Orthox are believed to be extensive at low altitudes in the heart of the humid tropics (Fig. 13).

Torrox. These are well-drained Oxisols low in organic matter and dry most of the year. Profiles resemble those of Orthox except that A horizons are more poorly expressed. The soils are believed to have been formed under more rainy climates of past eras. Total extent of the Torrox seems to be extremely small.

Ustox. These are well-drained Oxisols low to moderate in organic matter and dry for periods of at least 90 cumulative days each year. Profiles resemble those of Orthox, on the whole. Ustox are lower in organic matter than Humox, dry for longer periods each year than Humox and Orthox, and moist for longer periods than Torrox. Good data on extent are lacking, but Ustox are believed to be extensive.

Spodosols. These soils have B horizons with accumulations of one or both of organic matter and compounds of aluminum and iron. The accumulated substances are amorphous in nature. They impart red, brown, or black colors to the B horizons, which may have irregular lower boundaries with tongues extending downward a foot or more. If the soil has not been disturbed, the surface layer consists of both fresh and partly decayed litter. This rests on a very pale, leached E horizon overlying a highly contrasting B horizon. The Spodosols formed from sands under boreal coniferous forests have some of the most striking profiles in the world. Mostly, the soils are strongly acid because of the small supplies of bases.

Spodosols are most extensive in humid, cool climates, but some occur at low elevations under tropical and subtropical climates. The soils were largely under forest. The bulk of the Spodosols have been formed in sandy regoliths, with others in loamy regoliths. Land surfaces are fairly young.

Most Spodosols remain in forest, but some are cultivated in both cool and tropical regions. The variety of crops produced is large because of the wide climatic range under which the soils occur. The range in yields is also wide, being dependent on the combination of climatic conditions and prevailing level of technology. Production is modest for most Spodosols under cool climates and simple management. Production is high from some soils cultivated with complex management in tropical climates.

Aquods. These are seasonally wet Spodosols. The soils may be wet most of each year but not all of the time. The B horizons are black or dark brown in color and some are cemented. Aquods occupy depressional areas or wide flats from which water cannot escape easily (**Fig. 16**).

Ferrods. These are well-drained Spodosols having B horizons of iron accumulation with little organic matter. Appearance of the profile is much like that of Orthods.

Humods. These are well-drained Spodosols having B horizons of humus accumulation, usually black or dark brown in color. Aluminum usually accumulates with the humus but iron is lacking, especially from the upper part of B horizons. Where formed in white or nearly white sands, the soils have striking profiles, as in parts of western Europe.

Orthods. These are well-drained Spodosols having B horizons of humus, aluminum, and iron accumulation. The B horizons are mostly red or reddish in color and are friable. They grade downward into

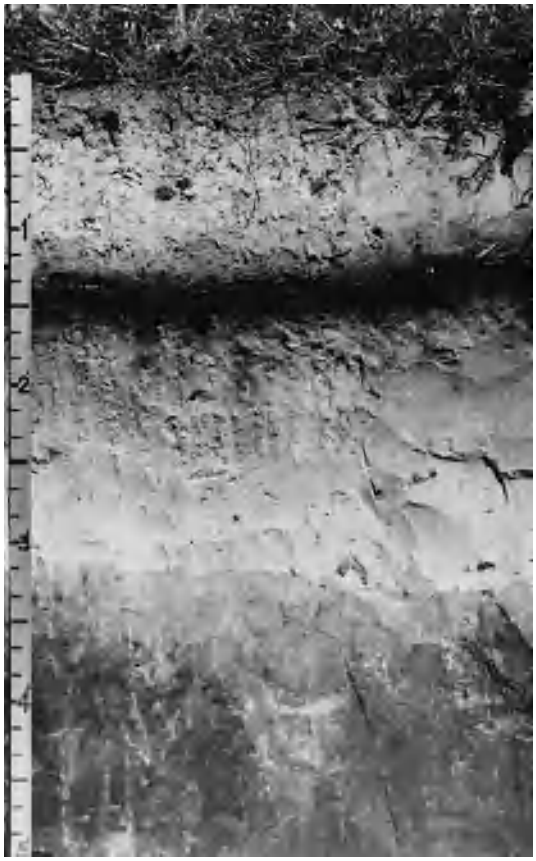


Fig. 16. Profile of an Aquod with distinct E horizon, dark B horizon at depth of 18 in. (45 cm), pale C horizon below, and part of buried profile below 4 ft (1.2 m); soil consists of sand; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

lighter-colored C horizons which are commonly less friable and may be very firm. Orthods form the most extensive suborder among the Spodosols, being widespread in Canada and Russia.

Ultisols. Like Alfisols, these soils have A and E horizons that have lost silicate clays, sesquioxides, and bases. Most A and E horizons are pale, though not all are. The B horizons have accumulations of silicate clays and low levels of exchangeable calcium and magnesium. The C horizons are usually lighter in color and lower in clay than the B horizons. Combined thickness of the A, E, and B horizons is greater, on the average, for Ultisols than for Alfisols. Ultisols are strongly acid throughout their profiles, reflecting the low levels of exchangeable bases.

Ultisols are most extensive under humid, warm-temperate climates but extend through the tropics. They are not found in cold regions. The largest bodies of the soils are in southeastern Asia, nearby islands, and the southeastern United States. The soils were usually under forest but some were covered by savanna vegetation. All were formed in strongly weathered regoliths on old land surfaces.

Many Ultisols remain in forest. Among those producing crops, a majority are used under some method of shifting cultivation. Production is limited in such circumstances. On the other hand, the variety of crops and yields obtained can be large if cultivators are in a position to apply complex technology to Ultisols.

Aquults. These are seasonally wet Ultisols, saturated with water for an important part of the year unless drained. Usually the soils have thin, dark A horizons, but they may be as thick as 20 in. (50 cm). Deeper profiles are gray, with or without red mottles. Aquults occur in depressions or on wide upland flats from which water moves very slowly.

Humults. These are well-drained Ultisols formed under rather high rainfall distributed evenly over the year. The soils are high in organic matter throughout their profiles, and most have darkened A horizons of moderate thickness. Deeper profiles tend to be brown, reddish-brown, or yellowish-brown in color. Humults are common in southeastern Brazil.

Udults. These are well-drained Ultisols of humid, warm-temperate and tropical regions. The soils are low or relatively low in humus and typically have thin, darkened A horizons. The B horizons are yellowish-red, red, brown, or yellowish-brown in color and are fairly thick. Rainfall is high enough and distributed evenly enough over the year so that soils are dry for only short periods. Udults are major soils in the southeastern parts of the United States and Asia (**Fig. 17**), and their total extent is large.

Ustults. These are well-drained Ultisols of warm-temperate and tropical climates with moderate or low rainfall. The soils are like the Udults in general appearance but are dry for appreciable periods each year. Examples of Ustults may be found in northeastern Australia. Total extent of the suborder is appreciably less than that of Udults.

Xerults. These are well-drained Ultisols of regions with warm, dry summers and cool, rainy winters. The soils are like Ustults in appearance but become and remain dry for longer periods in summers. Total extent is small.

Fig. 17. Profile of a Udult with pale A horizon 16 in. (40 cm) thick, darker B horizon higher in clay and iron oxides, and C horizon near bottom; numbers on scale indicate feet. 1 ft = 30 cm. (*Photograph by R. W. Simonson*)

Vertisols. These soils have faint horizonation for two main reasons. In the first place, Vertisols are formed in regoliths that are high in clay and therefore resistant to change. In the second place, the clay fraction in the soils has high levels of activity. The soils are therefore subject to marked swelling and shrinking as they wet and dry. Cracks formed as the soils become dry may extend to depths of several feet. Because of the shrinking and swelling of the soils, materials from deeper profiles are forced upward in places so that entire soils are slowly but continually overturned and mixed. Vertisols have therefore been called "self-swallowing" and "soils that plow themselves."

Some Vertisols have darkened A horizons, whereas others do not. All are low in organic matter and high in bases. Many are calcareous in deeper profiles. Most are neutral or mildly alkaline in reaction because of goodly supplies of bases.

Vertisols occur in warm-temperate and tropical climates with one or more dry seasons. The soils were under savanna vegetation for the most part with a

few in forest. Land surfaces are old or fairly old. Large bodies of Vertisols are found on the Deccan Plateau of India, in the Gezira of Sudan, and in Australia.

Because they are high in active clays, Vertisols are hard to cultivate. The soils therefore remain in savanna in many places, and the savannas are used for grazing or left alone. Large areas of the soils are cultivated, some with simple, bullock-drawn implements and others with large machinery. A wide variety of crops are produced, but yields are generally modest, especially for cultivators dependent on simple technology. The soils will, however, produce crops indefinitely under simple management.

Torrerts. These are Vertisols of arid regions, the driest soils of the order. Because the soils are dry most of the time, cracks that form tend to remain open. The soils do become wet enough at rare intervals to permit cracks to close. Torrerts are of limited extent.

Uderts. These are Vertisols of humid regions; each profile is moist in some part most of each year. The soils do dry out enough to permit formation of cracks once every year, as a rule. Uderts are moist more of each year than other Vertisols. The soils are of moderate extent.

Usterts. These are Vertisols of subhumid and semiarid regions, chiefly under climates with two rainy and two dry seasons each year. Cracks formed during dry seasons are open for at least 90 cumulative days per year. The Usterts are thus intermediate in moisture regimes between Uderts on the wet side and Torrets on the dry side. Usterts are dry for shorter periods during summer than Xererts. Usterts are extensive, represented by large bodies in Australia and India (**Fig. 18**).

Fig. 18. Profile of a Ustert with thick, dark A horizon, which is due in part to mixing and churning of soil mass and which grades into the lighter-colored C horizon; scale in feet. 1 ft = 30 cm. (*Photograph by E. H. Templin*)

Xererts. These are Vertisols of regions with warm, dry summers and cool, rainy winters—the mediterranean climates. Cracks formed as the soils dry out each summer remain open for at least 60 consecutive days. Xererts are higher in moisture than Torrerts and lower than Uderts. They have longer dry periods during warm seasons than Usterts. The soils are of limited extent.

Roy W. Simonson

Zonality

Many soils that are geographically associated on plains have common properties that are the result of formation in similar climates with similar vegetation. Because climate determines the natural vegetation to a large extent and because climate changes gradually with distance on plains, there are vast zones of uplands on which most soils have many common properties. This was first observed in Russia toward the end of the nineteenth century by V. V. Dokuchaev, the father of modern soil science. He also observed that on floodplains and steep slopes and in wet places the soils commonly lacked some or most of the properties of the upland soils. In mountainous areas, climate and vegetation tend to vary with altitude, and here the Russian students observed that many soils at the same altitude had many common properties. This they called vertical zonality in contrast with the lateral zonality of the soils of plains.

Zonal classification. These observations led N. M. Sivirtsev to propose in about 1900 that major kinds of soil could be classified as Zonal if their properties reflected the influence of climate and vegetation, as Azonal if they lacked well-defined horizons, and as Intrazonal if their properties resulted from some local factor such as a shallow ground water or unusual parent material.

This concept was not accepted for long in Russia. It was adopted in the United States in 1938 as a basis for classifying soil but was dropped in 1965. This was because the Zonal soils as a class could not be defined in terms of their properties and because they had no common properties that were not shared by some Intrazonal and Azonal soils. It was also learned that many of the properties that had been thought to reflect climate were actually the result of differences in age of the soils and of past climates that differed greatly from those of the present.

Zonality of soil distribution is important to students of geography in understanding differences in farming, grazing, and forestry practices in different parts of the world. To a very large extent, zonality is reflected but is not used directly in the soil classification used in the United States. The Entisols include most soils formerly called Azonal. Most of the soils formerly called Intrazonal are included in the orders of Vertisols, Inceptisols, and Histosols and in the aquic suborders such as Aquolls and Aqualfs. Zonal soils are mainly included in the other suborders in this classification.

The soil orders and suborders have been defined largely in terms of the common properties that result from soil formation in similar climates with

similar vegetation. Because these properties are important to the native vegetation, they have continuing importance to farming, ranching, and forestry. Also, because the properties are common to most of the soils of a given area, it is possible to make small-scale maps that show the distribution of soil orders and suborders with high accuracy.

Zonal properties. A few examples of zonal properties of soils and their relation to soil use follow. The Mollisols, formerly called Chernozemic soils, are rich in plant nutrients. Their natural ability to supply plant nutrients is the highest of any group of soils, but lack of moisture often limits plant growth. Among the Mollisols, the Udolls are associated with a humid climate and are used largely for corn (maize) and soybean production. Borolls have a cool climate and are used for spring wheat, flax, and other early maturing crops. Ustolls have a dry, warm climate and are used largely for winter wheat and sorghum without irrigation. Yields are erratic on these soils. They are moderately high in moist years, but crop failures are common in dry years. The drier Ustolls are used largely for grazing. Xerolls have a rainless summer, and crops must mature on moisture stored in the cool seasons. Xerolls are used largely for wheat and produce consistent yields.

The Alfisols, formerly a part of the Podzolic soil group, are lower in plant nutrients than Mollisols, particularly nitrogen and calcium, but supported a permanent agriculture before the development of fertilizers. With the use of modern fertilizers, yields of crops are comparable to those obtained on Mollisols. The Udalfs are largely in intensive cultivation and produce high yields of a wide variety of crops. Boralfs, like Borolls, have short growing seasons but have humid climates. They are used largely for small grains or forestry. Ustalfs are warm and dry for long periods. In the United States they are used for grazing, small grains, and irrigated crops. On other continents they are mostly intensively cultivated during the rainy season. Population density on Ustalfs in Africa is very high except in the areas of the tsetse fly. Xeralfs are used largely for wheat production or grazing because of their dry summers.

Ultisols, formerly called Latosolic soils, are warm, intensely leached, and very low in supplies of plant nutrients. Before the use of fertilizers, Ultisols could be farmed for only a few years after clearing and then had to revert to forest for a much longer period to permit the trees to concentrate plant nutrients at the surface in the leaf litter. With the use of fertilizers, Udults produce high yields of cotton, tobacco, maize, and forage. Ustults are dry for long periods but have good moisture supplies during a rainy season, typically during monsoon rains. Forests are deciduous, and cultivation is mostly shifting unless fertilizers are available.

Aridisols, formerly called Desertic soils, are high in some plant nutrients, particularly calcium and potassium, but are too dry to cultivate without irrigation. They are used for grazing to some extent, but large areas are idle. Under irrigation some Aridisols are highly productive, but large areas are unsuited to irrigation or lack sources of water. Guy D. Smith

Physical Properties

Physical properties of soil have critical importance to growth of plants and to the stability of cultural structures such as roads and buildings. Such properties commonly are considered to be (1) size and size distribution of primary particles and of secondary particles, or aggregates; (2) the consequent size, distribution, quantity, and continuity of pores; (3) the relative stability of the soil matrix against disruptive forces, both natural and cultural; (4) color and textural properties, which affect absorption and radiation of energy; and (5) the conductivity of the soil for water, gases, and heat. These usually would be considered as fixed properties of the soil matrix, but actually some are not fixed because of influence of water content. The additional property, water content—and its inverse, gas content—ordinarily is transient and is not thought of as a property in the same way as the others. However, water is an important constituent, despite its transient nature, and the degree to which it occupies the pore space generally dominates the dynamic properties of soil. Additionally, the properties listed above suggest a macroscopic homogeneity for soil which it may not necessarily have. In a broad sense, a soil may consist of layers or horizons, each consisting of roughly uniform soil materials of various types, which considered together may affect both mechanical properties and movement of water in a soil profile.

From a physical point of view it is primarily the dynamic properties of soil which affect plant growth and the strength of soil beneath roads and buildings. While these depend upon the chemical and mineralogical properties of particles, particle coatings, and other factors discussed above, water content usually is the dominant factor. Water content depends upon flow and retention properties, so that the relationship between water content and retentive forces associated with the matrix becomes a key physical property of a soil.

Structure and pore space. Although the mineralogical nature of the clay fraction and the textural classification of the soil are of basic importance to soil physical properties, other factors often have even greater importance. Decomposed organic matter, or humus, and certain inorganic compounds, such as iron and aluminum oxides, form coatings on soil particles, and these, along with some clays, bind soil particles together into granules or aggregates. The size and distribution of such aggregates are major factors in determining pore size and distribution in soil and are important in the physical behavior of soils. Additionally, fragmented organic materials, undecomposed or only partially so, if present in soil in significant quantity, also can play a major role in determining physical properties. The manner in which mineral soil particles are assembled and maintained in aggregate form, together with quantity, size, and distribution of fragments of partially decomposed

organic materials, is referred to as soil structure. Soil structure may develop as a consequence of natural processes such as wetting and drying, freezing and thawing, transport of minerals in moving water and their deposition, or mechanical forces exerted by plant roots or by other biological agents. Natural structure also may be modified by plowing, cultivation, and mechanical forces associated with cultural activities.

There are two aspects to the evaluation of soil structure which have major practical importance: measurements of pore size and configuration, and assessments of the stability of these characteristics against various natural and cultural forces which act to change the physical arrangement of primary particles and secondary aggregates. Pore size and configuration determine water retention and water and air transport properties of a soil and, in some cases, the ease with which plant roots penetrate and living organisms can move. The equivalent pore size of the largest water-filled pore in a soil sample (that is, the radius of a cylindrical pore which would behave with respect to water retention similarly to the largest water-filled soil pore) may be measured by using the hanging water column apparatus (**Fig. 19**), and the equation for capillary rise, $r = 2s/Dgh$, where s is the surface tension of water, D its density, g the acceleration of gravity, and h the height of rise of water in a capillary tube of radius r.

The porous plate apparatus for relating water-filled pore size to the force of removal exerted by the hanging water column (Fig. 19*a*) is based upon a capillary tube model where soil pores are presumed to be equivalent to small sections of capillary tubes (Fig. 19*b*). Pores in the porous plate are all considerably smaller than the size of the capillary tube (Fig. 19*a*), so that they remain water-filled as h is varied. More typically, the shape of air-water interfaces (Fig. 19*c*) has a radius of curvature, R, given as a combination of a positive radius r_1 measured axially and a negative radius r_2 at right angles, or $1/R = (1/r_1) + (1/r_2)$. A plot of water content of a soil sample in the hanging column apparatus against the height of the hanging water column characterizes the pore size distribution of the sample. A large change in water content with a small change in hanging water column length indicates a large volume of pores of similar size, whereas a small change in water content with a large change in hanging water column length indicates a relatively small volume associated with pores over a wide size range. More direct evidence of pore size and arrangement may be obtained by microscopic examination.

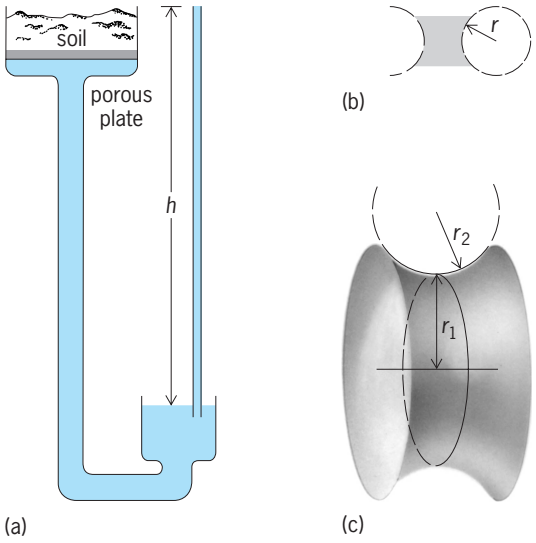


Fig. 19. Measurement of equivalent pore size. (*a*) Porous plate apparatus. (*b*) Schematic representations of the air-water interfaces and (*c*) the particle-water interfaces.

The second factor which is basic to determining the structural state of soil has to do with soil stability against disruptive forces. The existence of binding forces between clay particles themselves and between clay and other mineral surfaces, together with colloidal organic decomposition products, polysaccharides and polyuronides, and inorganic cementing substances, determines how well a particular geometrical organization of soil particles resists change against disruptive forces. Hence, structural evaluation usually involves both geometrical properties and some indication of how stable a particular arrangement of particles will be against the disruptive forces of a plow, falling raindrops, the weight of an animal, or some other force. Although the nature of the pore space, by implication at least, is a part of any characterization of soil structure, often the measurement involves application of a particular disruptive force and noting the degree to which an existing arrangement is destroyed.

Bulk density. An important property which can be used to characterize the structural state of a soil is the bulk density. Where the particle density of soil materials is uniform (often taken as 2.65 g/cm^3), it is possible to determine the total pore space fraction of a soil from measurement of the bulk density (mass of dry soil per unit bulk volume) and use of the formula, pore space fraction $= 1 - (D_b/D_p)$, where D_b and D_p are the bulk and particle densities. The presence of organic materials having particle densities differing appreciably from those of the mineral particles introduces some error into such computations. Agricultural soils in the plow depth ordinarily have bulk densities ranging from 1.0 to 1.5 g/cm^3 and a pore-space fraction from 0.4 to 0.6, or roughly half particles and half pore space.

Soil horizonation. From some perspectives the physical properties of soil may be presented adequately by a description of the properties of a sample of surface soil. However, horizonation, whether produced by natural or cultural processes, may profoundly affect water flow and retention near the surface or within the rooting zone of plants. Such influence is particularly noticeable at boundaries between materials having different pore sizes.

Horizonation may be the result of the method of original deposition, for example, alluvial or loess derived profiles, or it may develop over time as a consequence of weathering processes which differ near

the soil surface from those at depths beneath. Root growth and development and organic matter decomposition near the surface may promote both the development of aggregates and their stabilization. Also, fine soil particles, silts and clays, may be displaced downward through a profile by moving water, accumulating in layers below. Solutes carried in water also may dissolve minerals, carrying them in the direction of water flow and depositing them as precipitates or leaving them behind as water evaporates.

Horizonation also may occur in soil as a consequence of plowing and cultivation operations. It is not unusual for a plow pan to develop in cultivated soils (**Fig. 20**) as a consequence of the smearing action of an implement surface as it slides through the soil. Decreased porosity at an interface between disturbed and undisturbed soil can profoundly affect water flow and bring about both filtering of fine particles, which may be carried in moving water, and deposition of substances by precipitation. Also, soil may become severely compacted in the surface by vehicle and animal traffic, with a consequent decrease in porosity and increase in soil hardness.

Forces acting on soil water. Flow of water in soil depends upon forces existing within the soil matrix and upon forces associated with gravity acting upon the water itself. In saturated soil the force of gravity acting upon water within the soil and by standing water upon the soil surface constitutes the moving force, and it can cause flow in any direction depending upon the geometry of the system. In unsaturated soil, gravity still acts upon the water but, except in very wet soil, this force may be small compared to forces due to the attraction of solid surfaces for water. These are the same forces that cause water to rise in a capillary tube, and they are sometimes referred to as capillary forces. In soil, such forces are called matric forces. In addition to matric forces, three other forces may be present in a soil system: osmotic forces, relating to the attraction between solutes and water; local gas pressures, which may exert a force upon air-water interfaces within soil pores; and overburden forces, which may arise where substances dissolved or suspended in water add to its density and increase its weight. Overburden forces may exist also in saturated soil, but usually such forces are neglected under both saturated and unsaturated conditions.

The nature of the various forces acting on water in soil differ appreciably, and the direction in which each acts may be highly complicated. Hence, it is difficult to apply them quantitatively to soil water. To circumvent this difficulty, the contribution of the various forces to the energy state of soil water is considered, thus permitting the addition of the contributions of each type of force. This involves measuring the work required to remove a unit quantity of water, and is known as the potential. Such work depends upon the degree of wetness, so that the energy state is not a linear function of water content. The units of measurement are energy content per unit of liquid volume of water. The commonly used units for potential are the bar, the millibar, and centimeter of water. Above the water table, work is required to remove water, and by convention, potential is negative. The potential at a flat or free air–water interface is taken to be zero.

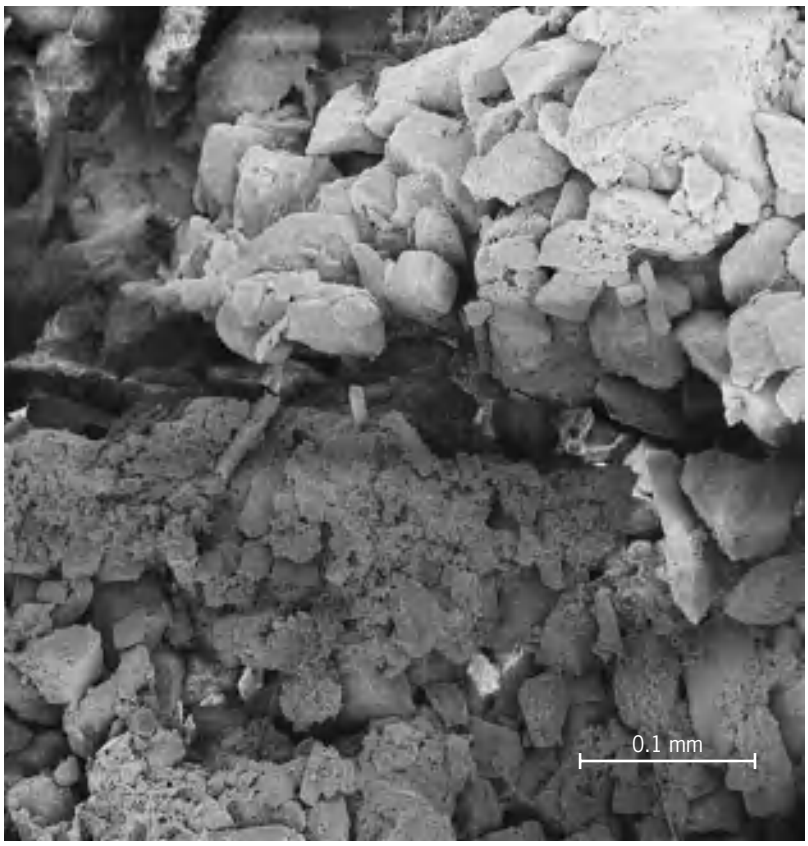


Fig. 20. Interface between disturbed soil and compact soil produced by the smearing action of a tillage implement as shown by a scanning electron micrograph. Water flow across such an interface is greatly reduced.

Practical problems of water flow in unsaturated soil involve mostly the matric and gravity potentials. For flow into roots, because they are semipermeable membranes, additional work must be done by a plant to remove water from those ions which do not readily pass through root membranes. Thus, in dealing with flow of water in soil-plant systems, osmotic potentials also must be considered. However, gravity potential often is small in unsaturated soil compared to matric and osmotic potentials, and sometimes may be neglected. Matric potential is a function of water content, pore size, and pore size distribution in the soil. As with capillary tubes, small pores hold water tightly, and considerable work must be done to remove it. The capillary rise equation may be used to relate the size of the largest water-filled pores (the size of a cylindrical pore with similar water-retentive properties) to the force required to remove water from wet soil, $h = 2s/dgr$ as described earlier. Such air-water interfaces in small pores are essentially hemispherical so that this radius also approximates the spherical equivalent of the irregularly shaped pore.

The capillary rise equation can be applied to a porous soil system in the wet range (Fig. 19). Pores in the porous plate are all small enough to remain

water-filled at the elevation of the hanging water column (smaller than the radius of a capillary tube which would raise water to this elevation). The radii of the air-water interfaces in the soil are given by the capillary rise equation. If the elevation h is decreased, larger pores become water-filled; if h is increased, only smaller pores remain water-filled. The elevation is an index of the matric potential, and the potential on a mass basis may be obtained by multiplying by the acceleration of gravity: hg (or if potential is on a volume basis, the value is hDg, where D is the density of water). The pressure in the water just below the interface as given by the equation $p = 2s/r$, and if the radius is taken as negative (the radius of a raindrop is positive and the pressure inside also is positive), the pressure may be seen to be negative or less than the pressure of the atmosphere. The use of the hanging water column apparatus (Fig. 19) is confined to the very wet range because of the limitation in the length of a hanging water column due to possible cavitation. However, positive pressures applied in a pressure chamber over a porous plate upon which samples are placed achieve the same result without any serious limitation to the range of measurements which can be made.

Matric potential in wet soil equals the work done per unit mass or volume to remove water from the air-water interface or against surface tension forces. As the soil becomes drier, larger and larger proportions of the water are associated with particle surfaces, and more work per unit quantity must be done to remove this water, which is tightly held by adsorptive forces in the particle surface. Hence, matric potential may be seen to involve both surface tension and the attraction of particle surfaces for water. Curves showing this relationship over a wide range of water content are obtained experimentally and used to characterize different soils (**Fig. 21**). However, since it is the porosity (a characteristic dependent not only upon textural classification but also upon the degree and kind of aggregation) that defines the relationship, different curves can exist for the same basic soil material treated differently.

The water potential–water content curves (Fig. 21) are desorption curves obtained in the process of removing water from saturated soil. A slightly different curve would exist for the wetting cycle because filling and emptying of large pores is controlled by the size of entryways rather than the size of the pore itself. Air is easily trapped in large pores during filling, and water is retained in large pores during drying, until entryways are emptied. This hysteresis phenomenon sometimes complicates use of water potential–water content curves, since it is impossible to determine one from the other without knowledge of the wetting history.

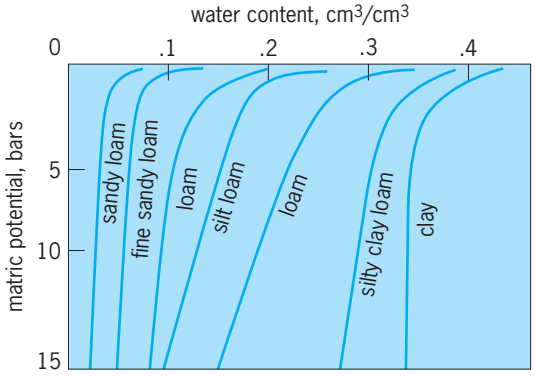


Fig. 21. Relationship between water potential (index of retentive forces) and water content for different soil materials. Wide variations often exist, even for materials in the same textural class. 1 bar = 10^2 kPa; 1 cm^3 = 0.6 $in.^3$

Water movement and retention. Water moves in soil in response to the net force acting. Above the water table in the region of a soil profile primarily involved with growing plants, matric forces are the major forces involved. Only near the water table where soil is very wet do gravitational forces play an important role. Where matric forces dominate, the water flow (flux) is the product of a conductivity factor and the moving force, which is the gradient of the matric potential. The conductivity factor is expressed as the product of a dimensionless number between 0 and 1 which expresses the degree to which the channel is filled with water and the value of the conductivity for saturated flow. This relationship describes saturated flow when the channels are full and where the matric potential is replaced by an equivalent potential derived from gravity and positive pressure, which is associated with saturation.

The channel-filling factor falls off rapidly as large pores empty, so that the cross section available for liquid flow is greatly reduced. As the soil desaturates, the pressure potential disappears and is replaced by a negative matric potential term. Matric forces are "pulling" rather than "pushing" forces, and water is pulled from regions of high potential into regions of low potential. Liquid water flows through water-filled interstitial space and along surfaces. The air-filled pores contribute nothing to liquid flow. Under such circumstances it may be seen that coarse materials, such as gravels, which would have high conductivity when saturated, would have extremely low conductivity at low water content. Thus water movement in unsaturated soil may be retarded in regions of large pores as well as in regions with fine pores. Aside from the influence of air-filled pores in reducing conductivity, conductivity also is reduced, beyond the amount expected from reduction of flow cross section, as pore size is reduced. The conductivity of capillary tubes varies with the fourth power of the capillary radius (Hagen-Poiseuille equation). On a unit area basis, the variation is with the square of the radius. Hence, with the force term held constant, reducing pore radius by a factor of 2 would reduce flow by a factor of 2^2, or 4.

Movement of water into soil, infiltration, and redistribution of water in soil following water addition slow down with time as a consequence of both changing gradients and reduction of unsaturated hydraulic conductivity as water moves from wet zones into drier zones. Slow redistribution is of particular

importance to consideration of water retention in soil. Water is nearly always moving, downward into dry soil or toward a water table, and upward in response to evaporative forces at the soil surface. Hence, soil has no unique retentive capacity. Water-retentive capacity is a dynamic property and must be defined in the context of change with time. A practical field capacity may be specified as the quantity of water in a defined depth of soil when rate of downward flow, or loss from a designated part of the profile, reduces to a value beyond which any further loss may be regarded as negligible. For agricultural purposes, and depending upon the nature of the profile, this rate would be reached in a period of time ranging from a fraction of a day to 10 or more days. The presence of soil horizons, involving either coarse or fine layers, reduces liquid water flow rates and increases water retention, generally reducing the time required to reach a negligible loss value.

An important application of unsaturated flow of water in soil concerns contamination of soil by radioactive materials or by other chemical or biological substances in spills. The seriousness of contamination depends upon the amount of water that is applied to soil as precipitation or through artificial means. From an agricultural point of view, slow movement of liquid water deep in the profile below the normal depth of roots may be of little consequence. However, when soil contains contaminants, particularly long-half-life isotopes, such flow could become important. Under some circumstances, the presence of layers of sand or gravel might constitute an effective barrier to the spread of some contaminants, but not when long periods might allow diffusion of substances in extremely thin films, which might under some conditions be present on particle surfaces or in vapor form.

Water vapor is present in soil air with relative humidity ranging from 98.9 to 100%, existing in soil wet enough to support plant growth. At uniform temperature the sum of matric and osmotic potentials, which affect evaporation, may be equated to relative humidity by the equation: matric + osmotic potential = $(RT/M) \ln (p/p_0)$, where R is the universal gas constant, T the Kelvin temperature, M the molecular weight of water, and p/p_0 the vapor pressure divided by the vapor pressure of saturated air at the same temperature, or the relative humidity. Where gradients of matric plus osmotic potential exist in soil, vapor pressure gradients also exist, and vapor flow can take place, even when liquid flow is negligibly small. If temperature gradients also exist, appreciable vapor flow can occur because of large vapor density gradients. However, in the absence of temperature gradients, vapor flow over large distances is small in dry soil. Because of low hydraulic conductivity values in dry soil, evaporative water loss from below depths of 25–50 cm (10–20 in.), in the absence of plants, generally is negligibly small, and moist soil, even in deserts, may be found below such depths. Flow of water in dry soil is primarily in vapor form, and the insulation properties of dry soil are high, so that high temperature gradients required for rapid diffusion of vapor do not exist. Where deep-rooted plants, particularly trees, are present without a water source, soil can be dried to great depths through the process of plant water uptake and evapotranspiration. Plant uptake has been used to remove unwanted chemical substances from soil and, in some coastal reclamation projects, to dry soil.

Air composition and flow. Gaseous composition of soil air, apart from water vapor discussed above, tends toward the composition of the atmosphere, with somewhat higher concentrations of carbon dioxide and lower concentrations of oxygen due to metabolic processes which utilize oxygen and give off carbon dioxide. Exchange with the atmosphere is most rapid in dry soil having high porosity and least rapid in dense or wet soil. The composition of the soil air affects plant growth and the growth of microorganisms and insects which inhabit the soil. Hence, limited air exchange with the atmosphere often has a marked effect on plant disease and plant growth generally. Oxygen requirements of plant roots vary with different plants, so that plant composition often is dependent upon aeration characteristics of soils.

Solute movement. Physical conditions of soils affect solute movement. Water is necessary in the vicinity of roots to solubilize nutrients contained in clays and other minerals and to form liquid paths through which the nutrients may diffuse and become available for absorption by roots. Also, solutes may be carried in water taken in through roots to form the transpirational stream. *See* PLANT MINERAL NUTRITION.

Solutes may be carried with moving water downward out of the rooting zone of plants and upward in moving water due to evaporation at the soil surface. In the latter case, transport to the surface and deposition through evaporation of pure water constitute a concentration process which may lead to high salt concentration in and on the soil surface. *See* ROOT (BOTANY).

Mechanical strength. Single-grain soil materials like dry sands have only a limited capability to withstand stress or compressive forces when unconfined. With small amounts of water added, this capability is increased because of surface tension forces in air–water interfaces which tend to hold particles together. Elimination of air–water interfaces by further additions of water, together with lubrication of particle surfaces, reduces the ability to withstand stress. Aggregated soil behaves somewhat in the same fashion except that stabilizing materials such as humus tend to bind particles into aggregates. Soil, particularly when moist, has some tensile strength. However, tensile strength is limited, as may be seen with swelling soils which may shrink and form large cracks upon drying. Compressive strength varies with degree of compaction brought about by animal or vehicle traffic, particularly at critical water contents somewhere between wet and dry. The water content for maximum compaction depends upon textural characteristics and types of clay minerals present. In a like manner an optimum water content generally exists for maximum effectiveness

of tillage operations designed to fragment the soil and reduce hardness.

Mechanical strength is an important consideration in agriculture as well as construction engineering. The agriculturalist needs soil which is strong enough to support small plants and trees, yet soft enough to permit easy movement of roots. Additionally, the soil should be sufficiently resistant to the mechanical forces of rain and running water to hold its position. The engineer needs soil with minimum shrink and swell properties and maximum strength over a wide range of water contents. *See* SOIL MECHANICS.

Soil temperature. Soil temperature depends upon absorption of solar radiation, reradiation from the surface, conductive exchange with the air, heat flow within, and the heat capacity of the soil. Soil color and surface texture influence both absorption and reradiation. Smooth, light-colored materials generally reflect light energy and are poor radiators, while rough, dark materials absorb or reradiate energy best. Thus, rough, dark soils tend to warm faster than smooth, light-colored materials. Organic residues on the soil surface play a major role in determining soil temperature through interception of both incoming and outgoing radiation and reduction in the velocity of air movement at the soil surface. Water content is the major factor involved in both heat transfer and heat retention, increasing both thermal conductivity and thermal capacity. Change of state of soil water—freezing, thawing, and evaporating—involves significant quantities of energy as latent heat. Soil temperature at a given depth and time depends on both heat conductance and storage, and complicated mathematical models are required for its prediction.

Plant growth and biological activity. Growth of plants, microorganisms, and insects in soil all involve establishment of an optimum physical environment. In turn, the presence in soil of organic materials and their decomposition products profoundly affects these physical properties. Broken or partially decomposed organic materials behave somewhat the same as mineral particles, and the decomposed material or humus acts as a binding substance imparting stability to mineral particle arrangements of soil aggregates. From a soil physical point of view, the major factors affecting biological activity are water, temperature, and gas composition. With some exceptions, soil supporting growing plants must have a matric-plus-osmotic potential ranging from about −15 bars (−1.5 megapascals) up to a small fraction of −1 bar (−0.1 MPa). If the soil is too wet, aeration becomes limiting. But, because aeration depends upon the quantity of air–filled pores, which will be different in different soils, it is difficult to specify a definite water potential value which will be limiting. Also, numerous agricultural plants grow best when the water potential is above −1 bar, even though plants might survive at potentials even below −15 bars.

Soils having mostly fine pores tend to remain wet in the spring and to warm slowly, so that many plants are delayed in growing because of low temperature. Optimum temperature for plants varies widely with species, and temperature thresholds exist for many plants; for example, corn and tomatoes grow poorly or not at all until day-time temperatures are well above 59°F (15°C), and growth is reduced if soil temperature greatly exceeds 86°F (30°C). However, many plants do well in cold soil.

Mechanical properties of soil can restrict root growth. Hard pans or plow pans formed by tillage operations often limit the growth of all but the hardiest of plants. Such pans may reduce the depth of the rooting zone. The hardness of the soil also is a factor in plowing and tillage operations. Since soil water content affects hardness, certain tillage operations are timed so as to take advantage of a water content that produces softer soil. If too wet, however, a soil may puddle badly when tilled. Thus an optimum water content for such operations exists. Appreciable attention is being given to no-till or minimum tillage to avoid creating hard soil by excessive traffic compaction and to maintain the protection of plant cover against erosion. Weed control, one of the important functions of tillage, is done chemically when these practices are followed. *See* HERBICIDE; PLANT GROWTH.

Walter H. Gardner

Spatial Variability

Soil properties vary vertically at a given site, horizontally from place to place, and both vertically and horizontally with time. Determination of the spatial variability of soil can be accomplished by kriging, a geostatistical technique that was developed to aid mining engineers make better estimates of ore deposits in unsampled locations. Geostatistics is the statistics of spatially correlated data; since soil properties are generally correlated spatially, geostatistics and kriging have become popular methods for describing and dealing with this complex aspect of soils. Kriging is a technique of making optimal, unbiased estimates of soil properties at unsampled locations.

Describing spatial variability of soil is important only insofar as it enables users of land resources to make better predictions of soil behavior and performance. The precision with which soil behavior and performance can be predicted depends on the uniformity of the land in question. While it may be possible to use an average value to estimate soil behavior or performance, it is often not very useful and sometimes dangerous to depend on averages, especially if different parts of a land parcel behave very differently. For example, a home owner who plans to install a septic tank needs to know the hydraulic characteristics of the plot and not the average value for the neighborhood. Since the performance of each plot as a septic tank field depends on the hydraulic behavior of the field, there is economic value in any method that enables a user to estimate the local value with greater precision than the regional mean. Kriging achieves this aim by using observed values in the neighborhood to estimate values at unsampled locations.

If all observed values are identical, the average of the observed values would be an excellent estimator of the same property at any point in the region. The

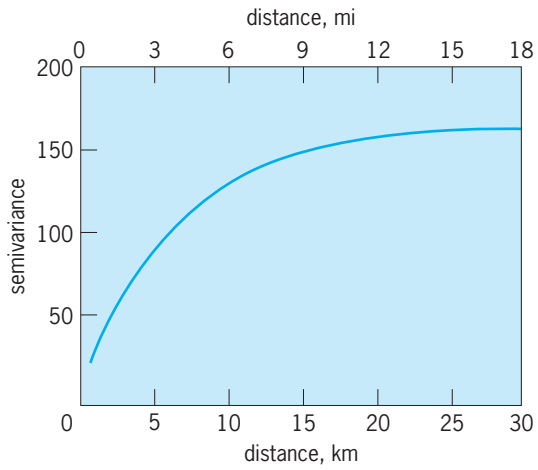


Fig. 22. Semivariogram of sand content showing a range of spatial depedence of about 9 mi (15 km). Data points represent the experimental values, and the curve is an exponential model.

problem arises when the variance in the observed data is high, and the user is unwilling to accept the mean as an estimator of values at unsampled locations. In such cases, kriging can provide better results, provided certain conditions are met.

Kriging depends on the fact that closely spaced samples tend to be more alike than samples separated by large distances. The difference in clay content between soil samples taken 1 m apart, for example, more likely will be smaller than that of the samples taken 1 km apart. In kriging, the clay content is expected to vary from point to point, but the difference in clay content between two points is expected to vary only with separation distance. Very often, however, the difference varies not only with separation distance but with orientation as well.

Considering the simpler case, in which the difference varies only with separation distance, that information can be captured and condensed in a statistical representation known as a semivariogram: each point on such a curve (**Fig. 22**) represents the sum of squared differences for a particular separation distance divided by twice the number of pairs. The semivariogram is calculated from Eq. (1), where

$$\gamma(h) = \frac{1}{2n}\sum[z(x) - z(x+h)]^2 \qquad (1)$$

$\gamma(h)$ is the semivariogram [$2\gamma(h)$ is the variogram], n is the number of pairs, z is the soil variable at location x, and h is the separation distance. In the "ideal" semivariogram, $\gamma(h) = 0$ when $h = 0$; that is, samples separated by zero distance are identical. In practice, the curve frequently intersects the vertical axis at a value above zero. The point of intersection is called the nugget variance, a term that links geostatistics to its early beginning in the gold mining industry. Another characteristic of the semivariogram is that it levels off to a value called the sill. It turns out that in the ideal case the sill corresponds to the ordinary sample variance. The sill, therefore, corresponds to variances associated with samples that are spatially independent. The maximum distance at which sample pairs are correlated is called the range. For sample distances shorter than the range, the sample property is said to be spatially dependent. Note that if by chance the sampling distance exceeds the range, one obtains a semivariogram in which the nugget variance and sill are the same. For purposes of kriging, separation distances within the range of spatial dependence are required. Choosing the minimum sampling distance so that it falls well within the range requires knowledge of the study area. Trends in the data and anisotropy are frequently encountered. Trends are regular changes in a value, such as soil temperature decreasing with elevation. Special techniques are required to eliminate the trend from the data. Anisotropy is encountered when the difference between sample values depends not only on separation distance but on orientation as well.

In order to use the structure in the variance for kriging, the experimental semivariogram is fitted to a model semivariogram. The three most commonly used models are the spherical, exponential, and linear models: the exponential model (**Fig. 23**) is fitted to the experimental semivariogram.

In kriging, local values are estimated from the weighted average of observed values within the range of the point to be estimated. The estimated value of the variable z at location x_0 is given by Eq. (2), where n is the number of observed values

$$z(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \qquad (2)$$

$z(x_i)$, and λ_i are weights applied to each $z(x_i)$. The weights are calculated so that the estimate is unbiased and the estimation variance is minimized. The estimation involves solving a set of $n + 1$ linear equations with $n + 1$ unknowns that describe the expected autocorrelation between observed values and the value to be estimated.

For purposes of mapping, the points to be estimated can be located in a square grid (Fig. 23). The

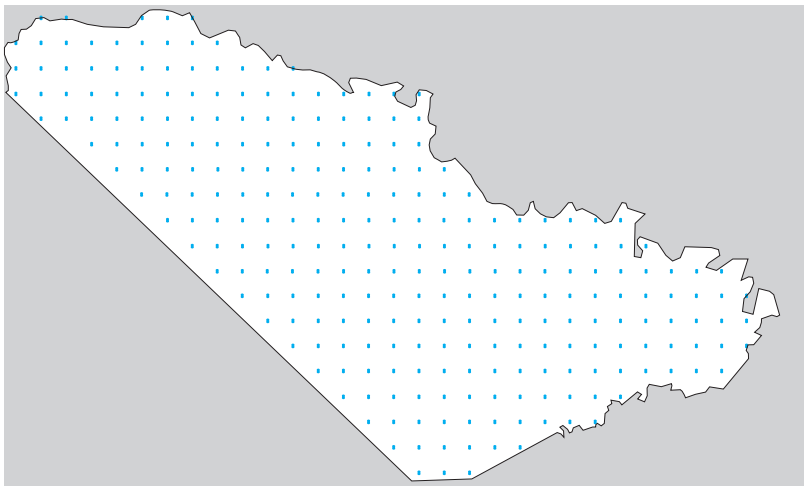


Fig. 23. Locations of 268 points for kriging sand content. The points are 1 km (0.6 mi) apart.

Fig. 24. Isarithm maps. (*a*) Map of actual sand content; values at the isarithms are in percent. (*b*) Map of the estimation variance of percent sand content; values at the isarithms are in percent squared.

estimated values and their corresponding estimation variance can then be used to produce isarithm maps (**Fig. 24**). Areas with low sampling densities will show high estimation variances, so that the estimation variance map can be used to guide subsequent sampling to improve map quality. If the variable in question is difficult or costly to measure, it is easier and cheaper to measure another variable that correlates highly with the variable of interest, and then compute linear estimates through the technique of cokriging. Local estimates of undersampled variables can also be improved through cokriging. Cokriging is, therefore, an interpolation technique that increases the accuracy of an estimated property by using additional information provided by the structure in the variance of a covarying property. Cokriging can be especially useful in soil science because most soil properties are spatially correlated and intercorrelated.

It is often more practical to estimate the local value of a block of land than a point in the landscape. This procedure, known as block kriging, provides a way to estimate an average value for an area of land. The area may be the size of a experimental plot, septic tank field, or a corn field. The difference between point kriging and block kriging is in the calculation of the weighing coefficients. In block kriging, the semivariances between observed points and the estimated values of point kriging are replaced by the average semivariances between the observed points and all points in the block. As in point kriging, the weights are calculated so that the estimation variance is minimized.

The ability to determine the property of a soil at any point or block in the landscape provides users of land resources with information that permits sounder economic and safer environmental land-use decisions to be made. Spatial variability can be costly to farmers and other users of land resources. Farmers who apply agricultural chemicals uniformly in a field with spatially varying soils will add too much to places where the requirement for the chemical is low, and too little where the requirement is high. Farmers who cultivate large tracts of land are finding it profitable to invest in farm machinery that can match application rates to spatially varying soil and crop requirements with a high degree of precision. While economic considerations serve as the main incentive, stricter environmental regulations may compel farmers to adopt precision farming, and kriging offers a way to predict optimum application rates of agricultural chemicals.

Goro Uehara

Management

Soil management can be defined simply as the manner in which people use soils to produce food, fiber, and forages. Soil management includes determination and use of many factors and practices, such as land survey maps, cropping systems, organic matter and tilth, soil fertility, salinity, and irrigation. Most often, it is not a single soil that is managed; management is carried out on a field or a portion of a landscape composed of a number of physical and biological features, like climate, vegetation, topography, and drainage. In a true sense, applying soil management strategies involves an integration of a number of factors. Such integration involves a difficult synthesis of many individual characteristics, both measured and observed, to arrive at a meaningful interpretation of how a soil responds to management. In fact, the value of any agricultural soil is determined by how well that soil responds to proper management.

Standards. Meaningful assessment of how soils are responding to a particular management system can be judged only if standards are established. There are two primary standards for judging quality of management: prevention of soil deterioration or degradation, and improving the soil system in ways that result in increased plant production. Examples of deterioration that must be prevented are excessive erosion, fertility depletion, and accumulations of salt within the rooting zone of plants. Examples relating to improving the soil system are conservation of precipitation, irrigation development, reclaiming salt-damaged land, and increased water use efficiency by proper fertilization. *See* AGRICULTURAL SOIL AND CROP PRACTICES.

The two goals are very broad and therefore require the integration of many aspects of science. Incorporation of the basic areas of chemistry, physics, and microbiology into management planning is essential. The sciences of ecology, genetics, and various types of engineering are also essential to modern soil management. Computer modeling will play an increasingly important role in soil management by making simulation models directly available to farm and ranch operators. *See* SIMULATION.

Land capability classification. The first step in planning a management system is assessment of the soil and environmental resources. This is accomplished through land classification—more specifically, through land capability classification, which groups land according to properties essential for identifying the opportunities or constraints the land offers for various uses.

Interpretations of land capability classifications are most often based on physical, chemical, and economic considerations. Classification provides assessment of land suitability for agricultural and other uses, identifies adapted crops, estimates yields of crops under defined systems of management, indicates presence of specific soil management problems, and delineates opportunities and limitations of various management practices. Through land capability classification, there usually are identified several alternative management strategies for a given type of land. From this information, analyses can be made to determine which land use or management strategy will be most desirable for both economic and wise land utilization.

The basic requirements for developing a good classification system are soil-survey and soil chemical and physical data; climatic, topographic, and hydrologic data; field and laboratory research data; long-time land-use records; and experience and observations of land response to various uses.

Soil surveys are basic in that they identify soil properties important for determining land-use capability. Surveys provide maps showing the location and boundaries of soils. Soil depth, texture, kinds of minerals, salinity and kinds of salts, and acidity are some properties utilized in determining land capability. The particular set of properties needed for determining land capability is dependent on the particular use.

Land classifications cannot be static because they depend upon interpretation. As technology changes and economic and social conditions change, interpretations change. With basic soil, climatic, and topographic data and maps, however, these interpretations can be revised easily. Combinations of alternative land-use and management strategies, as well as social and economic conditions, do change, but the physical and chemical factors of well-managed land do not change very much.

Land classifications have been developed for specific purposes, for example, tax assessment, sales and credit, soil and water conservation, irrigation potential and management, wildlife suitability, watershed management, recreational uses, industrial uses, or highway construction.

Effective extrapolation of management or use strategies to the land depends on an adequate inventory of basic data and how well an existing classification is chosen or a classification developed that fits the specific need at a given time and place. With appropriate facts and maps, through land classification, predictions can be made about the results of using a specific type of land in a particular way. Then, planning of land use includes the practical combinations of management practices required, and the effects of management on the quality of land resources. Without land classification and appropriate maps, it is difficult to extrapolate experience and research results to the land.

Land capability classification systems have been utilized for many years by the Soil Conservation Service, USDA, and other land resource planning and management agencies. Their utilization is being enhanced greatly through the use of computer data storage and management systems. *See* LAND-USE CLASSES.

Long-term impacts. Modern soil management practices have a cumulative effect on the soil's future productivity. The foremost management question for the soil manager is whether a given practice or system is causing soil degradation. Many current agricultural problems have developed because managers did not realize the long-term impact of their management techniques. Some practices may appear to be sustainable for periods of time equal to a human lifetime, and yet they could be causing slow soil degradation that will ruin the soil for long-term use. Therefore, long-term plans that identify causes of soil deterioration and avoid it successfully are essential to a nation's future agricultural productivity. For example, proper use of fertilizers and manures can enhance long-term productivity and maintain environmental quality. Poor management choices including fertilizers and manures can actually ruin a soil and damage the environment, almost irreparably. Civilizations have often missed the "clues" that a particular system was not sustainable. This has led to loss of their food supply, and with their being conquered eventually or ceasing to exist.

Soil water. Water conservation and efficient use of stored water are important issues within soil management. Since plant growth is limited by water supply in almost every climate situation at some point in the crop cycle, having adequate stored soil water is critical. Irrigation becomes important in climates where the natural precipitation will not meet plant demands or in areas where rainfall patterns are erratic.

Water conservation principles are the same in either rain-fed or irrigated situations, and can be divided into two phases: water capture in soil, and water retention in the soil. Water capture is maximized by maintaining open pores at the soil surface and protecting these surface pores from raindrop impact which would otherwise destroy surface

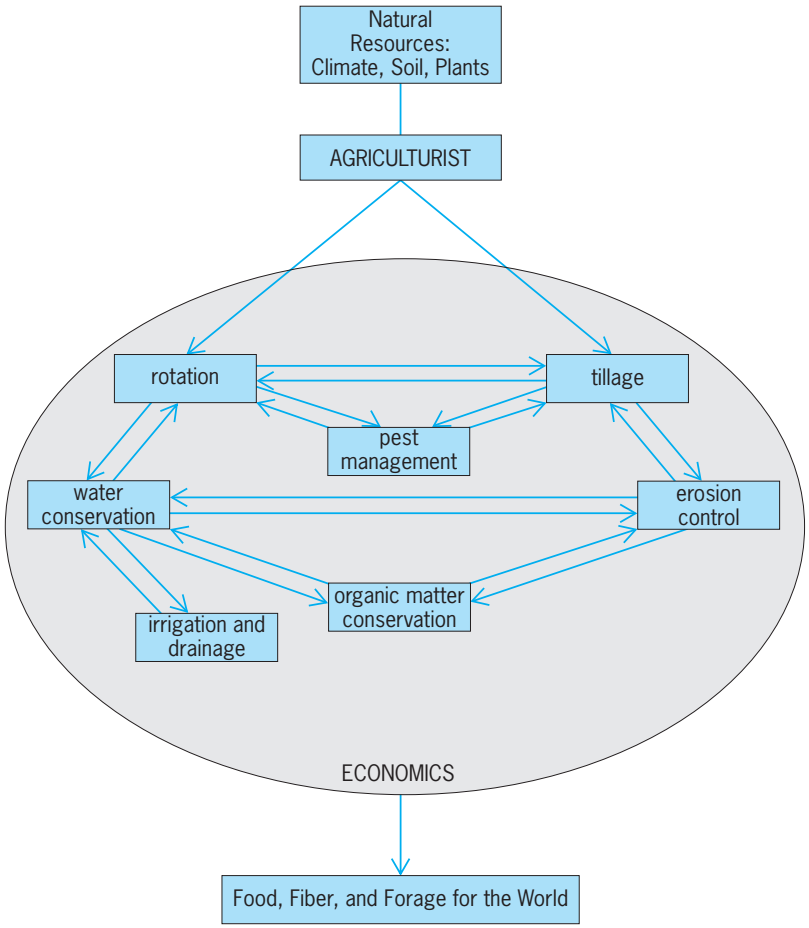


Fig. 25. Diagram showing an integrated system of soil management.

soil structure and ultimately plug the pores. Crop-residue mulches are effective protectors of structure, and these are best maintained by reducing tillage. Residues also slow runoff water and increase the opportunity for infiltration into the soil. The second phase of water conservation, retention of captured water, requires prevention of weed growth that would waste the captured water, and reduction of evaporation. Protecting the soil surface from exposure to sunlight and wind by maintaining crop residues at the surface is an important management tool for evaporation control. Residues serve as reflectors of light energy, as insulators to keep the soil surface cool, and as wind deflectors. Reducing and even eliminating tillage maximizes the crop residue on the soil surface and consequently maximizes water retention.

Irrigation is not always an option. Many locations in the world do not have adequate water supplies to allow irrigation, even though the crops could benefit from additional water. In some cases, underground water supplies are present but are so deep that the additional crop growth with irrigation would not offset the cost of pumping the water. Soil factors such as texture, depth, and salt content also dictate the feasibility of irrigation. Some very fine textured soils will not absorb water readily and are not good candidates for irrigation; others are so shallow that the storage capacity is too little. If soils have too many soluble salts, they are not suitable for irrigation. There are many combinations of climate, soils, water supply, and economic parameters that must be considered when determining if irrigation will be feasible or profitable.

All irrigation waters, whether they are diverted from rivers or pumped from the underground supply, contain some dissolved salts. Thus, when soils are irrigated they receive a dose of these salts. If the irrigation manager does not plan for these salts, the soil can become saline and lose productivity. *See* WATER CONSERVATION.

Integrated systems. Soil management involves integration (**Fig. 25**). The resource block includes climate, soil, and plants; these are the natural-resource inputs. Modifications can and will be made on them, but for the most part they are the constants of the system. The agriculturist enters the scheme at this point to impose a management system, which has as its goal the production of food, fiber, or forages. Figure 25 also shows that the cropping system imposed is directly related to the soil, climate, and plant resources. The manager implements a cropping system through choice of crop rotation and tillage system. Choice of tillage system is often conditioned by the rotation chosen.

Erosion control, organic matter conservation, and water conservation are the classical concerns of soil management; they are all highly linked. A change in one or more of these three factors creates feedbacks to the others. For example, control of erosion leads to direct savings in soil organic matter and soil water and thus soil conservation. Furthermore, the control of erosion, which relates to the first goal of preventing soil deterioration, is linked directly to the second goal, namely, improvement of the soil system. Erosion prevention techniques also improve soil water storage by reducing water runoff. Conservation of organic matter may increase water infiltration, which decreases runoff and thus decreases erosion. The other blocks in Fig. 25 represent pest management and irrigation and drainage. They are equally important, but have not been studied to as great an extent as the first three factors. In fact, pest and pesticide management are relatively new concepts in soil management. Soil tillage methods and rotations are intimately involved in pest and pesticide management.

Finally, basic economic principles are involved in all sound soil management systems, as indicated in Fig. 25 by the enclosure of all other factors by economics. This is true in even the most primitive civilizations. If the people cannot live on the food or income produced, the system cannot survive. In more developed nations, quality of management has even influenced the potential value of the land. The ability to cope with the problems of a particular soil eventually determines its economic value to society.

Forces external to the system, such as changes in government policy, can greatly impact the overall management system. For example, the government has funded land set-aside programs for multiple

reasons, ranging from attempts to decrease production of various crops to efforts to slow soil erosion. These interventions cause multiple, and often unexpected, interactions within the system. The programs encourage producers to take land from production of annual crops, such as corn, cotton, sorghum, and wheat, and to replace them with perennial grasses. During the contract period, the producer cannot harvest or graze the grass, and it is left as a cover crop on the soil.

Quite obviously these land set-aside programs have an economic impact on both the producer and local community. The producer experiences less risk because inputs to the system are low and the income is stable, no matter what the weather is. At the same time, the community merchants and the local labor market are often negatively impacted because the merchant sells fewer inputs and there is less need for seasonal labor. In contrast to the mixed economic outcome, the set-aside programs have nothing but positive influences on the soil itself. Soil erosion is slowed, relative to the conditions prior to the set-aside program, and soil organic matter contents stabilize or rise. However, these potentially positive soil changes may not be long-lasting because government policies change on a short-term basis (5–10 years) while soil changes require multiple decades to be of long-term value to the nation. Depending on the management scheme adopted by the producer, the benefits of 10 years in perennial grass can be lost in 1–2 years.

High-quality, efficient management of the natural resources of climate, water, and soil is needed to provide the food and fiber that will sustain life on Earth. Soil management decisions will have an important effect on the future food supply.

G. A. Peterson; W. O. Willis

Salinity

Soluble salt and exchangeable cation concentrations play major roles in determining the pH, physical characteristics, and chemical composition of soils. When a salt dissolves in water, it dissociates or separates into cations and anions. The predominant cations in salt-affected soils are calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^{+}), and potassium (K^{+}); the predominant anions are chloride (Cl^{-}), sulfate (SO_4^{2-}), carbonate (CO_3^{2-}), and bicarbonate (HCO_3^{-}).

Clays and organic matter contain negative electrically charged sites. In salt-affected soils, this charge is satisfied by calcium, magnesium, sodium, and potassium ions. The cations, bound to the exchange sites by electrical charge, are known as exchangeable cations because they can be removed from the charged surface only by replacement with another cation from the soil solution. *See* HUMUS; SOIL CHEMISTRY.

Classification. Each soil can be classified as normal, saline, sodic, or saline-sodic, based on its salt content and exchangeable cation ratios.

Normal soils. These soils do not contain sufficient soluble salts or exchangeable sodium to adversely affect plant growth or soil physical properties. Normal soils have saturation paste pH values less than 8.3 and saturation paste extract electrical conductivities of less than 4 deciSiemens per meter (dS/m) and an exchangeable sodium percentage less than 15. A saturation paste is made by mixing just enough distilled water with a soil sample to fill the voids without having excess water standing on the surface of a well-mixed sample after 4–16 h. These electrical conductivities are defined upper limits, but if salt-sensitive plants are grown on soils with an electrical conductivity of 3.5 dS/m, a significant yield reduction will result. Likewise, using a high-volume sprinkler system to irrigate a soil with an exchangeable sodium percentage of 12 could produce serious runoff rates because of low infiltration rates. *See* HYDROLYSIS; PH.

Saline soils. These soils contain sufficient soluble salts (electrical conductivity greater than 4 dS/m) in the upper root zone to reduce yields of most cultivated crops and ornamental plants. The exchangeable sodium percentage is less than 15, the sodium absorption ratio is less than 13, and the pH is less than 8.3. The predominant cations are calcium, magnesium, and, in a few cases, potassium. The predominant anions are chloride and sulfate. Water entry and movement through these soils is not inhibited by high exchangeable sodium concentrations. Osmotic effects and chloride toxicity are the predominant causes of plant growth reduction. *See* OSMOSIS.

Sodic soils. These soils are lower in soluble salts than saline soils (electrical conductivity less than 4 dS/m). The exchangeable sodium percentage is greater than 15 and the sodium absorption ratio of the saturation paste extract is greater than 13. The pH of the saturation paste is greater than 8.5. Bicarbonate, carbonate, and hydroxide (OH^{-}) ions are the predominant anions in these soils; these anions cause calcium to precipitate from solution as calcium carbonate ($CaCO_3$; lime). The combination of high exchangeable sodium percentage and pH, and low electrical conductivity and exchangeable calcium causes the clay and organic matter to disperse, which in turn destroys the soil structure or tilth, causing so-called slick spots. These spots have extremely low rates of water and air exchange. They often have a black, greasy, or oily-looking surface due to the dispersed organic matter. Vegetation may be absent because of low water infiltration and insufficient plant-available water.

Saline-sodic. These soils are similar to saline soils in that the electrical conductivity is greater than 4 dS/m and the pH is below 8.3. Saline-sodic soils differ from saline soils in that more than 15% of the exchangeable cations are sodium and the saturation-paste sodium absorption ratio is greater than 13. The anions are a mix of bicarbonate, chloride, and sulfate. As long as the electrical conductivity remains above 4 dS/m, infiltration rates and hydraulic conductivities are similar to those of normal or saline soils. Irrigating saline-sodic soils with water having low salt concentrations will convert them into sodic soils if they do not contain gypsum (a calcium sulfate mineral).

This happens as the electrical conductivity decreases without a decrease in the exchangeable sodium percentage, causing the undesirable properties of sodic soils to be expressed. It is not uncommon to have a mix of two or more classes of salt-affected soils within a field. Salt-affected soils tend to be highly variable from one part of a field to another.

Sources. Most soluble salts and exchangeable cations are derived from rock and mineral weathering of the soil parent materials. In high-rainfall, humid, and tropical areas, rain and melting snow leach the salts from the soil as they form. In arid and semiarid areas, the annual evapotranspiration potential is greater than the total annual precipitation, and the salts are not always leached from the soil as they are released. With time, they may accumulate in the root zone at concentration levels that affect plant growth.

Salts can also accumulate above shallow water tables as water moves to the soil surface by capillary rise (wicking) and evaporates, leaving the salts on or near the surface. Shallow water tables may occur naturally, induced by irrigating poorly drained areas, by irrigating upslope from lowlands, or by construction activity that blocks natural subsurface lateral drainage.

All natural waters contain dissolved salt. In many arid and semiarid areas, good-quality irrigation water (low in salts and low in sodium) is not available; consequently, irrigation water is used that contains more salts or sodium than is desirable. If sufficient water does not move through the soil and leach the salts below the root zone, salts or sodium will accumulate in the soil. It is often stated that under irrigation "hard water makes soft soils and soft water makes hard soils." This implies that irrigation waters containing predominantly calcium and magnesium salts (sodium absorption ratio less than 3 or 4) tend to promote a more friable soil condition than do waters with high sodium concentrations.

Four conditions must be satisfied in order to remove soluble salts and excess sodium from soils: (1) less salt must be added to the soil than is removed; (2) salts must be leached downward through the soil; (3) water moving upward from shallow water tables must be removed or intercepted to avoid additional salts moving back to the soil surface; and (4) in sodic and saline-sodic soils the exchangeable sodium must be replaced with another cation, preferably calcium, and the sodium leached out. Applications of soil amendments (gypsum, iron sulfate, sulfur, or sulfuric acid) are beneficial only on sodic soils when leaching also occurs and on leaching of saline-sodic soils that do not contain natural gypsum.

Saline and sodic soils are found primarily in arid and semiarid areas of the world. Exceptions are recently drained coastal areas, salt marshes, and soils formed in depressions from marine deposits where the weathering products are not leached from the soil. Sodic and saline-sodic soils can also form where mists are carried from ocean waters by the wind and deposited on soil surfaces in arid and semiarid coastal areas. Aridisols and Entisols include most salt-affected soils. Low rainfall and unweathered soil materials result in insufficient salts leaching from the root zone of these soils. Mollisols, Alfisols, and Vertisols also contain considerable saline and sodic soil areas.

Human activities such as spills or intentional dumping of salts or solutions from oil well drilling-mud ponds, mines, food-processing plants, municipal sewage water, power-plant cooling-tower water, or heavy applications of wood ash can induce saline and sodic conditions in any soil when soluble salts are applied faster than they are leached from the soil.

C. W. Robbins

Identification and mapping. Because of the highly variable nature of salt-affected soils and the expense of reclamation, field mapping is prerequisite to effective reclamation. Visual observation mapping of plant growth and soil appearance is the quickest and least expensive method for mapping low-, medium-, and high-salinity impacted areas of the landscape. Native plant species and stunted growth or leaf burn in cultivated crops are effective identifiers of saline and sodic areas. Saline soil areas will also have white salt crusts, while sodic areas will be barren with dark-to-black, oily-looking slick shining surfaces bordered by water-stressed plants. Remote-sensing procedures, such as aerial photography, videography, infrared thermometry and imaging, multispectrial scanners, microwave sensors, and time domain reflectometry, also are used for salinity mapping. Once the productive and problem areas are defined, deep sample holes can be bored to determine if a high water table is the salt source. Water samples from high water tables should be analyzed for salt concentration and type. Soil samples should also be analyzed to determine salt type and concentration. Exchangeable cation concentrations are also needed to determine the degree of the sodium problem. Irrigation-water salt concentration and cation ratios are also important factors in salinity management. Once the salt sources, concentrations, and cation ratios are determined, a reclamation plan can be developed; or if reclamation is not practical, crops tolerant to the conditions can be selected.

Reclamation. Four conditions must be satisfied in order to reclaim salt-affected soils by removing soluble salts and excess sodium: (1) less salt must be added to the soil than is removed; (2) salts must be leached downward through the soil; (3) water moving upward from shallow water tables must be removed or intercepted to prevent additional salts from moving back to the soil surface; and (4) in sodic and saline-sodic soils the exchangeable sodium must be replaced with another cation, preferably calcium, and the sodium leached out. Applications of soil amendments (gypsum, iron sulfate, sulfur, or sulfuric acid) are beneficial only on sodic soils when leaching also occurs and on leaching of saline-sodic soils that do not contain natural gypsum. Adding chemical amendments such as gypsum to saline soils only adds more salts and is not needed unless the water has a sodium adsorption ratio of 10 or greater.

C. W. Robbins

Effects on plants. Many ions are essential to plant growth as major or minor nutrients. However, when ion concentrations become too high, plant growth is adversely affected by either the toxic effect of a specific ion or the general effects of high ion concentrations. Salinity decreases plant growth through a combination of nonspecific ionic effects and by causing a decrease in the water potential of the soil, which is principally an osmotic effect.

The point at which salinity limits plant growth varies because plants have adapted to a wide range of salinity environments. The ocean, which has salt concentrations in excess of 35 parts per thousand (ppt), is abundant in plant life and contains over half of the Earth's plant biomass. Concentrations at which specific ions become harmful to plant growth also vary over several orders of magnitude. For example, boron is toxic to some plants at soil-water concentrations as low as 0.05 mol/m^3, whereas some plants can tolerate chloride concentrations as high as 20 mol/m^3. Sodium and chloride are the most abundant salinity-inducing ions in soils and water; however, significant amounts of calcium, magnesium, sulfate, carbonate, and bicarbonate ions are also found in nature. The proportions of these ions vary with respect to soil types and local geology.

Terrestrial plants that are tolerant of high concentrations of soluble salts in their root zones (the area around the root from which a plant extracts water) are known as halophytes. Halophytes can survive and complete their life cycles at optimum salt concentrations of 1.2–30 ppt in their root zone. Most terrestrial plant species are not adapted to high salt concentrations and may be considered glycophytes.

In agriculture, salt is a serious hazard in irrigated areas if growers do not leach their soils properly during irrigation, fail to provide adequate drainage for their crops, or allow their water tables to rise too near the surface. High concentrations of salts in the water used for irrigation may also damage crops or reduce yields. High-salinity waters may include ground and surface water and water recycled from municipal and industrial uses. Salinity acts as an environmental plant stress that may cause leaf damage of reduced growth or, at high concentrations, may be lethal. Plant responses to excess salts in their root zones or on leaf surfaces (from ocean sprays or irrigation) are quantitatively dependent on salt concentration, composition, and time of exposure. Plant sensitivity to salt also varies according to growth stage. Generally, plants are salt-tolerant during germination, are most sensitive during the seedling stage, and become more tolerant with maturity. Plants also may show increased salt sensitivity during their reproductive stage, as seen by decreased pollen viability and seed setting ability.

Crop salt tolerance. Salt tolerance is the capacity of a plant to endure excess salt in its environment. This characteristic is quantitative and influenced by many soil, climate, and cultural factors. Tolerance assessment may be based on the ability of the plant to survive, to produce high yields, or to withstand adverse growth reductions. In nature, the measure of tolerance may be the ability to survive, reproduce, and compete with other species; whereas in sustenance agriculture, tolerance may be related to both survival and productive yield. In commercial agriculture, the ability of the crop to withstand salt effects without reducing yields below the profit margin is the most important consideration. Thus crop salt tolerance is usefully described by two parameters: the threshold salinity (T) at which yield reduction is significantly measurable, and the rate (R) at which yield decreases with increasing salinity beyond the threshold (**Fig. 26**). The rate of yield decrease for most crops can be described simply as the slope of a straight line. Salinity concentrations are described as an index of the electrical conductivity of a soil-saturated paste in units of dS/m. Salt tolerance parameters are useful for predicting how one crop may compare with another under similar conditions. However, such assessments are general and relate to crop growth after germination and seedling establishment.

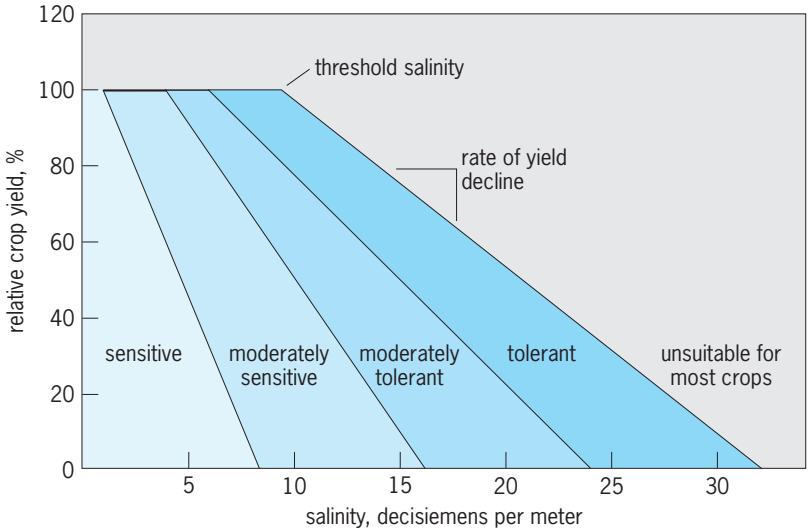


Fig. 26. Typical classifications for salt tolerance of crops based on their relative yields under nonsaline conditions in contrast to yields under increasing saline conditions.

Climate and agricultural management practices may reduce or increase the effects of salinity upon plants. Irrigation and management practices that leach salts away from, or maintain lower concentrations of salt in, the root zone during growth will reduce salt effects. Seed beds should be sloped or maintained in a manner that allows the irrigation water to move salts past the root zone. If excess salts in the seed bed are not kept low, the resulting reduction in plant stand will decrease yields far more than is predicted by salt tolerance parameters. Flood, furrow, drip, and sprinkler irrigations should also be applied at times and in ways that reduce salt accumulation on plant parts and within root zones. Climate factors such as high temperature, low humidity, and high wind speed will increase salt damage, whereas factors that reduce transpiration demand will reduce salt damage. Soil type is also an important factor. Sandy soils will not accumulate salts as readily and are easier to leach than clay soils. *See* IRRIGATION (AGRICULTURE).

Genetic variability. Many crops, such as sugarbeet, asparagus, date palm, cotton, and barley, are salt-tolerant and are standard crops in saline areas. Different types of beans and berries, as well as avocado and many fruit trees, are very sensitive to salt. In salt-sensitive species, sensitivity is often associated with the accumulation of a specific ion in leaves, usually chloride or sodium. Saline soil environments with high proportions of sulfate compared to chloride may have less severe salinity effects on chloride-sensitive plants. The ability to exclude specific ions at the level of the root or shoot is one of the major causes for genetic variability among plant varieties and ecotypes. Rice, although salt-sensitive, is grown on saline lands for reclamation because it has a shallow root zone and can be grown on flooded fields if water of good quality is available. Salt tolerance also varies less between cultivars and ecotypes of the same species.

Conventional breeding efforts to improve salt tolerance of crops include selection for more tolerant cultivars through hybridization among varieties of a species; hybridization of a cultivated species with related, wild salt-tolerant species to increase genetic variability prior to selection; and exploitation of the useful agronomic potential of wild halophytes. The ability of the grower to control the effects of salt in the field is of more greater consequence than the variability in salt tolerance among cultivars.

Morphological and physiological effects. Salinity reduces plant growth through both osmotic and ionic influences. The osmotic effects are a result of increased solute concentrations at the root-soil water interface, which create lower water potentials. Growth suppression is the result of total electrolyte concentration, soil water content, and matrix effects, and is manifested in reduced cell enlargement and metabolism. The plant suffers water stress for a short period until it can make some type of osmotic adjustment. Plants make this adjustment by accumulating more salt within their tissues (a halophytic response) or by the synthesis of organic solutes, which increases the osmotic potential of the cytoplasm so that water will flow into the root and plant tissues.

Ionic effects may be both general and specific. General ionic effects are the result of the increased ionic strength of the soil water. Ionic effects may interfere with the normal processes by which plants take up nutrients by changing the surface chemistry at the cell wall and plasma membrane. Specific ions may disrupt normal metabolic processes or upset nutrient balances. For instance, high sodium concentrations relative to other salts can disrupt root permeability to ions by displacing calcium in the plasma membrane. Upsetting calcium metabolism and nutrition within the cell may cause additional effects. At higher sodium-to-calcium ratios, soil structure, tilth, and permeability of the soil to water may be reduced (sodicity).

The specific physiological cause of growth reduction due to salt stress is undoubtedly complex in both the metabolic and genetic sense. Salt stress reduces plant growth primarily because it increases the metabolic energy needed to acquire water from the root zone and to make the biochemical and morphological adjustments necessary to maintain growth in a higher ionic environment. *See* PLANT-WATER RELATIONS; PLANTS, SALINE ENVIRONMENTS OF; SOIL FERTILITY.

Michael Shannon

Erosion

Soil erosion results from the detachment and transport of soil materials by water. Geologic erosion and erosion from human activities are the principal types. Long-term geologic erosion creates topographic features such as canyons, stream channels, and valleys. Removal of natural vegetation by human activities, such as farming, ranching, forestry, and construction, may also cause erosion.

Excessive erosion could threaten the world supply of agricultural and forest products. The efficiency of water conveyance and storage structures may be significantly impacted by sedimentation resulting from soil erosion. Excessive amounts of sediment in streams and rivers can reduce their suitability as a biological habitat and create water supply difficulties. Understanding the various types of erosion and the factors affecting erosion is necessary in identifying appropriate control practices.

Types. Erosion by water occurs when soil particles are detached from the soil surface and then transported by runoff. As runoff rate increases, small channels, called rills, begin to form. The region between rills is defined as the interrill area. When concentrated runoff is sufficiently large to cut deep channels, gully erosion occurs. Stream channel erosion may develop within a water course that has nearly continuous flow. Interrill erosion, rill erosion, gully erosion, and stream channel erosion each have distinct characteristics.

Interrill erosion. On interrill areas, raindrops impacting the soil surface serve to detach soil particles. Some of the soil particles are transported by thin interrill flow into rills. Residue materials from the previous crop that are left as a surface mulch are very effective in reducing interrill erosion. Raindrop energy is absorbed and dissipated by the residue mulch, thus protecting the soil surface.

Rill erosion. Soil materials removed by raindrop impact on interrill areas may eventually be delivered to rills where they are transported down a hillslope by rill flow. Rills are small enough to be removed by normal tillage operations. Substantial erosion may occur once rills have formed. If conservation measures are not employed to reduce rill erosion, rapid loss of soil productivity may result. Rill erosion can usually be controlled by contouring, strip cropping, and conservation tillage.

Gully erosion. Gullies are deep channels larger than rills that cannot be removed by tillage. Gully erosion generally occurs near the upper end of intermittent streams. Once they are formed, gullies become a permanent part of the landscape, and they may expand rapidly. Gully formation often develops where there is a water overfall causing the gully to move

upslope. Water moving through the gully may cause the channel to deepen. In addition, sections of the exposed banks may be undercut and slide into the gully, where they are later removed during large runoff events. Control measures, such as terraces or vegetated waterways, may be required to prevent gully erosion.

Stream channel erosion. Stream channel erosion results from the removal of soil from stream banks or beds. Runoff flowing over the side of the stream bank or scouring below the water surface can cause stream channels to erode, especially during severe floods. The major cause of erosion along stream banks is meandering. Stream channel erosion may increase when sediment delivery from upland areas is reduced through control practices or when upstream sediments are caught in water storage facilities.

Factors affecting erosion. The principal elements affecting soil erosion are rainfall characteristics, soil factors, topography, climate, and land use.

Rainfall characteristics. Runoff is rainfall that is neither absorbed by the soil nor accumulated on the surface but moves downslope. Rainfall rate and duration are important variables influencing runoff and erosion. Runoff occurs only when rainfall intensity exceeds soil infiltration rate, which decreases with time. Thus, no runoff may occur from a storm of short duration, while substantial runoff may result from a storm of the same intensity but of longer duration.

Both the rate and volume of runoff are influenced by rainfall intensity. Infiltration capacity is exceeded by a greater margin during a high-intensity storm than a less intense rainfall event. As a result, the high-intensity storm may produce a greater volume of runoff even though total precipitation was similar for the two events. The infiltration rate may also be substantially reduced by the destructive action of the storm on the soil surface.

Irrigation is used on some agricultural areas. Runoff may result from both irrigation and natural precipitation events. The runoff potential may be compounded on irrigated areas because of the increased quantities of water introduced through irrigation.

Soil factors. The physical, chemical and mineralogical characteristics of soils vary greatly, as does their susceptibility to erosion. Soil erodibility is influenced significantly by the size of primary soil particles, organic matter content, soil structure, and permeability. These soil characteristics affect the susceptibility of soil particles and aggregates to detachment.

For erosion to occur, runoff must be present. In general, as runoff rates become greater, erosion also increases. One of the most effective means of reducing erosion is to maintain high infiltration rates. Keeping crop residue materials on the soil surface to reduce sealing caused by raindrop impact helps to preserve high infiltration rates.

Topography. The degree and length of slope, and the size and shape of the watershed influence erosion. As slope gradient increases, the velocity of flowing water becomes greater. The ability of moving water to detach and transport soil particles increases substantially with larger flow velocity.

Rill erosion becomes greater on longer slopes because of an increased accumulation of overland flow. Concave slopes, with a smaller slope gradient at the bottom of the hillslope, are less erosive than convex slopes. Deposition frequently occurs at the bottom of concave slopes because of reduced transport capacity of flow.

Climate. The quantity of erosion that occurs from a given region is influenced by the total amount and intensity of rainfall. The dense vegetation found on areas that receive substantial rainfall reduces erosion potential. Regions with low rainfall and limited vegetation are often susceptible to erosion during high-intensity rain storms.

Runoff from melted snow and ice can cause serious erosion problems in colder climates. During several months of the year, frozen soil is not subject to erosion. However, if the snow cover melts rapidly and infiltration does not take place, substantial runoff may result. The rapid melting that may occur when rain falls upon a snow-covered surface can also produce significant runoff. Substantial erosion may occur as water moves over a thin layer of freshly thawed soil. In many colder climates, more erosion results from snowmelt than from rainfall.

Land use. Areas having complete ground cover throughout the year are least susceptible to erosion. Erosion from undisturbed forests is usually minimal, because a constant vegetative cover is maintained on the soil surface. On croplands, the amount of surface cover is influenced by the cropping and management conditions employed. One of the most critical periods exists after planting when residue cover is at a minimum and high-intensity rains frequently occur.

A study conducted in the southeastern United States demonstrated the effects of selected land use on runoff and soil loss (**Table 2**). The results showed that cultivated land left fallow with no vegetative cover is particularly vulnerable to erosion. Row crops such as corn and cotton grown continually on areas with steep slopes may also result in significant erosion. Planting row crops in rotation with grasses and legumes that maintain a dense surface cover substantially reduces erosion.

TABLE 2. Mean annual runoff and erosion under different land uses with mean annual rainfall of 1400 mm (55 in.)*

Land use	Runoff, mm (in.)	Soil loss, Mg/ha† (ton acre)
Rotation; coastal bermuda grass	70 (2.8)	5 (2)
and crimson clover after corn	70 (2.8)	5 (2)
Corn grown continuously	180 (7.1)	27 (12)
Cotton grown continuously	250 (9.8)	49 (22)
Fallow; cultivated with no vegetative cover	470 (19)	135 (60)

*After J. R. Carreker et al., *Soil and Water Management Systems for Sloping Land*, USDA, ARS-S-160, 1978.
†1 Mg = 10^6 grams.

Interseeding row crops with a legume can also be an effective conservation measure on areas that receive sufficient precipitation. The legume provides a protective surface cover during the critical planting period and also serves as a supplemental nitrogen source for the following cropping season. A herbicide is usually applied to kill the legume before the row crop is planted.

The dense sod found in pastures grown in humid areas is very effective in reducing erosion. Erosion is also minimal on natural rangelands where adequate surface cover is maintained. In regions with limited rainfall where bunch grasses are found, severe erosion may occur during intense storms from the exposed soil located between the bunches of grass. Reduction of vegetative cover through excessive grazing may also result in serious erosion.

On some rangeland areas, gravel and cobble materials are found throughout the soil profile. Since they are not easily transported by overland flow, gravel and cobble materials remain on the surface of eroded soils. This creates an armoring process in which the gravel and cobble materials reduce further erosion from rangeland soils.

Erosion is greatly diminished on forest lands because of the overhead canopy of trees and the surface layer of decaying organic matter. In an undisturbed forest, almost all erosion occurs from channel banks or adjacent steep slopes. Erosion rates may increase substantially on forest areas that are disturbed by timber harvesting, road construction, or fires.

Control. Contouring, strip cropping, conservation tillage, terraces, buffer strips, and stream channel erosion control measures have been used effectively to reduce the damage caused by soil erosion.

Contouring. Planting crops and performing tillage along the contours of the land can be an effective conservation measure. Surface runoff is confined in small depressions, thus reducing rill development. Ridge tillage systems are used to significantly increase the storage capacity of furrows. To maintain furrow storage capacity, row crops are planted on the top of the same furrow each year. As the slope gradient increases, the effectiveness of ridges in trapping runoff and reducing soil loss decreases.

Strip cropping. Strip cropping occurs when alternate strips of different crops are grown in the same field. The strips with the greatest vegetative cover reduce runoff velocity and capture soil eroded from upslope areas. The strip widths selected allow for the convenient use of farm equipment. For erosion control, the strips are usually planted on the contour in a rotation that shifts crops annually from one strip to the next.

Conservation tillage. Leaving a residue mulch from the previous crop on the soil surface greatly reduces erosion (**Fig. 27**). The reduction is related to the percent of residue cover left on the soil surface. Even small amounts of residue cover can cause substantial reductions in erosion. The amount of erosion protection provided with a given percent of surface cover is influenced by the type of residue material.

Conservation tillage has been defined as any tillage or planting system that leaves at least 30% of the soil surface covered with residue after planting (Fig. 27). When tillage is performed, implements are used that cause only minimal disturbance to the soil surface, thus maintaining existing crop residues. For some row crops, such as soybeans, no tillage is used before planting to leave sufficient residue cover to control erosion.

Terraces. Terraces are broad channels built perpendicular to the slope of steep land. The gentle grades used in terraces allow runoff to be carried around a hill at relatively low velocities, causing sediment to settle from the runoff water. Terraces usually empty onto grassed waterways or into underground pipes, thus preventing the formation of gullies.

Crops are usually planted parallel to the terrace channel, requiring the use of contour farming. Conservation tillage is also frequently used in conjunction with terracing. A significant investment is required to construct terraces, and farming operations are more difficult on terraced hillslopes. As a result, terraces are used only when other control measures cannot provide adequate erosion protection.

Buffer strips. Buffer strips are areas of land maintained with permanent vegetation, designed to intercept runoff. They are most effective when used in combination with other erosion control practices. Buffer strips are located at various positions along the landscape as part of a planned conservation system. The types of vegetation used in buffer strips is influenced by local conditions. Periodic maintenance may be required for sustained buffer strip performance. The types of buffer strips frequently used are contour buffer strips, filter strips, and grassed waterways.

Contour buffer strips containing perennial grasses are planted along steep slopes. The grass strips

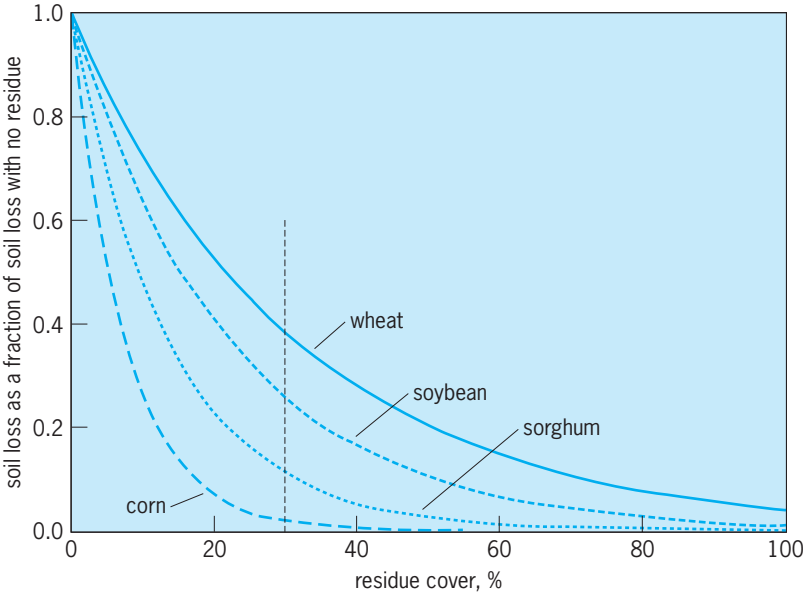


Fig. 27. Ratio of soil loss for given residue covers to soil loss with no cover. The vertical broken line indicates the residue cover necessary to be defined as conservation tillage. (*After T. S. Colvin and J. E. Gilley, Crop residue: Soil erosion combatant, Crops and Soils, 39(7):7–9, 1987*)

remove sediment from overland flow. The species of perennial grasses used and the spacing of the grass strips are tailored to local conditions. A narrow terrace may eventually form along the grass strip as a result of sedimentation. The expense of establishing contour buffer strips is substantially less than the cost for constructing terraces.

Filter strips can also be used to remove sediment from overland flow and provide increased infiltration. They are usually located on the edge of fields or adjacent to streams, ponds, or wetlands, and thus do not interfere significantly with normal farming operations. Filter strips are best suited for areas with gentle slopes where rilling is not a problem.

Grassed waterways can be used to convey runoff from terraces or other concentrated flow areas, and prevent channel erosion and gully formation. Sediment transported by overland flow is deposited in grassed waterways, thus reducing costly downstream sedimentation. A stable outlet below the grassed waterway is provided to reduce runoff velocity and disperse the flow before it enters a vegetated filter.

Stream channel erosion control measures. Vegetative, mechanical, or combined vegetative-mechanical means have been developed to reduce stream channel erosion. Depending on the size of the upstream drainage area, grading of the stream bank to a less severe slope may be necessary. Grass, shrubs, and trees have been successfully used to stabilize stream channels.

Dikes made of loose stone or rock piles are placed within the stream channel to divert the faster-flowing water away from the bank. Mechanical covers of stones, rocks, or other protective material may also be placed over the erodible bank. Areas with the greatest erosion hazard, such as the bottom of a stream bank, may be protected with a mechanical cover, while the upper portion of the stream banks is usually stabilized with vegetation. John E. Gilley

Wind erosion. Soil erosion by wind is a dynamic process. The results of wind erosion are evident when soil particles are dislodged from the soil surface, injected into the wind stream, and in some cases transported around the world as "dust" before being deposited. In the process of being dislodged by wind, soil particles are sorted according to size, like the winnowing of grain. Wind erosion is the geomorphological process responsible for the tremendous loess (wind-borne) deposits of highly fertile soil around the world. While wind erosion has always been an active process, human activities tend to accelerate it. As humans disturb large areas, loose soil particles may result. Large, coarse particles may move a few feet or be deposited in fence rows or road ditches at the edge of a field. When carried by strong winds, these coarse particles damage plants, abrade paint, break down soil crusts, and accelerate the wind erosion process. The loss of fine soil particles from a field reduces the capacity of the soil to produce crops; moreover, these particles reduce visibility and degrade air quality. *See* DUST STORM; LOESS.

When susceptible soils are exposed to erosive winds, ominous dust clouds occur. To effectively control the wind erosion that produces these dust clouds, all available resources, including the climate, soil, crop, and management systems, must be utilized. No single erosion control system will be equally effective for all wind erosion problem areas. For example, in regions that normally produce high-residue (vegetation) crops, the most effective wind erosion control systems maintain those residues either erect or on the soil surface. For semiarid regions growing crops that produce little residue, the most effective wind erosion control systems include a rough and cloddy soil surface and vegetative wind barriers. Options for humid or subhumid regions should include cover crops and residue management.

For all regions, if the erosive winds blow from the same direction, the benefits of residue management, soil roughness, or cover crops could be supplemented with strip cropping, annual or perennial crop wind barriers, and tree shelter belts. Farmers should be aware of all the wind erosion control practices that are available and then use whatever combination will be most effective for their conditions.

It is possible to estimate soil erosion with computer modes. Wind erosion models utilize weather, soil, crop, and management information in the calculation of predicted soil erosion losses. The losses can be estimated daily or for the entire crop-growing period.

Each erosion control strategy must consider the potential for rainwater to degrade soil surface roughness, temperature and rainfall requirements for growing cover crops, and the rate at which crop residues decompose.

Wind erosion damage can never be completely eliminated, but with careful planning and wise use of available resources (climate, soil, crop, and management) the impact of wind erosion on plants, soils, the atmosphere, and humans can be minimized. *See* EROSION; SOIL CONSERVATION. Donald Fryrear

Bibliography. P. W. Birkeland, *Soils and Geomorphology*, 1984; A. Blum, *Plant Breeding for Stress Environments*, 1988; N. C. Brady and R. R. Weil, *The Nature and Properties of Soils*, 11th ed., 1995; E. Bresler, B. L. McNeal, and D. L. Carter, *Saline and Sodic Soils*, 1982; H. D. Foth, *Fundamentals of Soil Science*, 8th ed., 1990; H. D. Foth, *Soil Fertility*, 1988; W. H. Gardner and J. C. Hsieh, *Water Movement in Soil*, 1960 (27-min time-lapsed video cassette, available in Soil Conservation Service and university audio-visual libraries, from Crop and Soil Sciences Club, Washington State University, P.O. Box 646420, Pullman, WA 99164-6420); C. T. Haan et al., *Design Hydrology and Sedimentology for Small Catchments*, Academic Press, San Diego, 1994; C. T. Haan, H. P. Johnson, and D. L. Brakensiek (eds.), *Hydrologic Modeling of Small Watersheds*, Amer. Soc. Agr. Eng. Monogr. 5, 1982; D. Hillel, *Environmental Soil Physics*, Academic Press, San Diego, 1998; N. Hudson, *Soil Conservation*, 3d ed., Iowa State University Press, Ames, 1995; E. H. Isaaks, *An Introduction to Applied Geostatistics*, 1990; H. Jenny,

The Soil Resource: Origin and Behavior, Ecological Studies, vol. 37, 1980; W. A. Jury, W. R. Gardner, and W. H. Gardner, *Soil Physics*, 5th ed., 1991; A. Klute (ed.), *Methods of Soil Analysis*, pt. 1: *Physical and Mineralogical Methods*, 2d ed., 1986; R. Lal, *Soil Erosion and Its Control*, 1988; A. Lieth and A. M. Massoun, *Towards Rational Use of High Salinity Tolerant Plants: Agriculture and Forestry under Marginal Soil Water Conditions*, 1993; L. Lyles, L. J. Hagan, and E. L. Skidmore, *Soil Conservation: Principles of Erosion by Wind*, Amer. Soc. Agron. Monogr. 23, 1983; National Academy of Sciences, *Soils of the Humid Tropics*, 1972; National Association of Review Appraisers Staff, *Land Classification for Land Uses Management and Valuation*, 1983; M. A. Nearing et al., A process-based soil erosion model for USDA: Water erosion prediction project technology, *Trans. Amer. Soc. Agr. Eng.*, 32(5):587–1593, 1989; F. J. Pierce and W. W. Frye (eds.), *Advances in Soil and Water Conservation*, Sleeping Bear Press, Chelsea, MI, 1998; K. G. Renard and G. R. Foster, *Soil Conservation: Principles of Erosion by Water*, Amer. Soc. Agron. Monogr. 23, 1983; G. Richardson and P. Mueller-Beilschmidt, *Winning with Water: Soil Moisture Monitoring for Efficient Irrigation*, 1988; G. O. Schwab et al., *Soil and Water Conservation Engineering*, 4th ed., John Wiley, Singapore, 1993; I. Shainberg and J. Shalhevet, *Soil Salinity under Irrigation: Processes and Management*, 1984; M. A. Sprague and G. B. Triplett (eds.), *No-Tillage and Surface Tillage Agriculture: The Tillage Revolution*, 1986; S. J. Stavarek and D. W. Rains, The development of tolerance to mineral stress, *HortScience*, 19:377–382, 1984; K. K. Tanji (ed.), *Agricultural Salinity Assessment and Management*, 1990; F. R. Troeh et al., *Soil and Water Conservation: Productivity and Environmental Protection*, 3d ed., Prentice Hall, Upper Saddle River, NJ, 1999; U.S. Department of Agriculture, *Soil: Yearbook of Agriculture*, 1957; U.S. Department of Agriculture Soil Conservation Service, *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 1988; W. H. Wischmeier and D. D. Smith, *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*, USDA Agr. Handb. 537, 1978.

Soil chemistry

The study of the composition and chemical properties of soil. Soil chemistry involves the detailed investigation of the nature of the solid matter from which soil is constituted and of the chemical processes that occur as a result of the action of hydrological, geological, and biological agents on the solid matter. Because of the broad diversity among soil components and the complexity of soil chemical processes, the application of concepts and methods employed in the chemistry of aqueous solutions, of amorphous and crystalline solids, and of solid surfaces is required. For a general discussion of the origin and classification of soils. *See* SOIL.

Elemental composition. The elemental composition of soil varies over a wide range, permitting only a few general statements to be made. Those soils that contain less than 12–20% organic carbon are termed mineral. (The exact percentage to consider in a specific case depends on drainage characteristics and clay content of the soil.) All other soils are termed organic. Carbon, oxygen, hydrogen, nitrogen, phosphorus, and sulfur are the most important constituents of organic soils and of soil organic matter in general. Carbon, oxygen, and hydrogen are most abundant; the content of nitrogen is often about one-tenth that of carbon, while the content of phosphorus or sulfur is usually less than one-fifth that of nitrogen. The number of organic compounds into which these elements are incorporated in soil is very large, and the elucidation of the chemistry of soil organic matter remains a challenging problem. *See* HUMUS.

Besides oxygen, the most abundant elements found in mineral soils are silicon, aluminum, and iron (**Table 1**). The distribution of chemical elements will vary considerably from soil to soil and, in general, will be different in a specific soil from the distribution of elements in the crustal rocks of the Earth. Often this difference may be understood in terms of pedogenic weathering processes and the chemical reactions that accompany them. Some examples are the accumulation of aluminum and iron in the Oxisols and of calcium in the Mollisols. The most important micro or trace elements in soil are boron, copper, manganese, molybdenum, and zinc, since these elements are essential in the nutrition of green plants. Also important are cobalt, which is essential in animal nutrition, and selenium, cadmium, and lead, which may accumulate to toxic levels in soil. The average natural distribution of trace elements in soil is not greatly different from that in crustal rocks (**Fig. 1**). This indicates that the total content of a trace element in soil usually reflects the content of that element in the soil parent material and, generally, that the trace element content of soil often is not affected directly by pedochemical processes.

The elemental composition of soil varies with depth below the land surface because of percolating water, chemical processes, and biological activity. The principal chemical processes are (1) hydrolysis (reaction with water to form products containing hydroxide ions), complexation (reaction of a metal with a ligand to form a product containing both), oxidation-reduction (changes in the oxidation state of an element), ion exchange (replacement of one ion by another on a solid surface), hydration (reaction with water to form a product containing water molecules), and flocculation-dispersion (settling-resuspension of solid particles in water). This last process affects soil particle removal by erosion or by translocation downward along soil pores. The principal effect of these processes is the appearance of illuvial horizons in which compounds such as aluminum and iron hydrous oxides, aluminosilicates, or calcium carbonate have been precipitated from

TABLE 1. Average percentages of the major elements and some microelements in subsurface soil clays and crustal rocks

Element	Soil order						
	Alfisol	Inceptisol	Mollisol	Oxisol	Spodosol	Ultisol	Crustal rocks
Silicon	19.20	24.69	23.01	12.43	5.79	16.02	27.72
Aluminum	12.38	19.61	10.29	19.33	15.86	17.49	8.20
Iron	8.04	3.81	6.83	10.83	3.29	11.96	4.10
Calcium	0.69	0.00	3.59	0.10	0.29	0.15	4.10
Magnesium	1.26	0.40	1.62	0.46	0.15	0.08	2.30
Sodium	0.18	2.52	0.04	0.00	0.27	0.06	2.30
Potassium	3.63	n.d.*	1.20	0.07	0.40	0.22	2.10
Titanium	0.40	0.28	0.44	1.32	0.16	0.50	0.56
Manganese	0.06	n.d.	0.06	0.08	0.06	0.05	0.10
Phosphorus	0.14	n.d.	0.14	0.27	0.17	0.12	0.10

*n.d. = no data.

solution or deposited from suspension. *See* WEATHERING PROCESSES.

Minerals. The minerals in soils are the products of physical, geochemical, and pedochemical weathering. Soil minerals may be either amorphous or crystalline. They may be classified further, approximately, as primary or secondary minerals, depending on whether they are inherited from parent rock or are produced by chemical weathering, respectively.

Primary minerals in soil. The bulk of the primary minerals that occur in soil are found in the silicate minerals, such as the olivines, garnets, pyroxenes, amphiboles, micas, feldspars, and quartz. The feldspars, micas, amphiboles, and pyroxenes commonly are hosts for trace elements that may be released slowly into the soil solution as weathering of these minerals continues. Chemical weathering of the silicate minerals is responsible for producing the most important secondary minerals in soil. *See* SILICATE MINERALS.

Secondary minerals in soil. The important secondary minerals that occur in soil are found in the clay fraction, sometimes in the form of coatings on other minerals. These include aluminum and iron hydrous oxides, carbonates, and aluminosilicates. The term allophane is applied to x-ray amorphous, hydrous aluminosilicates that are characterized by variable composition and a defect-riddled kaolinite structure containing aluminum in both tetrahedral and octahedral coordination. The significant crystalline aluminosilicates possess a layer structure; they are chlorite, halloysite, kaolinite, illite, smectite, and vermiculite. These clay minerals are identified in soil by means of the characteristic x-ray diffraction patterns they produce after certain pretreatments, although their positive identification may be difficult if two or more of the minerals are present at once. *See* CLAY MINERALS; X-RAY DIFFRACTION.

The distribution of secondary minerals varies among different soils and changes with depth below the surface of a given soil. However, under a leaching, well-oxidized environment, soil minerals do possess a differential susceptibility to decomposition, transformation, and disappearance from a soil profile. This has made possible the arrangement of the clay-sized soil minerals in the order of increasing resistance to chemical weathering known as the Jackson-Sherman weathering stages (**Table 2**). Minerals ranked near the top of the Jackson-Sherman list are present, therefore, in the clay fractions of slightly weathered soils (Entisols, Inceptisols, Aridisols, Gelisols), whereas minerals near the bottom of the list predominate in extensively weathered soils (Ultisols, Oxisols).

In zonal soils of humid-cool to subhumid-temperate regions, illite is the predominant clay mineral. Mixtures of kaolinite, vermiculite, and interstratified clay minerals are found in humid-temperate regions. In humid-warm regions, kaolinite, halloysite, allophane, gibbsite, and goethite are found. The mineralogical composition of the highly weathered and leached soils of the humid tropics is a subject of

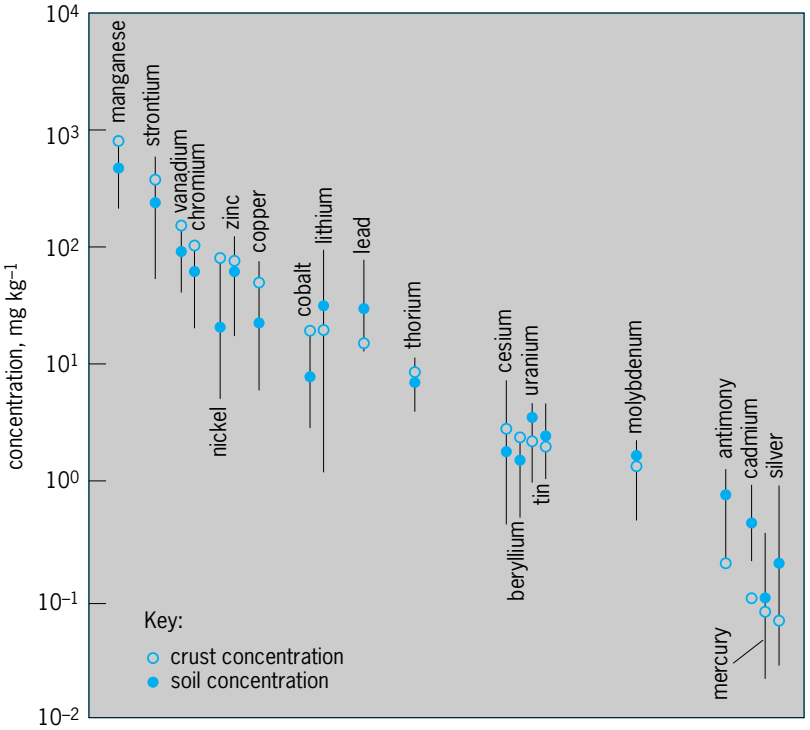


Fig. 1. Average concentrations and ranges of concentration of trace metals in surface soils compared with average crustal concentrations. (*After G. Sposito, Distribution of potentially hazardous trace metals, Metal Ions Biol. Sys., 20:1–20, 1986*)

TABLE 2. Jackson-Sherman weathering stages*

Characteristic minerals in soil clay fraction	Characteristic soil chemical and physical conditions
	Early Stage
Gypsum	Very low content of water and organic matter, very limited leaching
Carbonates	Reducing environments
Olivine/pyroxene/amphibole	Limited amount of time for weathering
Iron (II)-bearing micas	
Feldspars	
	Intermediate Stage
Quartz	Retention of sodium, potassium, calcium, magnesium, iron(II), and silica
Illite	Ineffective leaching and alkalinity
Vermiculite/chlorite	Igneous rock rich in calcium, magnesium, and iron(II), but no iron(II) oxides
Smectites	Silicates easily hydrolyzed
	Flocculation of silica
	Transport of silica into the weathering zone
	Advanced Stage
Kaolinite	Removal of sodium, potassium, calcium, and silica
Gibbsite	Effective leaching
Iron oxides (goethite, hematite)	Fresh-water oxidation of iron(II)
	Acidic compounds, low pH
Titanium oxides (anatase, rutile, ilmenite)	Dispersion of silica
	Aluminum-hydroxy polymers

*After M. L. Jackson and G. D. Sherman, Chemical weathering of minerals in soils, *Adv. Agron.,* 5:291–381, 1953, and M. L. Jackson, Clay transformations in soil genesis during the Quaternary, *Soil Sci.,* 99:15–22, 1965.

active investigation, in part because these soils (the Oxisols and Ultisols) are rapidly being converted to agricultural use by deforestation, with serious global change consequences. The soil minerals are dominated by iron and aluminum hydrous oxides, kaolinite, halloysite, and quartz. Weathering residues also are found in thin coatings on clay particle surfaces. Vermiculite and smectite with interlayer aluminum hydroxy polymers are common.

The chemical conditions favoring the genesis of kaolinite are the removal of the basic cations and Fe^{2+} by leaching, the addition of hydrogen ion (H^+) in fresh water, and a high aluminum-silicon (Al-Si) molar ratio. Smectite is favored by the retention of basic cations (arid conditions or poor drainage) and of silica. *See* GOETHITE; HALLOYSITE; ILLITE; KAOLINITE; MONTMORILLONITE; VERMICULITE.

Cation exchange. A portion of the chemical elements in soil is in the form of cations that are not components of inorganic salts but that can be replaced reversibly by the cations of leaching salt solutions. These cations are said to be readily exchangeable, and their total quantity, usually expressed in units of centimoles of proton charge [cmol(p^+)] per kilogram of dry soil, is termed the cation-exchange capacity (CEC) of the soil. The cation exchange capacity ordinarily is measured by leaching a known amount of soil with a salt solution, such as sodium acetate at pH 8.2, followed by an additional leaching with alcohol to remove residual salt, then determining the quantity of replacing cation [such as sodium (Na^+)] in the soil. However, the quantity of cations remaining after such treatment does not have a unique value characteristic of the soil alone, but depends as well on the concentration, the ionic composition, and the pH of the leaching solution. The cation exchange capacity of a soil generally will vary directly with the amounts of clay and organic matter present and with the distribution of clay minerals. *See* PH.

Soils that are less weathered because of recent origin, low precipitation, or temperate to cold climate have as exchangeable cations largely calcium (Ca^{2+}) and magnesium (Mg^{2+}). Some soils of dry areas contain significant amounts of exchangeable sodium. Extensively weathered soils, unless formed from basic parent material, have 20–95% of their exchangeable cations as aluminum. Prolonged leaching with fresh water supplies hydrogen (H^+) ions that eventually penetrate and disrupt the structures of soil aluminosilicates, thereby releasing aluminum cations, some of which remain in exchangeable form (**Table 3**).

The usual meaning of the term ion-exchange reaction in soil chemistry is the replacement of one adsorbed, readily exchangeable ion by another. On the molecular level, this means that ion exchange is a surface phenomenon involving charged species in outer-sphere complexes or in the diffuse-ion swarm. In practice, this conceptualization is adhered to only approximately. Cation-exchange reaction on soil humus, for example, include protons that may be adsorbed in inner-sphere surface complexes. Common extracting solutions for soil cation-exchange capacity measurements may displace metal cations from inner-sphere surface complexes as well as readily exchangeable metal cations. Careful experimental methods that will quantitate only readily

TABLE 3. Typical distributions of exchangeable cations in some soil orders*

Soil order (pH value)	CEC, $cmol(p^+) \cdot kg^{-1}$	Percentages				
		Calcium	Magnesium	Sodium	Potassium	Hydrogen + aluminum
Alfisol (5.4)	38	23	11	0	2	64
Aridisol (6.3)	4	52	26	5	17	0
Entisol (6.4)	5	37	23	<1	6	33
Inceptisol (5.3)	104	6	1	<1	<1	91
Mollisol (6.7)	25	67	12	<1	2	18
Spodosol (3.6)	106	5	6	<1	1	87
Oxisol (4.9)	27	31	8	<1	2	58
Ultisol (3.5)	24	11	3	<1	<1	84

*After R. L. Donahue, R. W. Miller, and J. C. Shickluna, *Soils,* Prentice-Hall, 1983.

exchangeable ions or will partition adsorbed ions accurately into readily exchangeable and specifically adsorbed species remain the focus of research. *See* CHEMICAL EQUILIBRIUM; ION EXCHANGE.

Isomorphic substitution in clay minerals. One of the most important sources of cation exchange capacity in soils is the negative charge that occurs on 2:1 clay mineral surfaces because of isomorphic substitution. The replacement of the silicon cation (Si^{4+}) in the tetrahedral sheet by aluminum cation (Al^{3+}) or of Al^{3+} in the octahedral sheet by magnesium cation (Mg^{2+}) and iron(II) cation (Fe^{2+}) results in a permanent, negative surface charge on the clay mineral surface that is more or less localized in hexagonal cavities formed by rings of oxygen atoms. This negative charge is balanced either by the formation of surface complexes between the cavities and cations taken from the soil solution or by swarms of cations remaining free in the soil solution. Extensive isomorphic substitution occurs in the tetrahedral sheet of vermiculite, with the result that potassium cation (K^+) can form inner-sphere surface complexes stabilized by the fact that this cation does not hydrate easily and fits almost perfectly into the hexagonal cavity in the clay mineral surface. On the other hand, isomorphic substitution occurs principally in the octahedral sheet of the smectite, montmorillonite, and the resultant electrostatic force binding a cation is weaker than for vermiculite because it must act through a greater distance. Cations such as calcium (Ca^{2+}) which hydrate strongly tend to form outer-sphere surface complexes with the hexagonal cavities in a 2:1 clay mineral surface.

Ionizable hydroxyl groups on clay minerals. At the edges of the structural layers in the crystalline clay minerals and on the surfaces of the amorphous clay minerals, hydroxyl (OH) groups may be found bonded to exposed Al^{3+} cations. These hydroxyl groups can act as Brönsted acids, ionizing in aqueous solution when the pH is sufficiently high. Thus a pH-dependent, negative surface charge can develop that will contribute to the cation exchange capacity. The mineral surfaces for which this pH dependence occurs can form surface complexes with cations. When this occurs, the complexed cation (for example, H^+, Na^+, or Ca^{2+}) is termed a potential-determining ion for the clay mineral surface, and its degree of surface complexation is governed by its electrochemical potential. The importance of pH-dependent charge versus permanent charge is greater in 1:1 layer silicates than in 2:1 layer silicates. For montmorillonite the pH-dependent cation exchange capacity may be 20% of the total cation exchange capacity; for illite it may be 40%; while for kaolinite and allophane it is essentially 100%.

Ionizable functional groups in organic matter. The organic matter in soil contains two types of functional groups that contribute importantly to the cation exchange capacity: aromatic and aliphatic carboxyls and phenolic hydroxyls. When the pH is greater than 4, these acidic groups ionize in aqueous solution and provide a source of negative charge for metal cation adsorption. The cations of metals in group 1 and 2 (for example, Na^+ and Ca^{2+}) of the periodic table are readily exchangeable after adsorption by soil organic matter, whereas those of the transition metals [for example, iron(III) (Fe^{3+}) and copper (Cu^{2+})] often are not. The contribution to the cation exchange capacity from organic functional groups can be large at high pH values.

Anion exchange. The stoichiometric exchange of the anions in soil for those in a leaching salt solution is a phenomenon limited to chloride and nitrate in the general scheme of anion reactions with soils. Under acid conditions (pH < 5) exposed hydroxyl groups at the edges of the structural sheets or on the surfaces of clay-sized particles become protonated and thereby acquire a positive charge. The degree of protonation is a sensitive function of pH, the ionic strength of the leaching solution, and the nature of the clay-sized particle. The magnitude of the anion-exchange capacity (AEC) usually varies from near 0 at pH 9 for any soil colloid to as much as 50 cmol(−)/kg of allophanic clay at about pH 4. [cmol(−)/kg means centimoles of negative charge per kilogram.] Smectite and other clay minerals with high, pH-invariant cation exchange capacity values do not usually adsorb exchangeable anions to any degree. Anion-exchange capacity may be measured conveniently by shaking a sample of soil for 1 h in sodium chloride (NaCl) solution at a chosen pH, filtering, and displacing the adsorbed chloride ion (Cl^-). Anion-exchange

capacity is then the number of cmoles of Cl displaced per kilogram of soil.

Negative anion adsorption. Soils whose cation exchange capacity is large often display a significant negative adsorption of anions: the concentration of anions in a solution contacting a soil suspension is larger than that of the anions in the liquid phase of the suspension. This phenomenon may be understood simply on the basis of the presence of negative charge on the surface of the solid-phase particles in the suspension. This surface charge attracts cations and repels anions. The principal effects of this repulsion are to reduce the anion exchange capacity and to increase the ease with which anions may be leached from a soil. Negative cation adsorption occurs in soils whose anion-exchange capacity is significant.

Specific anion adsorption and precipitation. Anion exchange in the classic sense applies primarily to halide and nitrate ions. For other ions, in particular, borate, molybdate, selenite, and orthophosphate, the reaction with the solid matter in soil involves specific adsorption or precipitation. The selenite ion (SeO_3^{2-}), for example, can react with the accessible aluminum hydroxy ions of clay minerals and with hydrous aluminum and iron oxides in soil to form inner-sphere surface complexes by ligand exchange with OH groups. The orthophosphate ion can react the same way or form x-ray amorphous analogs of the known crystalline aluminum and iron phosphate minerals by precipitation. The nature of the reaction is strongly dependent on pH, the metal cations in solution, the acidity of the added phosphate compound, and the structure of the solid phase with which the phosphate ion reacts. Under conditions of a relatively high pH, low acidity of the added phosphate, or high degree of crystallinity of the solid phase reactant, phosphate ion (PO_4^{3-}) will tend to be "specifically adsorbed" by Al or Fe ions at the surface of the solid through ligand exchange. If these conditions are reversed, the combination of PO_4^{3-} with the solid phase may result in a nearly complete destruction of the reactant solid and the formation of an amorphous aluminum or iron phosphate. This is an area of active research in soil chemistry.

Soil solution. The solution in the pore space of soil acquires its chemical properties through time-varying inputs and outputs of matter and energy that are mediated by the several parts of the hydrologic cycle and by processes originating in the biosphere (**Fig. 2**). The soil solution thus is a dynamic and open natural water system whose composition reflects the many reactions that can occur simultaneously between an aqueous solution and an assembly of mineral and organic solid phases that varies with both time and space. This type of complexity is not matched normally in any chemical laboratory experiment, but nonetheless must be amenable to analysis in terms of chemical principles. An understanding of the soil solution in terms of chemical properties has proven to be essential to progress in the maintenance of soil fertility and the quality of runoff and drainage waters.

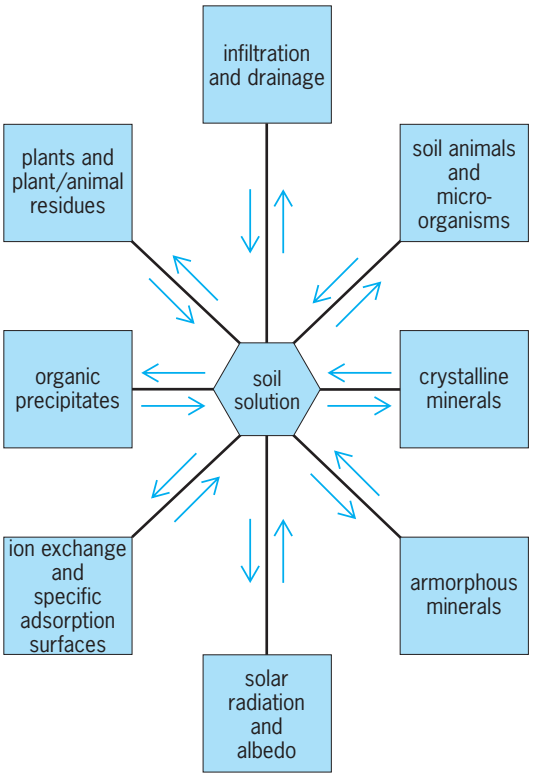


Fig. 2. Factors influencing the chemistry of a soil solution. (*After J. F. Hodson, Chemistry of the micronutrients in soils, Adv. Agron., 15:119–159, 1963*)

Chemical speciation of macrosolutes. The macrosolute composition of a soil solution will vary depending on pH, pE (negative common logarithm of the electron activity), organic matter content, input of chemical elements from the biosphere (including humans), and effectiveness of leaching. Under conditions of near-neutral pH, high pE, low organic matter content, no solute input from agriculture, and good but not excessive drainage, the expected macrosolutes are Ca^{2+}, K^+, Mg^{2+}, Na^+, Cl^-, biocarbonate ion (HCO_3^-), silicic acid [$Si(OH)_4$], and sulfate ion (SO_4^{2-}). If the pH is low, H^+ and Al^{3+} should be added to this list; if it is high, carbonate ion (CO_3^{2-}) should be added. If the soil has been fertilized, nitrate ion (NO_3^-) and hydrogen phosphate ion ($H_2PO_4^-$) become important. If the drainage is excessive, Al may be abundant and one or more of the solutes in the original list may be insignificant. If the drainage is poor and, therefore, the pE is low, SO_4^{2-} will be replaced by S^{2-} and CO_3^{2-} should be added. If the organic matter content is high, organic solutes become important. Combinations of these different environmental conditions will change the original list of solutes in still other ways (for example, low pE and nitrogen fertilizer addition could add NH_4^+ to the list).

The chemical speciation of macrosolutes (that is, their distribution among the free-ionic, complexed, precipitated, and adsorbed forms) depends on the nature of the solid matter in the soil, the composition of metals and ligands in solution and their concentration, and the pE value. Clearly the macrosolutes

themselves are interdependent in determining their speciation, and even the three factors just mentioned cannot be considered in complete isolation from one another. Nevertheless, it is possible to make some very broad statements about the macrochemical species to be expected by employing the principle of hard and soft acids and bases. *See* ACID AND BASE.

The macrosolute metal cations are hard acids. This means that they generally tend to form chemical bonds of a simple electrostatic type and, therefore, that their reactivities with any ligand (including those on a mineral surface) should be predictable on the basis of ionic charge and radius along with their common property of low polarizability (polarizability refers to the tendency of an ion to distort its electron cloud in the presence of an electric field). In a soil solution, the macrosolute metal cations will tend to (1) form complexes and sparingly soluble precipitates primarily with oxygen-containing ligands (hard bases), such as H_2O, OH^-, HCO_3^-, CO_3^{2-}, $H_2PO_4^-$, PO_4^{3-}, SO_4^{2-} and nonhydroxylated mineral surfaces: the stabilities of these complexes will tend to increase with the ratio of ionic charge to ionic radius of the metal; usually no reaction will occur with soft or nearly soft bases such as S^{2-} and Cl^-; (2) form complexes with carboxyl groups, but not with organic ligands containing only nitrogen and sulfur electron donors.

These generalizations make it possible to enumerate the probable complexes and precipitates of the macrosolute metals in a soil solution of known composition (including adsorbing surfaces). The exact speciation of the metals then can be calculated if stability constants are available for the expected chemical reactions. Usually a large set of nonlinear algebraic equations must be solved and ionic strength corrections must be performed.

Chemical speciation of microsolutes. The important microsolutes in soil include trace metals, such as iron, copper, and zinc, and trace element oxyanions, such as those formed by arsenic, boron, molybdenum, and selenium. The tableau of microsolutes in a given soil solution is more dependent on inputs from the lithosphere and the biosphere (particularly humans) and less on proton or electron activity and hydrologic factors than is the composition of macrosolutes. The trace metals present, for example, usually are derived from the chemical weathering of specific parent rocks, from the application of fertilizers, pesticides, and urban wastes, and from air pollution.

The most general features of the chemical speciation of the microsolutes also may be predicted on the basis of the principle of hard and soft acids and bases. The trace metal cations are soft or nearly soft acids, and the trace element anions are hard or nearly hard bases. The exception to this statement is Fe^{3+}, which is a hard acid. This means that the trace metal cations, except for Fe^{3+}, will tend to form chemical bonds of a covalent type whose strength will depend much more on detailed electronic structural considerations than on cationic size or charge. For the anions, the implication is that they will combine strongly with hard-acid metal cations, just as do other oxygen-containing ligands. Trace metal speciation thus presents a more complicated problem than does that of the macrosolute metals. Generally, the trace metal cations in a soil solution will (1) form complexes and insoluble precipitates more readily with the inorganic ligands Cl^- and S^{2-} and with sites on mineral surfaces that can bind covalently, than with oxyanions; stronger complexes also will form more readily with organic functional groups containing S, P, and N donors than with carboxyl groups; (2) tend to follow the Irving-Williams order in regard to the stabilities of strong complexes: $Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+} > Cd^{2+}$.

These broad predictions imply that trace-metal speciation in soil solutions will depend sensitively on the content and type of organic matter, the percentage of kaolinitic and amorphous hydrous oxide minerals, the pH, the pE, and the ionic strength. For example, a low solubility of the micronutrients Cu and Zn should occur for soils high in immobile organic matter (but not too low in pH) or for soil solutions with low pE values. Moreover, the solubilities of the trace metals should increase significantly with an increase in chloride concentration or in organic solutes. Since the complexes formed in these cases would reduce the ionic charge on the soluble trace metal species, a decrease in the affinity of a negative charge adsorbing surface (for example, that of montmorillonite) should also occur. Research has indicated that complex formation also decreases the bioavailability of a trace metal to a plant grown in soil. The uptake of a trace metal by a plant appears to be strongly correlated with the thermodynamic activity of the free metal cation species (**Fig. 3**).

Clay-organic complexes. The clay minerals in soils often are observed to be intimately associated with carbonaceous materials. These materials may be residues from plant or animal decomposition, herbicides or other pesticides, organic polymers and polyelectrolytes, surfactant compounds, or microbial metabolites. The complexes which they form

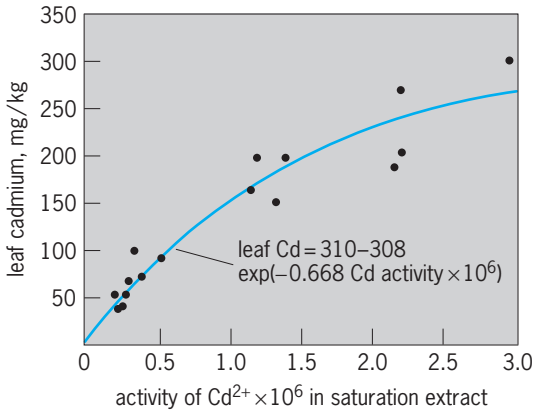


Fig. 3. Relationship between leaf concentration of a trace metal and its thermodynamic activity in the soil solution. The example shows Cd^{2+} uptake by Swiss chard, with the curve through the data represented by the equation. (*After F. T. Bingham, J. E. Strong, and G. Sposito, Influence of chloride salinity on cadmium uptake by Swiss chard, Soil Sci., 136:160–165, 1983*)

with clay minerals bear importantly on soil fertility [particularly in the rhizosphere (the near-root zone)], soil structure, soil moisture and aeration characteristics, the biological activity of organic compounds applied or disposed on soil, and the degradation of solid and liquid organic wastes in the soil environment. Generally, the organic component of a naturally occurring clay-organic complex will be of a very complicated nature that defies conclusive structural determination. Therefore, in order to obtain fundamental information about the mechanisms of bonding between clay minerals and organic matter in soil, a major line of research has involved the study of the reactions of known organic compounds with single types of clay minerals. On the basis of these studies, some important bonding mechanisms have been identified. They are expected to apply to the associations between clay minerals and organic matter in nature whenever the appropriate mineral species and organic functional groups are present.

Bonding mechanisms. The principal mechanisms through which organic compounds may bind to clay minerals have been elucidated largely by spectroscopic studies. They may be classified as follows:

1. Organic cation adsorption can occur, through protonated amine or carbonyl groups, onto any constant-charge clay mineral surface. The protonation of the functional groups may be either a pH effect or an acceptance of a proton that was formerly occupying an exchange site, was associated with a water molecule hydrating a metal cation, or was bound on another adsorbed organic cation. The affinity of an organic cation for a permanent-charge surface depends on molecular size, the nature of the functional groups present, and molecular configuration. Steric effects can be particularly significant because of the localized character of exchange sites and the quasirigid hydration envelope built up on a clay mineral surface. The stability of the water structure on a smectite surface is, in fact, great enough to require interstratified layers of either adsorbed metal cations or adsorbed organic cations when the clay mineral surfaces are only partially saturated with the organic compound. A mixture of the two types of cation in a single interlayer region disrupts the water structure too much to be stable.

2. Polar organic functional groups can bind to adsorbed cations through simple ion-dipole forces or complex formation involving covalent bonds (inner-sphere surface complex). The ion-dipole mechanism is to be expected, of course, for hard-acid metal cations, such as Ca^{2+} and Al^{3+}, whereas the formation of covalent bonds is to be expected for soft-acid metal cations, such as Cu^{2+}. As the organic functional groups often have soft-base character, the strength of binding by this mechanism should be greatest for exchangeable transition-metal cations. A sharp exception to this rule could occur with exposed Al^{3+} (or Fe^{3+}) in amorphous aluminosilicates that bind organic matter containing large numbers of carboxyl groups.

3. Large organic molecules can bind effectively to a clay mineral surface through hydrogen bonding. This bonding can involve a water bridge from a hydrated exchangeable cation to an oxygen-containing functional group (outer-sphere surface complex), a hydrogen bond from a more acidic functional group adsorbed directly on the clay mineral surface to a less acidic free one containing oxygen, or a direct hydrogen bond to a surface oxygen or hydroxyl plane in the clay mineral. If the exchangeable metal cation is a hard acid, the first type of bond will by far dominate the third type in importance. Direct hydrogen bonds to a plane of surface atoms would be accompanied by weaker dipole-dipole (that is, van der Waals) interactions, in general. This type of binding should be most important when very large organic molecules associate with a clay mineral surface containing relatively few exchange sites.

Catalysis reactions. Clay minerals have been shown often to catalyze reactions involving organic compounds. This catalytic function appears to be connected intimately with the presence of exchangeable metal cations and may be separated into two distinct types. The first type relates to the fact that the water molecules hydrating the exchangeable cations tend to dissociate protons very readily and, therefore, to endow the clay mineral surface with a pronounced acidity that increases markedly with desiccation. This enhanced proton-donating capability of the clay mineral, which will be greater the harder an acid the exchangeable cation is, serves a catalytic function in, for example, the surface protonation of amines and amino acids or in the abiotic degradation of pesticides.

A second type of catalytic function derives from the formation of inner-sphere organic complexes with exchangeable cations. This mechanism, which should be more significant the softer an acid the adsorbed metal cation is (again excepting Al^{3+} and Fe^{3+}), appears to play a basic role in, for example, the stabilization of humic compounds against degradation. These and other reactions catalyzed by clay minerals may prove to be very important in understanding how soil organic matter forms and how molecules of biological significance can be synthesized abiotically. *See* CATALYSIS; CHEMICAL BONDING; COORDINATION CHEMISTRY; COORDINATION COMPLEXES.

Garrison Sposito

Plant Nutrition

Soil chemistry controls the availability of plant nutrients within the soil and thus influences plant growth, yield, and nutritional value for human or animal consumption. The nutritional deficiencies or toxicities inherent to the morphology of the soil are known as mineral stress; they often represent a serious constraint for crop production and development. Globally, mineral stress occurs on about 23% of the world's soils. Plant nutritional problems are especially severe on highly weathered soils in tropical or subtropical geographic areas. This occurs because phosphate is made unavailable by iron and aluminum

oxides and because the clay minerals often have low cation exchange capacities.

The predominant soil chemical problems on well-aerated soils include acidity, salinity (high salt), and sodicity (high sodium). Important soil chemical processes include dissolution and precipitation of readily soluble minerals. The primary soil chemical problems in waterlogged soils include rapid disappearance of molecular oxygen, denitrification, and increased toxic concentrations of manganese and iron.

Nutrient movement. Soil chemical reactions occur continuously to maintain equilibria between available nutrients in soil solution and those existing as minerals within the soil or on exchange sites located on organic matter or clay minerals. Diffusion gradients (**Fig. 4***a*) and mass flow (Fig. 4*b*) are two primary mechanisms for nutrient movement to plant roots. Phosphorus and potassium move primarily by diffusion, but calcium and magnesium are usually supplied by mass flow. Although root interception (Fig. 4*c*) occurs, this process does not significantly contribute to plant nutrient accumulation.

Nutrient-specific effects. Carbon, hydrogen, and oxygen are supplied to plants by water and carbon dioxide (CO_2). Provided photosynthesis and transpiration are not limited, soil chemistry has very little effect on these nutrients.

Nitrogen (N) is assimilated from soil solution as ammonium (NH_4^+) or nitrate (NO_3^-) ions. Legumes also assimilate nitrogen through symbiotic fixation. The nitrogen cycle is biologically complex, but provided soils are not anaerobic and causing denitrification, soil chemistry has less affect on nitrogen than on a nutrient such as phosphorus. One exception is that the balance between NH_4^+ and NO_3^- forms can influence plant availability of nutrients such as magnesium. If NH_4^+ ions interfere with magnesium accumulation in plants, then availability of magnesium to grazing animals may be inadequate, leading to hypomagnesemia. *See* NITROGEN; NITROGEN CYCLE; NITROGEN FIXATION.

Phosphorus concentrations in soil solution are very low when compared with amounts present in organic matter and with amounts in distinct phosphate compounds such as monocalcium phosphate, dicalcium phosphate, octacalcium phosphate, hydroxyapatite, and fluorapatite. Phosphorus also exists as surface films coating soil aggregates and individual particles of sand, silt, and clay.

Phosphorus nutrition of plants is strongly affected by soil chemical conditions, especially soil reaction or pH (**Fig. 5**). Acid soils (pH < 5.0) have reduced availability of phosphorus because relatively insoluble iron and aluminum phosphate compounds are formed. In alkaline soils (pH > 7.0), insoluble calcium phosphate compounds, including octacalcium and hydroxyapatite, can substantially reduce available phosphorus. Primary geographic regions for alkaline phosphorus problems are the semiarid and arid regions, where free calcium carbonate ($CaCO_3$) is often found in the plant root zone. Phosphorus deficiencies associated with acid soil conditions are most prevalent in areas with high rainfall, including the southern United States as well as tropical or subtropical areas.

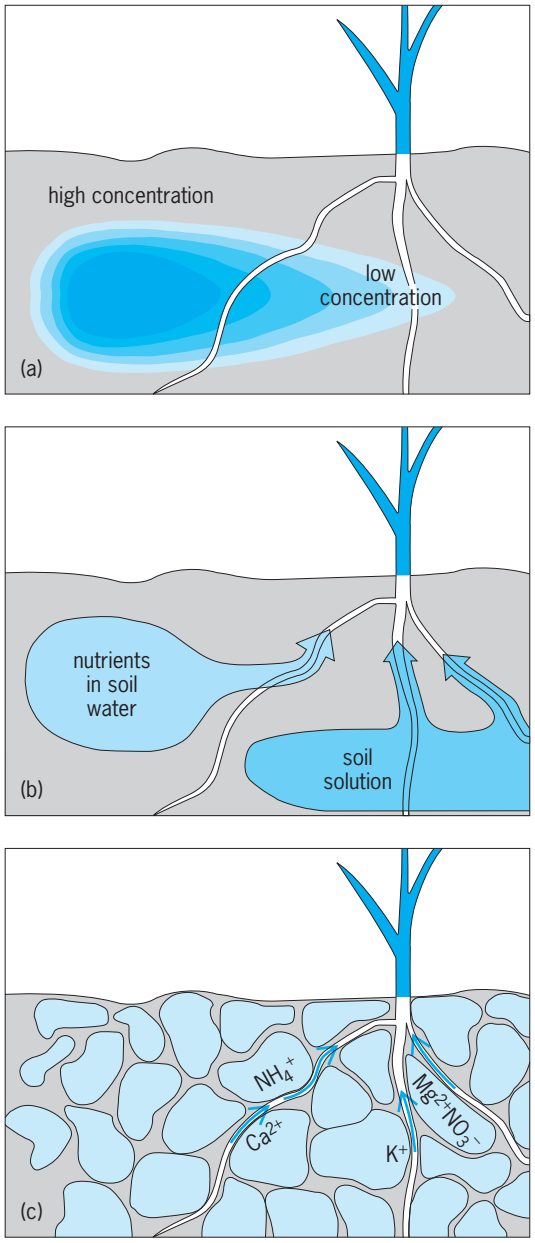


Fig. 4. Mechanisms of nutrient movement. (*a*) Diffusion; gradient movement. Fertilizer moves from high to low concentration. (*b*) Mass flow; water use controlled. (*c*) Root interception; contact required. NH_4^+ = ammonium ion; Ca^{2+} = calcium ion; K^+ = potassium ion; Mg^{2+} = magnesium ion; NO_3^- = nitrate ion.

Potassium is usually the predominant cation in plants, even though calcium is usually the predominant cation in soil. Deficiencies occur on all soils but are most prevalent on sandy soils that are low in organic matter, chalky soils, and soils derived from peat. Excess potassium can exist because of either natural mineral deposits or excessive additions of fertilizer or manure. Maintaining a balance among potassium, calcium, and magnesium is critical,

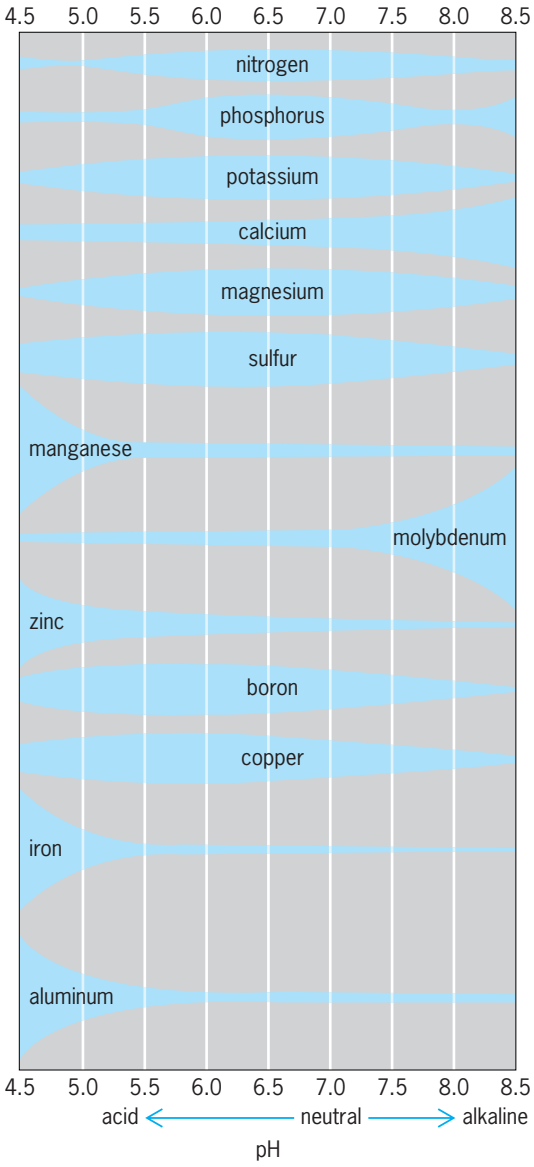


Fig. 5. Relationship of soil pH and plant nutrient availability, which is increased for the major plant nutrients when the soil pH is held between 6.5 and 7.0. Height of the tinted regions represents availability at a given pH level.

because imbalances can affect plant growth and development, and also can induce hypomagnesemia or grass tetany in cattle grazing forages on those soils.

Unless soils are extremely acidic, they can supply sufficient calcium for optimum plant growth. An exception is soils formed from serpentine minerals. Those soils often have a chemical imbalance between calcium and magnesium, and they have true calcium deficiencies. However, serpentine soils have many other chemical problems, including toxic levels of nickel and other heavy metals, as well as deficiencies of molybdenum.

Soil chemical processes generally supply magnesium in quantities sufficient to meet plant demand. Exceptions and areas where fertilizer response to magnesium may be found include sandy soils with low concentrations of magnesium, acid soils that have been treated with only calcitic limestone, and soils that have very high concentrations of potassium.

Saline and sodic soils. Saline, saline-sodic, and sodic soils are found in arid and semiarid regions where water loss through evaporation exceeds loss through leaching. This results in an accumulation of soluble salts, precipitated in the surface horizon and attached to cation exchange sites.

Saline soils contain enough chlorides and sulfates of sodium, calcium, and magnesium to interfere with growth of most plants. These are known as white-alkali soils, because the salts often give the soil light color. Saline-sodic soils also contain high amounts of neutral soluble salts and enough sodium to affect most plants seriously. These soils are most difficult to reclaim, because if they are leached with pure water, sodium ions will cause the clays to disperse and completely seal the soil surface. Reclamation must be done by using water with high concentrations of calcium or magnesium. Sodic soils do not contain large amounts of neutral soluble salts but have concentrations of sodium. They often have a very poor physical condition, because sodium disperses the clay particles. They are often known as black-alkali soils, because high concentrations of sodium carbonate (Na_2CO_3) disperse organic matter near the soil surface.

Micronutrients. Plant micronutrients are essential elements found in very low concentrations within plant tissues. These include boron, copper, iron, manganese, molybdenum, and zinc. Soil chemical processes that affect micronutrient availability include formation or dissolution of precipitates that contain these nutrients. Micronutrient availability is significantly influenced by soil pH (Fig. 5). Soil organic matter or microbial and root exudations also affect availability of micronutrients through formation of soluble coordination or chelation compounds. Any soil chemical change that influences formation of these compounds significantly affects micronutrient availability. *See* MYCORRHIZAE; RHIZOSPHERE.

Greenhouse culture. The principles of soil chemistry also apply to greenhouse growth media and potting mixes for home use. However, those media have physical and chemical properties that are distinctly different from natural soils. Since the mid-1990s, greenhouse operators have changed from using media containing soil to peat- or bark-based mixes containing manufactured materials such as perlite, vermiculite, or expanded polystyrene beads. These soilless growth media have good water-holding and aeration properties but limited nutrient holding capacities. As a result, fertility and plant nutrient management are more important than formerly.

Nutrient concentrations in the soil solution around the root are critical in such soilless growth media and depend upon the water-holding capacity of the medium. As water content increases, the nutrient concentration in solution decreases, because at complete saturation the water content is approximately four times the level at the permanent wilting point

and about two times that held against gravity. Total soluble nutrient concentration at saturation is one-fourth that at the permanent wilting point and one-half that at water held against gravity.

Developing a single set of fertilization guidelines is difficult because soilless growth media vary widely in bulk density (weight per unit volume). However, by analyzing a saturated media extract for pH, soluble salt content, and available nutrient levels, a balanced fertility program can be developed.

As with natural soils, growth-media pH influences the availability and uptake of all essential plant nutrients (Fig. 5). For peat-based media, however, the most desirable pH is 5.6–5.8 for most plants. If values exceed 6.5, micronutrient deficiencies may occur, and if the values are below 5.3, calcium or magnesium deficiencies or manganese toxicity may occur. All soluble ions or nutrients such as nitrate, ammonium, potassium, calcium, magnesium, chloride, and sulfate contribute to the soluble salt content of a growth media or water. The accumulation of these materials to toxic levels can be minimized by adding enough water to cause some leaching.

Whether in natural soil or soilless media, plants that have insufficient nutrients to maintain proper growth and development present several symptoms that can be used to help diagnose the primary problem. Beginning with the older leaves, plants deficient in nitrogen become light green in color. This occurs because the nitrate form of nitrogen is soluble and therefore mobile in the growth medium, so with watering, some may leach out. A fairly constant level

TABLE 4. Typical quantities of plant nutrients accumulated by harvested and nonharvested portions of various feed, food, and tree crops

Crop	Quantity, kg/ha*					
	Yield†	N‡	P	K	Mg	S
Corn grain	12,555	168	43	53	20	17
Stem, leaves, etc.		130	13	194	52	20
Wheat grain	5,380	100	22	25	13	6
Straw		50	5	126	13	17
Oat grain	3,590	90	12	19	6	9
Straw		40	7	116	17	12
Barley grain	5,380	123	20	32	9	11
Straw		45	7	107	10	11
Canola grain	2,350	74	16	15	—	13
Straw		44	7	62	—	10
Rice grain	7,850	86	23	26	9	6
Straw		39	7	112	7	8
Flax grain	1,255	45	8	14	—	3
Straw		16	2	28	—	3
Sorghum grain	8,970	134	30	28	16	25
Stem, leaves, etc.		146	15	158	34	18
Soybean grain	4,035	269§	24	78	19	13
Stem, leaves, etc.		94	8	54	11	14
Cotton lint + seed	1,680	105	19	41	12	8
Stalks, leaves, etc.		96	12	76	27	26
Sugarbeet roots	67,260	140	7	232	30	11
Tops		146	12	279	59	39
Sugarcane stalks	224,200	179	44	312	45	60
Tops and trash		224	32	255	67	36
Peanuts (nuts)	4,485	157§	11	32	6	11
Vines		112	8	140	22	12
Potatoes (tubers)	56,050	194	36	261	16	17
Vines		108	8	246	40	8
Sunflower seeds	3,925	140	30	36	13	7
Stems, leaves, etc.		57	5	83	34	11
Apples (fruit)	28,025	22	4	46	2	—
Leaves, new wood, etc.		90	19	121	25	—
Peaches (fruit)	33,630	39	5	60	13	—
Leaves, new wood, etc.		67	15	51	11	—
Oranges (fruit)	60,535	100	11	151	11	8
Leaves, new wood, etc.		196	16	152	31	24
Tomatoes (fruit)	67,260	112	11	187	9	24
Vines		90	12	112	22	22
Fababean (grain)	6,000	240	30	70	4	9
Field pea (grain)	6,000	187	29	75	8	16
Chickpea (grain)	1,500	91	6	50	11	9
Lentil (grain)	2,000	100	12	65	—	—
Alfalfa	17,935	504§	39	446	45	45
Tall fescue	7,850	170	32	172	14	—

*1 hectare = 10,000 square meters. Dividing by 10 gives values in grams per square meter.
†Yields are for the harvested portion of each crop (that is, grain for corn, wheat, oats, barley, canola, rice, flax, sorghum, soybean; lint + seed for cotton; roots for sugarbeet; stalks for sugarcane; leaves for tobacco; nuts for peanut; tubers for potatoes; seeds for sunflower; fruit for apples, peaches, oranges, tomatoes; and whole plant for alfalfa and tall fescue).
‡Accumulations given are total amounts taken up by the plants in both the harvested and the aboveground unharvested portions.
§Legumes such as soybean, peanut, and alfalfa can get most of their nitrogen from the air.

of available nitrogen can be maintained by injecting additional nitrogen into the watering system. An adequate phosphorus supply is important for root system development, rapid growth, and flower quality in floral plants. Deficient plants exhibit slow root and top growth.

The nutrient most often limiting in greenhouse fertility programs is potassium. The lower or oldest leaves of plants deficient in potassium show marginal yellowing or chlorosis. Many greenhouse plants have a potassium requirements equal to or greater than their nitrogen requirement.

Calcium availability for plant uptake is dependent on pH of the growth media and levels of other cations present, especially potassium and magnesium. Calcium deficiency results in abnormal growth or death of the growing tip.

Nutrient balance is just as important in soilless mixtures as in natural soils. The balance among potassium, calcium, and magnesium is especially important since these nutrients compete for similar uptake sites at plant root surfaces. Maintaining a balanced level of nutrients also enables plants to tolerate stresses such as low fertility or high levels of soluble salts. *See* PLANT MINERAL NUTRITION.

Douglas Karlen

Nutrient accumulation and removal. The relationship between soil chemistry and plant nutrition is evident when the elemental composition of plants is determined and the amount of each nutrient within the vegetative material (biomass) is computed [concentration of a specific element (mass per unit dry matter) times the amount of biomass (dry matter per unit area or container for greenhouse conditions)]. These determinations, made for plants grown in either soil or greenhouse potting media, are used in several ways. By measuring nutrient accumulation several times throughout a plant's life cycle, rates of nutrient accumulation can be computed for either the whole plant or for its parts (leaves, stalks, seeds). Nutrient accumulation measurements are used to determine how nutrient availability (supply), movement, and other factors (water, temperature, light) are affecting physiological processes within plants grown in a specific type of soil, potting media, or solution culture (hydroponics). Computing the quantity of nutrients accumulated by plants (**Table 4**) is useful for determining fertilizer needs, estimating nitrogen fixation, and predicting food or feed value. It is also useful for understanding nutrient cycling between plants, soil micro- and macrofauna (mites, collembola, ants, beatles, earthworms), soil flora (algae, fungi, bacteria), and the mineral fraction (sand, silt, and clay) of a soil or greenhouse media (which may contain inert materials as a portion of the mineral fraction). *See* BIOMASS.

Differences in the quantity of nutrients accumulated within vegetative (stems, stalks, straw, leaves, vines, or new wood) or reproductive (grain, seeds, or fruit) plant parts, especially nitrogen, phosphorus, and potassium, illustrate the various physiological roles that each element has in supporting plant growth and development. Quantities of nitrogen and phosphorus tend to be greater in the grain or reproductive parts where protein and energy reserves are stored, while potassium accumulation is generally much greater in the vegetative material where it plays a vital role in cellular physiology and stem strength. Total nitrogen accumulation differs in legumes (soybean, peanuts, alfalfa), corn (or sorghum), and grass crops (wheat, oat, barley, rice). Also, it is the relative quantity or amount of each nutrient accumulated by plants (Table 4), and not their physiological importance, that provides the rationale for classifying nitrogen, phosphorus, and potassium as macronutrients; magnesium, sulfur, and calcium as secondary nutrients; and boron, copper, iron, manganese, and zinc as micronutrients by agronomists, soil scientists, and other plant nutrition professionals. *See* BORON.

Predicting the feed or food value of plants for animal and human consumption is another reason for measuring nutrient accumulation. However, with regard to nutritional value, the quantity of nutrients within a plant available to the animal (including humans) consuming that plant will usually be substantially lower than the total nutrients accumulated by the plant. In addition to soil chemistry effects, postharvest handling can significantly change the bioavailability of nutrients, minerals, amino acids, and vitamins. Soil chemistry, plant chemistry, and external factors (water, heat, light stress) are complex interactions. The simple statement that healthy soils produce healthy plants which sustain healthy animals and people is very difficult to demonstrate quantitatively.

Worldwide, nutritional deficiencies (of calories and protein) are still among the most crucial human problems. Thus research and technology efforts focusing on the judicious use of fertilizers to increase food, feed, and fiber production, without negatively impacting water or air resources, remain important. *See* MALNUTRITION.

Douglas Karlen

Animal Nutrition

Mineral elements that are essential for ruminant animals, such as cattle and sheep, are commonly arranged into two groups: macronutrients and micronutrients (**Table 5**). The groups are distinguished by relative dietary requirements and amounts of mineral elements in animal tissues. For example, the macronutrients are required in the diet at concentrations greater than 100 micrograms per gram (parts per million), and the micronutrients, or trace elements, are required at concentrations less than 100 ppm. The elements calcium, phosphorus, magnesium, potassium, sodium, sulfur, and chlorine are considered macronutrients because tissue concentrations and dietary requirements of these essential elements are relatively high. In contrast, elements that generally occur at relatively low concentrations in living tissues and in diets are designated trace elements. The trace elements are chromium, cobalt, copper, iodine, iron, manganese, molybdenum, selenium, and zinc. Additionally, since about 1970, other trace elements have been shown to be required

TABLE 5. Critical concentrations of mineral elements in forages and dietary amounts necessary to meet requirements for normal growth and development of ruminant animals

Element	Critical concentration*		Dietary requirement†		Potentially toxic dietary concentration	
	Grass	Legume	Cattle	Sheep	Cattle	Sheep
Macronutrients						
	g/kg dry weight (*parts per thousand*)‡					
Calcium	2	18	4	2	20	20
Chlorine	0.3§	0.3§	2	1	4	4
Magnesium	1	3	2	1	6	6
Phosphorus	2.5	2.5	2	2	10	6
Potassium	25	25	8	3	30	30
Sodium	NR¶	NR¶	1.5	1	3	3
Sulfur	2	2.5	1.5	1.5	4	4
Micronutrients						
	μg/g dry weight (*parts per million*)					
Copper	4	11	8	6	100	25
Iodine	NR¶	NR¶	0.1	0.1	50	50
Iron	50	30	40	40	1000	1000
Manganese	20	25	25	40	1000	1000
Molybdenum	0.1	0.5	0.1	0.1	5	6
Selenium	NR¶	NR¶	0.1	0.1	5	3
Zinc	12	20	25	30	500	1000

* Concentration of nutrient required to be nonlimiting to growth.
† Dietary requirement for normal growth and development.
‡ 1 g/kg = 1 part per thousand = 0.1%.
§ Concentration not adequately defined.
¶ Not required for most higher plant species.

or beneficial to some species of animals. Trace elements in this group are arsenic, boron, lead, lithium, nickel, silicon, tin, and vanadium. The beneficial effects of many of the elements in this third group have been established mainly in laboratory animals under highly controlled environmental conditions. At present, there is no information to indicate that deficiencies have occurred in grazing ruminant animals under practical conditions. *See* AGRICULTURAL SCIENCE (ANIMAL).

Table 5 shows the mineral elements considered by most authorities to be essential to the nutritional health of cattle and sheep. Mineral requirements of animals are often expressed as amounts needed per day or as a portion (percentage or ppm) of the dry matter of the diet. In general, mineral requirements are affected by numerous animal factors, including species, breed, sex, age, weight, pregnancy, and stage of lactation. Within the body, mineral elements have three broad functions: (1) some elements, particularly trace elements, either are integral constituents of enzymes or activate specific enzymes; (2) some elements are important constituents of the structural components of organs and tissues; and (3) some elements are important components of fluids and tissues where they are concerned with maintaining osmotic pressure, pH balance, impulse conduction, and membrane permeability. **Table 6** shows some generalized concentrations of macrominerals in selected tissues and fluids of grazing animals. Total amounts and concentrations of mineral elements in animal tissues are variable and are influenced by many factors, including age, sex, hormones, nutrient intake, and nutrient interactions. Consequently, amounts of minerals reported for tissues may not represent all cases or all ages of animals.

Grasses, legumes, and forbs are the main sources of mineral nutrients consumed by grazing ruminants, such as cattle, sheep, and goats. Also, these animals may obtain some minerals from the water they drink and the soil they ingest, either passively (because of soil deposition on foliage) or actively (licking or eating the soil directly). Plants obtain most of their mineral nutrients through absorption from the soil, but some elements may be absorbed through the leaves after atmospheric deposition. Complex factors regulate the movement of mineral elements from rocks to soil, uptake of mineral nutrients by plants from the soil, transport of mineral nutrients from roots to other parts of the plant, and deposition of mineral nutrients in various components of the plant. Consequently, concentrations of mineral nutrients in plant foods and forages are highly variable and difficult to predict (**Fig. 6**). *See* SOIL FERTILITY.

The amount of an element required in plant tissues to prevent growth or yield retardation in the plant is called the critical concentration. Critical concentrations of elements may deviate between plant species and the values may vary, depending on the stage of growth and anatomical part of the plant that is sampled. Therefore, the values presented in Table 5 represent approximate amounts for numerous species and should not be used to

TABLE 6. Concentrations of mineral elements in selected tissues and fluids of animals*

Element	Tissue or fluid: Blood plasma	Muscle	Milk	Wool
Macrominerals				
	mg/100 ml	*mg/g dry wt*	*g/L*	
Calcium	9–12	0.2–0.6	1.2	
Chlorine	365	3–4	1	
Phosphorus	4–9	9–13	0.95	
Magnesium	1.8–3.2	1–1.5	0.12	
Potassium	15–16	15–17	1.4	
Sodium	325	3–5	0.5	
Micronutrients				
	μg/100 ml	*μg/g dry wt*	*μg/L*	*μg/g*
Cobalt	0.04–0.08	0.1	0.5	
Copper	90–120	4	40–200	4–140
Iodine	0.08–60	< 1	20–150	
Iron	90–270	200	20–600	50
Manganese	0.5	3	20–50	18
Molybdenum	1–5	0.1–5	18–120	1–15
Selenium	2–30	0.04–0.2	10	0.1
Zinc	70–120	80–240	3000–5000	80–220

*Concentrations of mineral elements in animal tissues are highly variable and may be affected by numerous factors, such as sex, age, hormones, dietary intake, and nutrient interactions.

compare individual plant species. Many of the minerals essential to animals occur at sufficient concentrations in practical feedstuffs and forages to meet the needs of cattle and sheep. However, in some circumstances sufficient concentrations of mineral elements may accumulate in grasses to support optimal growth of the plants, but still not provide sufficient amounts of some minerals to prevent potential deficiency problems in grazing animals. Moreover, some plants may accumulate quantities of mineral elements that may be toxic when consumed by grazing animals.

In order to establish the minimum dietary mineral concentrations necessary to meet the requirements of animals, it is necessary to have information on the bioavailability of mineral nutrients in various feeds and forages. Bioavailability refers to that proportion of the total amount of a mineral element present in

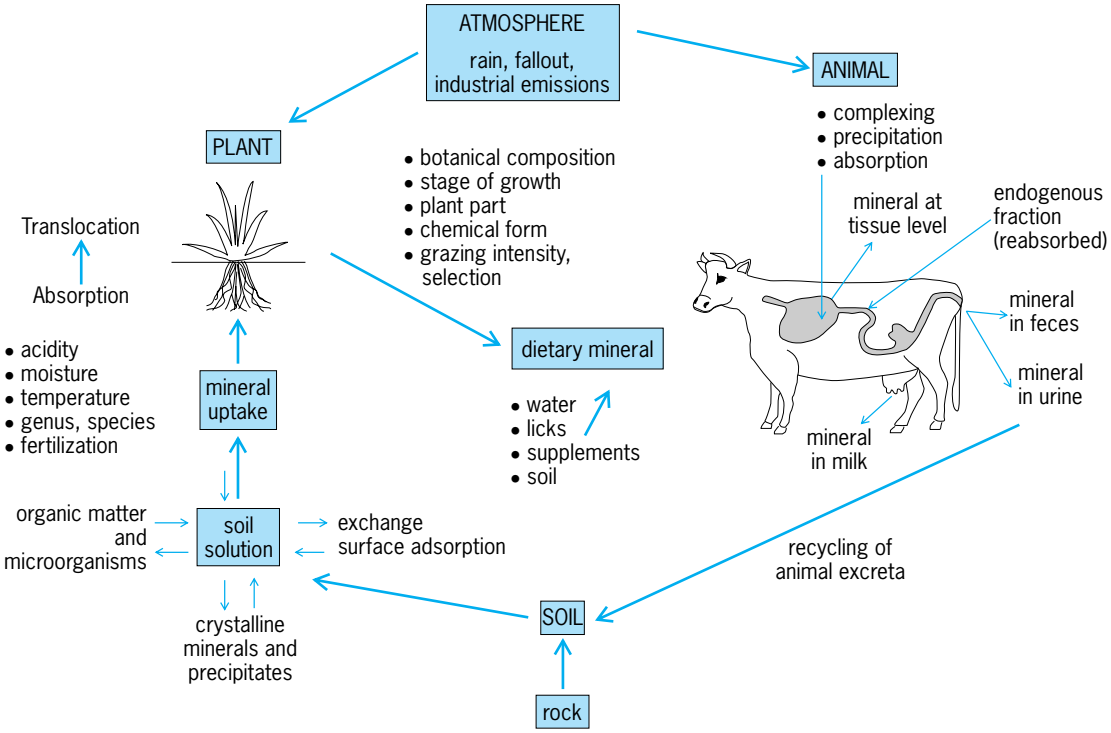


Fig. 6. Soil-plant-animal mineral cycle. Heavy arrows represent major pathways or shifts. Light arrows represent relatively minor pathways.

feeds or forages that is potentially absorbable in a metabolically active form. Moreover, as generally assessed, bioavailability is not a fundamental property of a feed or forage, but represents the response of a test animal to the diet or food. It would be very time-consuming and expensive to determine the bioavailability of mineral elements in all forages. Therefore, general estimates of mineral bioavailability, together with estimates of the amount of food consumed, have been used to approximate the minimal dietary concentrations of minerals necessary to meet the physiological needs of animals. Table 5 provides information on the amount of minerals required in the diet to meet the needs of growing animals. *See* ANIMAL GROWTH; NUTRITION.

Macronutrients. Macronutrients are the mineral elements required in ruminant animals' diet at concentrations greater than 100 μg/g.

Calcium. Calcium, the most plentiful mineral element in the body, is involved in numerous functions, such as blood clotting, bone structure, membrane permeability, muscle contraction, nerve impulse transmission, stability and activity of some enzymes, and secretion of specific hormones. Most of the calcium in the body (about 98%) occurs as part of the structural components of bone and teeth, and the remainder is found in extracellular fluids and soft tissues. The amount of calcium consumed from forage is affected by exchangeable calcium levels in soil, climatic conditions, plant species eaten, stage of plant growth, and portion of the plant consumed. *See* CALCIUM.

Phosphorus. Usually, phosphorus requirements and metabolism are considered with calcium because the two elements function together in bone. About 80–85% of the phosphorus in the body occurs in bone and teeth with the remainder in soft tissues. In addition to its role in bone structure, phosphorus is involved in several critical functions, including cell growth and differentiation, energy use, membrane structure, and maintenance of osmotic and pH balance. Phosphorus absorption is influenced by the amount and source of phosphorus consumed, the age of the animal, and dietary levels of fat, macrominerals (particularly calcium, magnesium, and potassium), and trace minerals (especially iron and manganese). In grazing animals, phosphorus deficiency may result in reduced growth, depraved or abnormal appetite, lowered reproduction, and impaired bone mineralization. A phosphorus deficiency may occur if the forage level is below 0.2%, a condition that may be found in tropical regions. *See* PHOSPHORUS.

Magnesium. Magnesium is involved in some important physiological functions, including maintenance of electrical potentials and transmission of neuromuscular impulses, structural integrity of bones and teeth, and activation of numerous enzyme systems. Also, as an ATP-magnesium complex, magnesium participates in all biosynthetic process. In cattle and sheep, about 60–70% of total body magnesium is in bone, 15% in muscle, 15% in other soft tissues, and 1% in extracellular fluids. In young animals, as much as 30% of the magnesium in bone can be mobilized during dietary magnesium deficiencies, but the value in adult animals is only about 2%. Grasses may have magnesium concentrations that are optimal for plant growth but lower than the requirements for grazing animals. *See* MAGNESIUM.

Potassium. Potassium, located mainly in intracellular fluids, affects osmotic pressure and pH balance within cells. Potassium also is involved in enzymatic reactions, nerve impulse transmission, muscle contraction, and transport of oxygen and carbon dioxide in blood. Secretion of potassium in milk increases the need for potassium by lactating animals, compared to growing animals. Forages usually contain sufficient potassium to meet the needs of grazing animals, but feed concentrates, such as corn silage, may not contain enough potassium to meet the needs of ruminant animals. Additionally, because potassium concentrations decline with increased maturity of forages, marginal potassium deficiencies, characterized by decreased food intake and reduced weight gain, may occur in animals grazing mature forages in winter or in drought periods. Severe potassium deficiencies are not likely to occur in grazing animals. *See* POTASSIUM.

Sodium and chlorine. In extracellular fluids, sodium is the major positively charged ion (cation) and chloride is the major negatively charged ion (anion). These elements function in controlling body fluid balance, regulating osmotic pressure, and maintaining pH balance. Also, sodium is involved in nerve impulse transmission, muscle contraction, and various transport processes. Chloride is necessary for the production of hydrochloric acid in gastric juice produced by the stomach and for the activation of the digestive enzyme amylase. Grazing beef cattle and sheep generally require less sodium than do lactating dairy cows because the cows secrete substantial amounts of sodium in milk. Sodium is not considered an essential nutrient for most forage species, and the sodium content of forages varies markedly. If sodium intake meets requirements, it is generally assumed that chlorine intake is sufficient because chlorine is usually associated with sodium. *See* SODIUM.

Sulfur. Sulfur is an integral component of several amino acids, particularly methionine, cysteine, and cystine, and it is part of the vitamins thiamin and biotin. Moreover, sulfur occurs in many compounds in the form of sulfate. One such compound, chondroitin sulfate, is an important part of blood vessels, bone, cartilage, and tendons. Moderate sulfur deficiency reduces food intake and lowers rates of protein synthesis by microorganisms in the rumen. When provided sufficient amounts of sulfur, the ruminal microorganisms can synthesize all of the sulfur-containing compounds required by the animal. Mature forages and forages grown in areas with sulfur-deficient soils may not provide enough sulfur to meet requirements. Dietary sulfur requirements are difficult to establish because of complex quantitative interactions between sulfur, copper, and molybdenum. Increasing the sulfur content of the diet lowers the absorption of copper. Additionally, sulfur

interacts with molybdenum to reduce copper retention. *See* SULFUR.

Micronutrients (trace elements). Micronutrients are the mineral elements required in ruminant animals' diet at concentrations less than 100 ppm.

Cobalt. Although cobalt is an indispensable micromineral, it is unique in that the requirement by most mammals is not for an ionic form of the element but for a preformed cobalt compound, vitamin B_{12}, also called cobalamin. Lack of this vitamin is the cause of signs of cobalt deficiency. Ruminant animals are unique in that vitamin B_{12} is synthesized by the microorganisms in the rumen. However, ruminal microorganisms are not able to synthesize sufficient quantities of vitamin B_{12} to meet the needs of the animal when dietary cobalt intake is inadequate, such as in areas of the world where soils are cobalt-deficient. In the United States, the sandy soils of the southeastern Coastal Plain and the leached soils derived from granite in New England are generally cobalt-deficient, and forages growing on these soils contain low amounts of cobalt. In lambs and calves, cobalt deficiency is likely to occur when forage concentrations of cobalt are below 0.1 ppm (dry-weight basis). Fertilizing pastures with cobalt or providing supplemental dietary cobalt are methods used to prevent cobalt deficiency in grazing ruminant animals. Establishing that certain pathological signs in grazing animals were attributable to cobalt deficiency is a well-documented example of a direct relationship between soil properties and animal health. *See* COBALT; VITAMIN.

Copper. Copper is an essential component of many enzymes. The amount of dietary copper required by grazing ruminants varies and is highly influenced by the amount of sulfur and molybdenum in the diet. Sulfur in dietary compounds may be reduced to sulfides by rumen microorganisms, and the sulfides may react with copper to form copper sulfide, a relatively insoluble compound that is not absorbed by the animal. Additionally, molybdenum in the diet may interact with sulfides in the rumen to form compounds called thiomolybdates. Then, copper may react with the thiomolybdates to form molybdenum-copper-sulfur compounds that are not absorbed. Moreover, some of the thiomolybdates may be absorbed, and further adversely affect copper metabolism in the body. The copper content of forages depends on the plant species and amount of available copper in the soil; grasses usually contain less copper than do legumes, such as clover. *See* COPPER.

Iron. Iron is an essential component of hemoglobin and myoglobin, the proteins involved in oxygen transport in blood and muscles, respectively. Within cells, iron is a component of compounds (cytochromes and proteins) that are involved in the use of oxygen to produce energy. Also, some enzymes contain iron or are activated by iron. For grazing ruminants, the dietary requirement for iron is about 50 ppm. Most forages and feedstuffs contain more than enough iron to meet the needs of grazing animals. Additionally, grazing animals may obtain iron from their drinking water or from eating soil. Excessively high intake of iron may depress the absorption of dietary copper. *See* IRON.

Manganese. Manganese is an essential component of some enzymes and an activator of other enzymes. Some of the metal activations are nonspecific; that is, for some enzyme activations, manganese can be replaced by other mineral ions, particularly magnesium. Manganese deficiency results in skeletal abnormalities in young animals, and affects reproduction in mature animals. The amount of manganese required for normal skeletal development is greater than the amount needed for maximum growth, but is less than the amount needed for optimal reproductive performance. Thus, the dietary manganese requirement for growing animals is about 20 ppm but that for breeding animals is about 40 ppm. Forage concentrations of manganese vary with plant species and are affected by soil acidity, soil moisture, and soil manganese content. Generally, forages contain sufficient amounts of manganese to meet the needs of grazing ruminants. *See* MANGANESE.

Molybdenum. Molybdenum is an essential component of several oxidase enzymes, but the dietary requirements for this element are not well established. Molybdenum requirements are low, and under practical grazing conditions there is no evidence that molybdenum deficiency occurs in grazing ruminant animals. However, in some cases supplemental molybdenum may increase ruminal microbial activity and thereby increase cellulose digestibility. Concentrations of molybdenum in forages vary greatly. Forages growing on moist (poorly drained) soils with high organic matter and neutral or alkaline pH are likely to have high concentrations of molybdenum. *See* MOLYBDENUM.

Selenium. Selenium is an indispensable part of an enzyme that protects cells from damage caused by some reactions that occur within cells. Additionally, selenium is an integral component of an enzyme responsible for converting thyroxine, a hormone that contains iodine, to its more active form, triiodothyronine. In grazing ruminant animals, selenium deficiency generally does not occur when the selenium concentration in forage is at least 0.1 ppm. However, there are regions in the United States and Canada where forages on range lands or in pastures contain less than adequate amounts of selenium because the selenium content of the soil is low. In contrast, there are other regions where soil selenium levels are high and plants growing on these seleniferous soils may accumulate very high amounts of selenium. Excessive consumption of these selenium accumulator plants may cause selenium toxicity. Maps produced by the U.S. Department of Agriculture show the regions where selenium deficiency and selenium toxicity are likely to affect the health of grazing ruminant animals. *See* SELENIUM.

Zinc. Zinc is an essential trace element because it is a component of a number of enzymes and because it is necessary for the activation of other enzymes. Also, zinc is important for the proper functioning of the immune system. The dietary zinc requirement for grazing ruminant animals has been set at 30 ppm.

However, the amount of zinc required by grazing ruminants is not as well defined as it is for other animals, partly because zinc status is difficult to assess. The amount of zinc in forages is affected by soil zinc content, soil acidity, stage of plant growth, and plant species. Grasses generally contain less zinc than do legumes, such as clover. Generally, the requirement for zinc in animal diets is higher than forage zinc concentrations necessary for normal plant growth. Forages in some areas, particularly those growing on calcareous soils with alkaline pH may not contain enough zinc to fully meet the needs of all grazing animals, and marginal or subclinical zinc deficiencies may occur. *See* ZINC.

William A. House

Bibliography. C. B. Ammerman, D. H. Baker and A. J. Lewis (eds.), *Bioavailability of Nutrients for Animals*, 1995; K. C. Beeson and G. Matrone, *The Soil Factor in Nutrition: Animal and Human*, 1976; C. A. Black, *Soil Fertility Evaluation and Control*, 1993; J. J. R. Frausto da Silva and R. J. P. Williams, *The Biological Chemistry of the Elements*, 1991; M. B. McBride, *Environmental Soil Chemistry*, 1994; H. Marschner, *Mineral Nutrition of Higher Plants*, 2d ed., 1995; L. R. McDowell, *Minerals in Animal and Human Nutrition*, 1992; K. Mengel and E. A. Kirkby, *Principles of Plant Nutrition*, 4th ed., 1987; F. M. M. Morel and J. G. Hering, *Principles and Applications of Aquatic Chemistry*, 1993; J. J. Mortvedt et al. (eds.), *Micronutrients in Agriculture*, 2d ed., Soil Science Society of America, Madison, WI, 1991; National Research Council (NRC), *Nutrient Requirements of Beef Cattle*, 7th ed., 1996; NRC, *Nutrient Requirements of Sheep*, 6th ed., 1985; NRC, *Nutrient Requirements of Dairy Cattle*, 6th ed., 1988; E. W. Russell, *Russell's Soil Conditions and Plant Growth*, 11th ed., 1988; D. L. Sparks, *Environmental Soil Chemistry*, 1995; G. Sposito, *The Chemistry of Soils*, 1989; W. Stumm, *Chemistry of the Soil-Water Interface*, 1992; R. L. Westerman (ed.), *Soil Testing and Plant Analysis*, 3d ed., Soil Science Society of America, Madison, WI, 1990.

Soil conservation

The practice of arresting or minimizing artificially accelerated soil deterioration. Its importance has grown because cultivation of soils for agricultural production, deforestation and forest cutting, grazing of natural range, and other disturbances of the natural cover and position of the soil have increased greatly since the middle of the nineteenth century in response to the growth in world population and human technical capacity. Accelerated soil deterioration has been the unfortunate consequence.

Erosion extent and intensity. Accelerated erosion has been known throughout history wherever people have tilled or grazed slopes or semiarid soils. Much evidence of accelerated erosion exists in the eastern and central Mediterranean basin, Iraq, China, and many other areas. Wherever the balance of nature is a delicate one, as on steep slopes in regions of intense rainstorms, such as those along the Caribbean coast of Venezuela, or in semiarid regions of high rainfall variability, grazing and cultivation eventually have had to contend with serious or disabling erosion. Irrigation works of the Tigris and Euphrates valleys are thought to have suffered from the sedimentation caused by quickened erosion on the rangelands of upstream areas in ancient times. Hill sections of Israel, Jordan, Syria, Greece, and central and southern Italy experienced serious and often virtually total soil loss down to solid bedrock from almost total deforestation, overgrazing by sheep and goats, and other land-use mismanagement many centuries ago. Accelerated water erosion on the hills of southern China and wind erosion in northwestern China also date far back. The economic and social effects of these soil movements have been much debated, but their impact appears to have been serious in many cases, such as that of classical Rome and that of China several centuries ago.

The exact extent of accelerated soil erosion in the world today is not known, particularly as far as the rate of soil movement is concerned. However, it may be safely said that nearly every semiarid area with cultivation or long-continued grazing, every hill land with moderate to dense settlement in humid middle latitude and subtropical climates, and all cultivated or grazed hill lands in the Mediterranean climate areas suffer to some degree from such erosion. These recognized and often growing problems of erosion occur in such culturally diverse areas as southern China, the Indian plateau, South Africa, Russia and Ukraine, and Spain.

Within the United States the most critical areas have been the hill lands of the interior Southeast, the Great Plains, the Palouse area hills of the Pacific Northwest, southern California hills, and slope lands of the Midwest. The high-intensity rainstorms of the Southeast and the cyclical droughts of the Plains have predisposed these two larger areas to erosion. The light-textured A horizon formed under the Plains grass cover was particularly susceptible to wind removal, while the high clay content of many southeastern soils predisposed them to water movement. These natural susceptibilities were repeatedly brought into play by agricultural systems which stressed corn, cotton, and tobacco in the Southeast, corn in the Midwest, and intensive grazing and small grains on the Plains, the Palouse, and in California. The bare soil surface left by the traditional intertilled cultivation of the Southeast furnished almost ideal conditions for water erosion and at the same time caused heavy nutrient depletion of soils thus cropped. The bare fields during seasons between crops have also been susceptible to soil depletion. Bare fields have been especially disastrous to maintenance of soil cover during droughts of the Plains. Soil mismanagement thus has been a common occurrence in parts of the United States where stability of soil cover hangs in delicate balance. *See* EROSION.

Types of soil deterioration. Soil may deteriorate either by physical movement of soil particles from a given site or by depletion of the water-soluble

Fig. 1. Erosion of sandstone caused by strong wind and occasional hard rain in an arid region. (*USDA*)

elements in the soil which contribute to the nourishment of crop plants, grasses, trees, and other economically usable vegetation. The physical movement generally is referred to as erosion. Wind, water, glacial ice, animals, and tools in use may be agents of erosion. For purposes of soil conservation, the two most important agents of erosion are wind and water, especially as their effects are intensified by the disturbance of natural cover or soil position. Water erosion always implies the movement of soil downgrade from its original site. Eroded sediments may be deposited relatively close to their original location, or they may be moved all the way to a final resting place on the ocean floor. Wind erosion, on the other hand, may move sediments in any direction, depositing them quite without regard to surface configuration. Both processes, along with erosion by glacial ice, are part of the normal physiographic (or geologic) processes which are continuously acting upon the surface of the Earth. The action of both wind and water is vividly illustrated in the scenery of arid regions (**Fig. 1**). Soil conservation is not so much concerned with these normal processes as with the new force given to them by human land-use practices. *See* LAND-USE PLANNING.

Accelerated erosion may be induced by any land-use practice which denudes the soil surfaces of vegetative cover (**Fig. 2**). If the soil is to be moved by water, it must be on a slope. The cultivation of a corn or a cotton field is a clear example of such a practice. Corn and cotton are row crops; cultivation of any row crop on a slope without soil-conserving practices is an invitation to accelerated erosion. Cultivation of other crops, like the small grains, also may induce accelerated erosion, especially where fields

Fig. 2. Improper land use. Corn rows planted up and down the slope rather than on the contour. Note better growth of plants in bottom (deeper) soil in foreground as compared to stunted growth of plants on slope. (*USDA*)

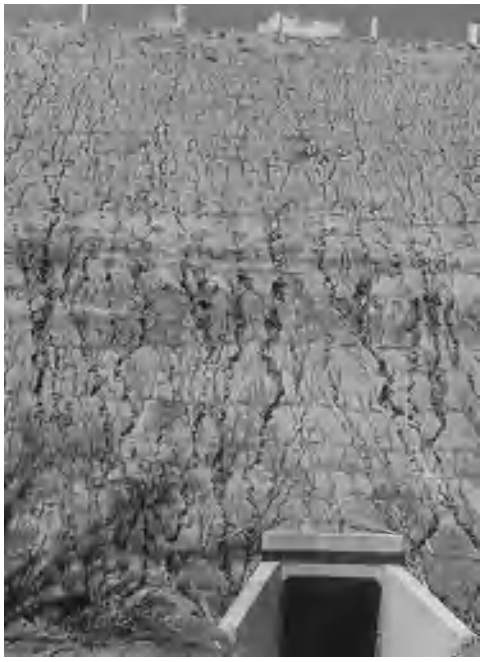


Fig. 3. Rill erosion on highway fill. The slopes have been seeded (horizontal lines) with annual lespedeza to bind and stabilize soil. (*USDA*)

are kept bare between crops to store moisture. Forest cutting, overgrazing, grading for highway use, urban land use, or preparation for other large-scale engineering works also may speed the natural erosion of soil (**Fig. 3**).

Where and when the soil surface is denuded, the movement of soil particles may proceed through splash erosion, sheet erosion, rill erosion, gullying, and wind movement (**Fig. 4**). Splash erosion is the minute displacement of surface particles caused by the impact of falling rain. Sheet erosion is the gradual downslope migration of surface particles, partly with the aid of splash, but not in any defined rill or channel. Rills are tiny channels formed where small amounts of water concentrate in flow. Gullies are V- or U-shaped channels of varying depths and sizes. A gully is formed where water concentrates in a rivulet or larger stream during periods of storm. It may be linear or dendritic (branched) in pattern, and with the right slope and soil conditions may reach depths of 50 ft (15 m) or more. Gullying is the most serious form of water erosion because of the sharp physical change it causes in the contour of the land, and because of its nearly complete removal of the soil cover in all horizons. On the edges of the more permanent stream channels, bank erosion is another form of soil movement.

Depletion of soil nutrients obviously is a part of soil erosion. However, such depletion may take place in the absence of any noticeable amount of erosion. The disappearance of naturally stored nitrogen, potash, phosphate, and some trace elements from the soil also affects the usability of the soil for human purposes. The natural fertility of virgin soils always is depleted over time as cultivation continues, but the rate of depletion is highly dependent on management practices. *See* PLANT MINERAL NUTRITION; SOIL; SOIL FERTILITY.

Causes of soil mismanagement. One of the chief causes of erosion-inducing agricultural practices in the United States has been ignorance of their consequences. The cultivation methods of the settlers of western European stock who set the pattern of land use in the United States came from physical environments which, because of the mild nature of their rainstorms and their prevailing soil textures, were in general far less susceptible to erosion than were the environments into which they migrated in North America. The principal European-sown grains, moreover, were wheat, barley, oats, and rye which covered the soil surface completely as they grew. Corn, cotton, and tobacco, however, were crops unfamiliar to European agriculture. In eastern North America the combination of European cultivation methods and American intertilled crops resulted in generations of soil mismanagement. In later years the Plains environment, with its alternation of drought

Fig. 4. Two most serious types of erosion. (*a*) Sheet erosion as a result of downhill straight-row cultivation. Note onions washed completely out of ground. (*b*) Gully erosion destroying rich farmland and threatening highway. (*USDA*)

Fig. 5. Wind erosion. The accumulation of topsoil was blown from the bare field on right. (*USDA*)

and plentiful moisture, was also an unfamiliar one to settlers from western Europe.

Conservational methods of land use were slow to develop, and mismanagement was tolerated because of the abundance of land in the eighteenth and nineteenth centuries. One of the cheapest methods of obtaining soil nutrients for crops was to move on to another farm or to another region. Until the twentieth century, land in the United States was cheap, and for a period it could be obtained by merely giving assurance of settlement and cultivation. With low capital investments, many farmers had little stimulus to look upon their land as a vehicle for permanent production. Following the Civil War, tenant cultivators and sharecroppers presented another type of situation in the Southeast, where stimulus toward conservational soil management was lacking. Management of millions of acres of farmland in the Southeast was left in the hands of people who had no security in their occupancy, who often were illiterate, and whose terms of tenancy and meager training forced them to concentrate on corn, cotton, and tobacco as crops.

On the Plains and in other susceptible western areas, small-grain monoculture, particularly of wheat, encouraged the exposure of the uncovered soil surface so much of the time that water and wind inevitably took their toll (**Fig. 5**). On rangelands, the high percentage of public range (for whose management little individual responsibility could be felt), lack of knowledge as to the precipitation cycle and range capacity, and the desire to maximize profits every year contributed to a slower but equally sure denudation of cover. *See* AGRICULTURAL SOIL AND CROP PRACTICES.

Finally, the United States has experienced extensive erosion in mountain areas because of forest mismanagement. Clearcutting of steep slopes, forest burning for grazing purposes, inadequate fire protection, and shifting cultivation of forest lands have allowed vast quantities of soil to wash out of the slope sites where they could have produced timber and other forest values indefinitely. In the United States the central and southern Appalachian area and the southern part of California have suffered severely in this respect, but all hill or mountain forest areas except the Pacific Northwest have had such losses. *See* FOREST MANAGEMENT.

Economic and social consequences. Where the geographical incidence of soil erosion has been extensive, the damages have been of the deepest social consequence. Advanced stages of erosion may remove all soil and therefore all capacity for production. More frequently it removes the most productive layers of the soil—those having the highest capacity for retention of moisture, the highest soil nutrient content, and the most ready response to artificial fertilization. Where gullying or dune formation takes place, erosion may make cultivation physically difficult or impossible. Thus, depending on extent, accelerated erosion may affect productivity over a wide area. At its worst, it may cause the total disappearance of productivity, as on the now bare limestone slopes of many Mediterranean mountains. At the other extreme may be the slight depression of crop yields which may follow the progress of sheet erosion over short periods. In the case of forest soil losses, except where the entire soil cover disappears, the effects may not be felt for decades, corresponding to the growth cycle of given tree species. Agriculturally, however, losses are apt to be felt within a matter of a few years.

Moderate to slight erosion cannot be regarded as having serious social consequences, except over many decades. As an income depressant, however, it does prevent a community from reaching full productive potentiality. More severe erosion has led to very damaging social dislocation. For those who choose to remain in an eroding area or who do not have the capacity to move, or for whom migration may be politically impossible, the course of events is fateful. Declining income leads to less means to

cope with farming problems, to poor nutrition and poor health, and finally to family existence at the subsistence level. Communities made up of a high proportion of such families do not have the capacity to support public services, even elementary education. Unless the cycle is broken by outside financial and technical assistance or by the discovery of other resources, the end is a subsistence community whose numbers decline as the capacity of the land is further reduced under the impact of subsistence cultivation. This has been illustrated in the hill and mountain lands of the southeastern United States, in Italy, Greece, Mexico, China, and elsewhere for many millions of farming people. Illiteracy, short lifespans, nutritional and other disease prevalence, poor communications, and isolation from the rest of the world have been the marks of such communities. Where they are politically related to weak national governments, indefinite stagnation and decline may be forecast. Where they are part of a vigorous political system, their rehabilitation can be accomplished only through extensive investment contributed by the nation at large. In the absence of rehabilitation, these communities may constitute a continued financial drain on the nation for social services such as education, public health, roads, and other public needs.

Effects on other resources. Accelerated erosion may have consequences which reach far beyond the lands on which the erosion takes place and the communities associated with them. During periods of heavy wind erosion, for example, the dust fall may be of economic importance over a wide area beyond that from which the soil cover has been removed. The most pervasive and widespread effects, however, are those associated with water erosion. Removal of upstream cover changes the regimen of streams below the eroding area. Low flows are likely to be lower and their period longer where upper watersheds are denuded than where normal vegetative cover exists. Whereas flood crests are not necessarily higher in eroding areas, damages may be heightened in the valleys below eroding watersheds because of the increased deposition of sediment of different sizes, the rapid lifting of channels above floodplains, and the choking of irrigation canals.

A long chain of other effects also ensues. Because of the extremes of low water in denuded areas during dry seasons, water transportation is made difficult or impossible without regulation, fish and wildlife support is endangered or disappears, the capacity of streams to carry sewage and other wastes safely may be seriously reduced, recreational values are destroyed, and run-of-the-river hydroelectric generation reaches a very low level. Artificial storage becomes necessary to derive the services from water which are economically possible and needed. But even the possibilities of storage eventually may disappear when erosion of upper watersheds continues. Reservoirs may be filled with the moving sediment and lose their capacity to reduce flood crests, store flood waters, and augment low flows. A dramatic example of such ongoing storage reduction with ominous implications for the future is Lake Nasser, one of the largest artificial lakes in the world, impounded by the great dam on the Nile River at Aswan, Egypt. For this reason, plans for permanent water regulation in a given river basin must always include watershed treatment where eroding lands are in evidence. *See* WATER CONSERVATION.

Conservation measures and technology. Measures of soil management designed to reduce the effects of accelerated erosion have been known in both the Western world and in the Far East since long before the beginning of the Christian Era. The value of forests for watershed protection was known in China at least 10 centuries ago. The most important of the ancient measures on agricultural lands was terrace construction, although actual physical restoration of soil to original sites also has been practiced. Terrace construction in the Mediterranean countries, China, Japan, and the Philippines represents the most impressive remaking of the face of the Earth before the days of modern earth-moving equipment (**Fig. 6**). Certain land management practices that were soil-conserving have been a part of western European agriculture for centuries, principally those centering on livestock husbandry and crop rotation. Conservational management of the soil was known in colonial Virginia and by Thomas Jefferson and others during the early years of the United States. However, it is principally since 1920 that techniques for soil conservation have been developed for many types of environment in terms of an integrated approach. The measures include farm, range, and forest management practices, and the building of engineered structures on land and in stream channels.

A first and most important step in conservational management is the determination of land capability—the type of land use and economic production to which a plot is suited by slope, soil type,

Fig. 6. Ifugao rice terraces, Philippines. (*Philippine Embassy, Washington, DC*)

Fig. 7. Land capability classes. Suitable for cultivation: 1, requires good soil management practices only; 2, moderate conservation practices necessary; 3, intensive conservation practices necessary; 4, perennial vegetation—infrequent cultivation. Unsuitable for cultivation (pasture, hay, woodland, and wildlife): 5, no restrictions in use; 6, moderate restrictions in use; 7, severe restrictions in use; 8, best suited for wildlife and recreation. (*USDA*)

drainage, precipitation, wind exposure, and other natural attributes. The objective of such determination is to achieve permanent productive use as nearly as possible. The U.S. Natural Resources Conservation Service has developed a widely employed system for such determination (**Fig. 7**). In this system, eight classes of land are recognized within United States territory. Four classes represent land suited to cultivation, from the class 1 flat or nearly flat land suited to unrestricted cultivation, to the steeper or eroded class 4 land that must be cultivated only infrequently and with extreme care. Three additional classes are grazing or forestry land, with varying degrees of restriction on use. The eighth class is suited only to watershed, recreation, or wildlife support. The aim in the United States has been to map all lands from field study of their capabilities, and to adjust land use to the indicated capability as it becomes economically possible for the farm, range, or forest operator to put conservational use into force.

Once the capability of land has been determined, specific measures of management come into play. For class 1 land, few special practices are necessary. After the natural soil nutrient minerals begin to decline under cultivation, the addition of organic or inorganic fertilizers becomes necessary. The return of organic wastes, such as manure, to the soil is also required to maintain favorable texture and optimum moisture-holding capacity. Beyond these measures, little need be added to the normal operation of cultivation.

On class 2, 3, and 4 lands, artificial fertilization will be required, but special measures of conservational management must be added. The physical conservation ideal is the maintenance of such land under cover for as much of the time as possible. This can be done where pasture and forage crops are suited to the farm economy. However, continuous cover often is neither economically desirable nor possible. Consequently, a variety of devices have been invented to minimize the erosional results from tillage and small grain or row crop growth. Tillage itself has become an increasingly important conservational measure since it can affect the relative degree of moisture infiltration and soil grain aggregation, and therefore erosion. Moreover, forms of conservation tillage that disturb the soil surface as little as possible have become very significant. Tillage machinery that can work the subsurface soil without turning over the topsoil is employed, as are planters that inject seed, fertilizer, and herbicides directly into the subsurface through small holes. Where wind erosion is the danger, straw mulches or row or basin listing (furrowing) may be employed and alternating strips of grass and open-field crops planted. Fields in danger of water erosion are plowed on the contour (not up- and downslope), and if cultivation by lister (a double-moldboard plow) is also employed, water storage capacity of the furrows will be increased. Strip cropping, in which alternate strips of different crops are planted on the contour, may also be employed. Crop rotations that provide for strips of

closely planted legumes and perennial grasses alternating with grains such as wheat and barley and with intertilled strips are particularly effective in reducing soil erosion. Fields may also be terraced, and the terraces strip-cropped.

Design of conservational cultivation also includes provision for grass-covered waterways to collect drainage from terraces and carry it into stream courses without erosion. Where suitable conditions of slope and soil permeability are found, shallow retention structures may also be constructed to promote water infiltration. These are of special value where insufficient soil moisture is a problem at times. Additional moisture always encourages more vigorous cover growth.

The measures just described for class 2, 3, and 4 lands may be considered preventive, but there are also rehabilitation measures where fields have already suffered from erosion but offer possibilities of restoration. Grading with mechanical equipment and the construction of small check dams across former gullies are examples.

For the remaining four classes of land, whose principal uses depend on the continuous maintenance of cover, management is more important than physical conditioning. In some cases, however, water retention structures, check dams, and other physical devices for retarding erosion may be applied on forest lands and rangelands. In the United States such structures are not often found in forest lands, although they have been commonly employed in Japanese forests. In forestry the conservational management objective is one of maximum production of wood and other services while maintaining continuous soil cover. The same is true for grass and other forage plants on managed grazing lands. For rangelands, adjustment of use is particularly difficult, because grazing must be tolerated only to the extent that the range plants still retain sufficient vitality to withstand a period of drought which may arrive at any time.

A last set of erosion-control measures is directed toward minimizing stream bank erosion, which may be large over the length of a long stream. This may be done through revetments, retaining walls, and jetties, which slow down current undercutting banks and hold sand and silt in which soil-binding willows, kudzu, and other vegetation may become established. Sediment detention reservoirs also reduce the erosive power of the current, and catchment basins or flood control storage helps reduce high flows. *See* AGRICULTURE; FOREST AND FORESTRY; LAND-USE CLASSES; RETAINING WALL; REVETMENT; RIVER ENGINEERING.

Irrigated agriculture. The impact of soil management on soil productivity is extremely dramatic when soils are irrigated. Flood irrigation has been utilized in many parts of the world, with stream water being diverted through canals into the furrows of individual plots of land. The enhanced population-support capacity in such regions helped give rise to ancient culture hearths such as those in the Nile and Tigris–Euphrates valleys and to the growth of large population numbers in the wet monsoon floodplains of India and China. Until the twentieth century, flood irrigation was the predominant method of managing water on soils, but with the explosion of technology came another form of water management—center-pivot irrigation—in which water is pumped to the surface from a central well and into a large sprinkler on wheels which is made to rotate slowly around the wellhead. Other, high-technology methods of irrigation have also appeared, such as fertigation in the Negev Desert of southern Israel, where computer-controlled desert water is mixed with fertilizers and then piped directly onto the root system of each crop plant. The twentieth century saw a startling worldwide increase in the total area under irrigation. World Watch Institute in Washington, DC, estimates that the Earth's irrigated acreage at the end of the twentieth century reached 630 million acres (255 million hectares), a global increase of some 35% in just 20 years.

Highly significant is the fact that this increase in irrigated soils is being counterbalanced in part by salinization, where plant growth is somewhat inhibited by accumulating salts both on and below the soil surface. Salinization occurs because all irrigation waters contain at least some dissolved salts that are left behind as the water either transpires or evaporates. Salinization is a serious problem in soil management. Soil conservation on irrigated soils can be effected through the control and even the diminution of salinization. Salinity control and desalinization require that additional amounts of water be applied to soils in order to leach accumulating salts downward beyond the root zone of crops. This process depends on the placing of drainage pipes some 10 ft (3 m) or more below the surface. As desalinization is expensive and poses the risk of soil waterlogging, it is less commonly implemented than most of the other conservation measures discussed. *See* DESALINATION; FLOODPLAIN.

Donald J. Patton

Bibliography. J. L. Dodd, Desertification and degradation in sub-Saharan Africa, *BioScience*, 44:28–34, January 1994; G. Gardner, Shrinking fields: Cropland loss in a world of eight billion, *WorldWatch Paper*, no. 131, WorldWatch Institute, Washington, DC, July 1996; I. W. Heathcote, *Environmental Problem-Solving: A Case Study Approach*, McGraw-Hill, New York, 1997; R. L. McConnell and D. C. Abel, *Environmental Issues: Measuring Analyzing and Evaluating*, Prentice Hall, Upper Saddle River, NJ, 1999; D. Pimental et al., Environmental and economic costs of soil erosion and conservation benefits, *Science*, 267:1117–1123, February 24, 1995; J. D. Soule and J. Piper, *Farming in Nature's Image*, Island Press, Washington, DC, 1992.

Soil degradation

Loss in the quality or productivity of soil that is often the result of human activities, such as agriculture practices, deforestation, mining, waste disposal, and chemical spills. Degradation is attributed to changes in soil nutrient status, biota, loss of organic matter,

deterioration of soil structure, and toxicity due to accumulations of naturally occurring or anthropogenic (human-made) materials. The effects of soil degradation include loss of agricultural productivity, negative impacts on the environment and economic stability, and exploitation of marginally suitable or virgin land. Opinions as to the present threat from soil degradation range widely, with estimates of up to 17% of the Earth's 11.5 billion hectares (28.4 billion acres) of vegetated land considered by some to be degraded. Though the actual extent of the problem is unclear, it is certain that soil degradation remains a significant issue, particularly for impoverished regions. *See* SOIL.

Soil biology. Healthy soil is rich with life, particularly microorganisms. Because soil is highly compartmentalized, it offers diverse habitats that can accommodate many microbial lifestyles. Based on deoxyribonucleic acid (DNA) analysis, it appears that some 10,000 species of microorganisms can be present in a single gram (0.04 ounce) of soil. Bacteria are the most numerous microorganisms in soil, whereas fungi represent the largest fraction of the microbial biomass (living mass). Other biota comprising the soil community include algae, protozoa, nematodes, and a variety of larger organisms, such as arthropods and earthworms. Globally, it has been estimated that the approximately 26×10^{28} (260 million billion trillion) prokaryotic microbial cells present in terrestrial environments contain 26, 6.2, and 0.65 trillion kilograms (57, 14, and 1.4 trillion pounds) of the world's carbon, nitrogen, and phosphorus, respectively, rivaling the importance of plants in the global nitrogen and phosphorus budget.

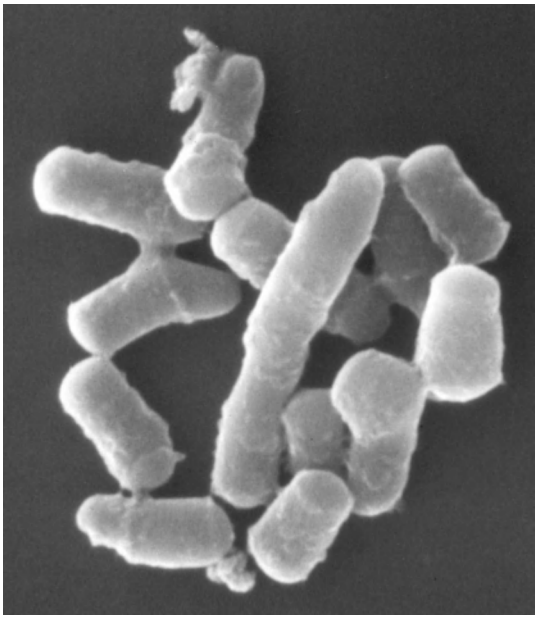
The soil microbial community is essential to the cycling of nutrients and the decomposition of wastes; thus hindrance of microbial activities may have serious ecological results. Fortunately, soil microbial communities are usually resilient and able to recover from most environmental insults, such as deposition of toxic chemicals. However, some terrestrial ecosystems, such as tundra or soils low in nutrient-holding capacity, are more fragile than others and thus recover more slowly from perturbations. Soil microbial activity is generally closely related to the amount of soil organic matter; thus degradative forces such as erosion that deplete organic matter also result in long-term losses of microbial activity.

While microorganisms are the driving force in maintaining soil quality and productivity, agricultural practices may stimulate soil microorganisms that have detrimental effects on humans. Some fertilizer components, such as urea, rapidly undergo microbial oxidation to form nitrate, which is mobile in the environment and causes health concerns, particularly for infants. This problem can be exacerbated by application of fertilizers in the fall, when there is no crop present to take up the nutrients. Administration of feeding antibiotics to livestock as a prophylaxis or to achieve modest gains in animal growth may promote the occurrence of antibiotic resistance among terrestrial organisms, which may serve as an environmental reservoir for resistance genes that could be transferred to microbial pathogens. Excessive concentration of animals may also directly introduce microbial pathogens into surface water or shallow ground water, frequently resulting in soil and water contamination with nutrients, a form of chemical degradation. *See* SOIL ECOLOGY; SOIL MICROBIOLOGY.

Chemical degradation. Failing to replace nutrients removed with crop harvests depletes the soil over the long term. Excessive leaching of the basic cations that buffer soil pH may result in soil acidification, changing the solubility, and thus availability, of certain nutrients to plants. Soil testing is meant to determine the appropriate rate of nutrient addition; however, procedures used in soil testing vary from region to region and may not accurately predict fertilizer needs. In sub-Saharan Africa, soils are becoming depleted due to insufficient fertilization, whereas in some parts of North America excessive fertilization results in contamination of water and perhaps changes in soil function. Even prudent use of fertilizers and liming materials to maintain or enhance productivity may mask the effects of soil degradation resulting from other agricultural practices. It is also possible for land management practices to result in concentrations of chemicals that are toxic to plants. For example, in arid lands salts may accumulate at the surface due to irrigation practices. Other land management practices, such as clear-cutting of forests, may result in long-term changes in nutrient cycling. *See* SOIL CHEMISTRY; SOIL FERTILITY.

Mining. Historically the mining of coal and various metal ores has been among the most devastating forms of land use. Strip (surface) mining can expose large amounts of reduced (decreased oxidation state) minerals in the form of mine tailings and rubble. In the case of coal mining, pyrite (FeS_2) is the source of the problem. A combination of chemical oxidation and microbial oxidation of iron and sulfur in the pyrite releases sulfuric acid (H_2SO_4) and ferric hydroxide ($Fe[OH]_3$) that drain away from the site in storm water. The acidic mine tailings are extremely difficult to revegetate, and without intervention remain exposed, continuing to produce acidic drainage until oxidizable material is depleted. Left untreated, strip mines may require 50 to 150 years for recovery. Reclamation of strip mines involves establishment of a rooting medium, typically by the addition of topsoil. Vegetation is selected for tolerance to acidic conditions. Most developed nations now require complete reclamation of strip mines by responsible parties. *See* AUTOXIDATION; LAND RECLAMATION; SURFACE MINING.

Toxic chemical spills. Hundreds of thousands of different organic compounds are synthesized commercially for a variety of purposes. Many of these compounds eventually find their way into soil ecosystems, often without incident. In some cases, such as petroleum spills, the compounds are toxic to soil organisms when present at high enough concentrations, and thus produce large shifts in microbial community structure, especially in fragile tundra soils.



Scanning electron micrograph of a *Rhodococcus* species (bacterium) isolated from an environmental sample. This organism can degrade a diverse array of aromatic pollutants, including quinoline, pyridine, catechol, benzoate, and protocatechuic acid.

Spillage of other toxic chemicals can cause similar long-term effects on microbiology. Eventually, microorganisms present in the soil degrade most organic contaminants, thus alleviating toxic effects. The time required for microorganisms to degrade contaminants depends on both the chemicals and the organisms involved. In some cases, organisms originating from other sites can be used as inocula to enhance degradation of contaminants. For example, a species of the bacterium *Rhodococcus* (see **illustration**) is capable of degrading quinoline, a toxic organic compound associated with synfuel waste water.

Unlike organic contaminants, which are usually degraded to nontoxic forms, toxic metals become essentially permanent features of soils once introduced. Fortunately, adsorption of metals to soil colloids decreases the availability of the metals for movement in the environment or uptake by plants, animals, and microorganisms. The presence of toxic metals provides genetic pressure for the selection of metal-resistant microorganisms, which have been detected at metal-contaminated sites. Metal contamination may be either natural or anthropogenic.

Pesticide use. Pesticides are chemicals designed specifically to kill or suppress targeted groups of organisms, such as weeds, insects, or in some cases microorganisms. As a result, there has been concern that pesticide use will eventually result in changes in the functioning of the soil microbial community. Numerous investigations have been completed with very little evidence for such effects; however, it is possible that some organisms are affected with no measurable impact on the overall function of the microbial community. Until recently, it was difficult to examine the impacts of pesticides on soil microorganisms, because most cannot as yet be cultivated. Molecular biology tools now available allow collection of considerable information about microbial communities without the necessity of cultivating the organisms.

Effects of pesticides on soil microbial communities are likely to be transient. Though most pesticides are xenobiotic (foreign to life) and thus did not exist in nature until recently, there are microorganisms able to degrade most of these substances, sometimes with no prior exposure to them. *See* AGRICULTURAL CHEMISTRY; PESTICIDE.

Physical degradation. Physical effects of soil degradation, such as erosion, are more obvious than other forms of degradation and thus are the most widely documented. Soils vary widely in the relative proportions and organization of the solid and pore spaces (structure). Soil structure, in part, determines the suitability of land for crop production or its vulnerability to erosion. Misuse of soil often results in compaction or densification (loss of pore space). Compaction affects the flow of water through soil and impedes growth of plant roots. Compacted soils generate greater runoff during storms and tend to be sensitive to erosion. Erosion results not only in lost soil productivity but also contamination of lakes and streams with sediment and associated agrochemicals. Based on river sedimentation rates, it appears that human-induced erosion on average roughly equals natural erosion, around 11.2 billion metric tons (11 billion tons) per year on a global basis, though the relative rates may very considerably from place to place. *See* EROSION.

Soil degradation and climate change. Wind erosion of unprotected soil can induce the development of desertlike conditions (desertification). It is apparent that this process was involved in the disappearance of several ancient civilizations, including the Harappan (western India), Mesopotamian (western Asia), and Mayan (Central America) cultures. Desertification continues to be a problem in Africa, Kazakhstan, Uzbekistan, and northern China. Soil degradation can affect not only regional climate, but also global climate. In turn, global climate change, particularly global warming, may promote soil degradation and worsen existing soil problems. Higher temperatures promote microbial decomposition of organic matter, leading to decreased soil productivity and increased susceptibility to erosion. Organic matter decomposition releases carbon dioxide into the atmosphere, exacerbating warming. Changes in rainfall patterns associated with global warming may also promote soil erosion by lengthening dry periods and increasing the intensity of rainy periods. It is expected that deserts will expand because the interiors of continents will likely be drier. *See* CLIMATE MODIFICATION; DESERTIFICATION; GLOBAL CLIMATE CHANGE.

Soil conservation. Certain agricultural practices, such as no-till farming, reduce erosion losses; in the United States, conservation tillage has decreased water erosion by as much as 40%. Terracing and the use of cover crops also promote soil conservation. In

wealthy nations, addition of crop nutrients based on sound and suitable soil test results is the most promising way to reduce the problem of excess fertilization, whereas more affordable fertility practices, such as crop rotations and the use of nitrogen-fixing trees have been successful in poor regions. *See* AGRICULTURAL AND SOIL PRACTICES; COVER CROPS.

Sometimes, practices used to treat soil degradation may actually damage soils in other ways. For example, remediation of soil contamination by incineration or pyrolysis essentially destroys the soil along with the contaminants. Less invasive approaches, such as bioremediation (use of microorganisms or plants to clean up contamination), depend on natural processes to degrade contaminants and generally do not further degrade soil properties. Increased understanding of soil degradation processes has made it technically feasible to avoid many of the practices that damage soil; however, cultural and economic factors continue to promote degradative land-use choices. *See* BIODEGRADATION; SOIL CONSERVATION.

Gerald K. Sims

Bibliography. J. Kaiser, Wounding Earth's fragile skin, *Science*, 304(5677):1616–1618, 2004; R. Lal and B. A. Stewart, *Soil Management Experimental Basis for Sustainability and Environmental Quality*, CRC Press, Boca Raton, FL, 1995; G. W. Olson, Archaeology: Lessons on future soil use, *J. Soil Water Conserv.*, 36:261–64, 1981; C. A. Seybold, J. E. Herrick, and J. J. Brejda, Soil resilience: A fundamental component of soil quality, *Soil Sci.*, 164:224–234, 1999; G. K. Sims, Biological degradation of soil, *Adv. Soil Sci.*, 11:289–330, 1990; W. B. Whitman, D. C. Coleman, and W. J. Wiebe, Prokaryotes: The unseen majority, *Proc. Nat. Acad. Sci. USA*, 95:6578–6583, 1998.

Soil ecology

The study of the interactions among soil organisms, and between biotic and abiotic aspects of the soil environment. Soil is made up of a multitude of physical, chemical, and biological entities, with many interactions occurring among them. Soil is a variable mixture of broken and weathered minerals and decaying organic matter. Together with the proper amounts of air and water, it supplies, in part, sustenance for plants as well as mechanical support.

Soil is predominantly a sand-silt-clay matrix, containing living materials (biomass) including microorganisms, roots, and animals, plus dead organic matter, with varying amounts of gases and liquids. Soils, in addition to the three geometric dimensions, are greatly influenced by the fourth dimension, time.

Soils are the result of several factors, including climate, organisms, parent material, and topography (relief), acting through time (**Fig. 1**). These factors affect major processes, such as primary production, decomposition, and nutrient cycling, which lead to the development of ecosystem properties unique to a soil type. Such characteristics as cation-exchange capacity, texture, structure, and organic matter status are the outcome of these major processes operating as constrained by the controlling factors.

Soil profile. Abiotic and biotic factors lead to certain chemical changes in the top few decimeters (8–10 in.) of soil. In many soils, particularly in more mesic or moist regions of the world, there is a certain amount of leaching and redeposition of nutrients, often accompanied by a distinct color change. Thus, the air-litter surface is at the top of the profile, below are the zones of litter, fermentation, and humification (termed 00, 01, and 02, respectively), and below these is the mineral soil surface, which contains most of the organic matter (A horizon). The B horizon is next, containing deeper-dwelling organisms and somewhat weathered material. This is followed by the C horizon, the unconsolidated mineral material above bedrock.

The work of the soil ecologist is made easier by the fact that the surface 10–15 cm (4–6 in.) of the A horizon has the majority of plant roots, microorganisms, and fauna. A majority of the biological-chemical activities occur in this surface layer; indeed, most of the microorganism- and alga-feeding fauna, such as protozoa and rotifers and tardigrades, are in the top 1–2 cm (0.4–0.8 in.). The continual input of leaf, twig, and root materials, and the algal and cyanobacterial production and turnover make this region prime for biological activity.

The clay fraction in soil, so important to microorganism life and plant activity because of its nutrient content, comprises particles less than 2 micrometers in diameter. Unlike the sand-silt minerals, clays are weathered forms of primary minerals; they are therefore referred to as secondary minerals. Coarse clay particles (0.5 μm) often are derived from quartz and mica; finer clays (0.1 μm) are clay minerals or weathered products of these (such as hydrated ferric, aluminum, titanium, and manganese oxides). The effects of coarse and fine clays on organic matter dynamics have been studied intensively. Labile (that is, rapid turnover) constituents of organic matter may be preferentially adsorbed onto fine clay particles, and thus may be a prime source of energy for the soil microorganisms.

Input of organic matter to soil is one of the major agents of soil structure. The organic matter comes from both living and dead sources (roots, leaves, microorganisms, and fauna). Various physical processes, such as deformation and compression by roots and soil fauna and by freezing-thawing and wetting-drying also have significant influences on soil structure. Plant roots and resistant organic breakdown products (humus) are important in the formation of aggregates. Bacteria, fungi, and their metabolic products are equally important in promoting granulation. The interaction of organic matter and mineral components of soils has a profound influence on cation adsorption capabilities. This property is important to both agrosystem and natural ecosystem management. *See* AGROECOSYSTEM; HUMUS; SOIL CHEMISTRY.

Biological aspects of soil. The biological aspects of soil consist of major organic inputs, decomposition

by primary decomposers (bacteria, fungi, and actinomycetes), and interactions between microorganisms and fauna (secondary decomposers) which feed on them (Fig. 1).

The detritus decomposition pathway occurs on or within the soil after plant materials (litter, roots, sloughed cells, and soluble compounds) become available through death or senescence. Plant products are used by microorganisms (primary decomposers). These are eaten by the fauna, which thus affect flows of nutrients, particularly nitrogen, phosphorus, and sulfur. The immobilization of nutrients into plants or microorganisms and their subsequent mineralization are critical pathways. These processes are diagrammed in **Fig. 2**, showing flows out from the microorganisms and fauna via mineralization or direct losses into organic pools. The labile inorganic pool is the principal one that permits subsequent microorganism and plant existence. Scarcity of some nutrient often limits production. Most importantly, it is the rates of flux into and out of these labile inorganic pools which enable ecosystems to function successfully. *See* ECOSYSTEM; SYSTEMS ECOLOGY.

Model of soil interactions. Conceptual models are useful for envisioning ecosystem processes. The array of microorganisms, fauna, and roots all provide organic material belowground, which may move through the underground ecosystem in defined ways (**Fig. 3**). Pathways of organic matter utilization by primary decomposers include bacteria, actinomycetes, or saprophytic fungi. There is considerable grazing by such bacterial feeders as protozoa, nematodes, tardigrades, and enchytraeids. The final major primary decomposer group is the fungi, a wide array of saprophytic forms and vesicular-arbuscular mycorrhizae which take up nutrients and are mutualistic with the plant roots. The fungal grazers include stylet-bearing nematodes, collembolans, mites, and certain protozoa that are spore grazers. There is a general group of predators such as large predatory mites, nematodes, or centipedes which prey on bacteria-feeders and fungus-feeders and presumably some top predators as well.

While fauna in general account for less than 5% of the total detritus-decomposer respiration and even less biomass (see **table**), their indirect, catalytic role in decomposition is considerably greater in many ecosystems. Such indirect roles include feeding and its effects on microorganism (prey) populations, translocation of nutrients and microorganisms to different locations in the soil profile, and even, in some cases, immobilization of nitrogen and phosphorus in feces or in nests, such as those made by termites.

Earthworms play an important role in mixing and comminuting soil. They ingest large amounts of soil; a single worm such as *Lumbricus terrestris* can ingest several times its own body mass per day in a grassland or pasture. Hence, earthworms have a significant impact on nutrient transformation and availability. In addition to direct ingestion of leaf litter and other organic debris pulled from the soil surface, the earthworms will ingest and generally triturate any

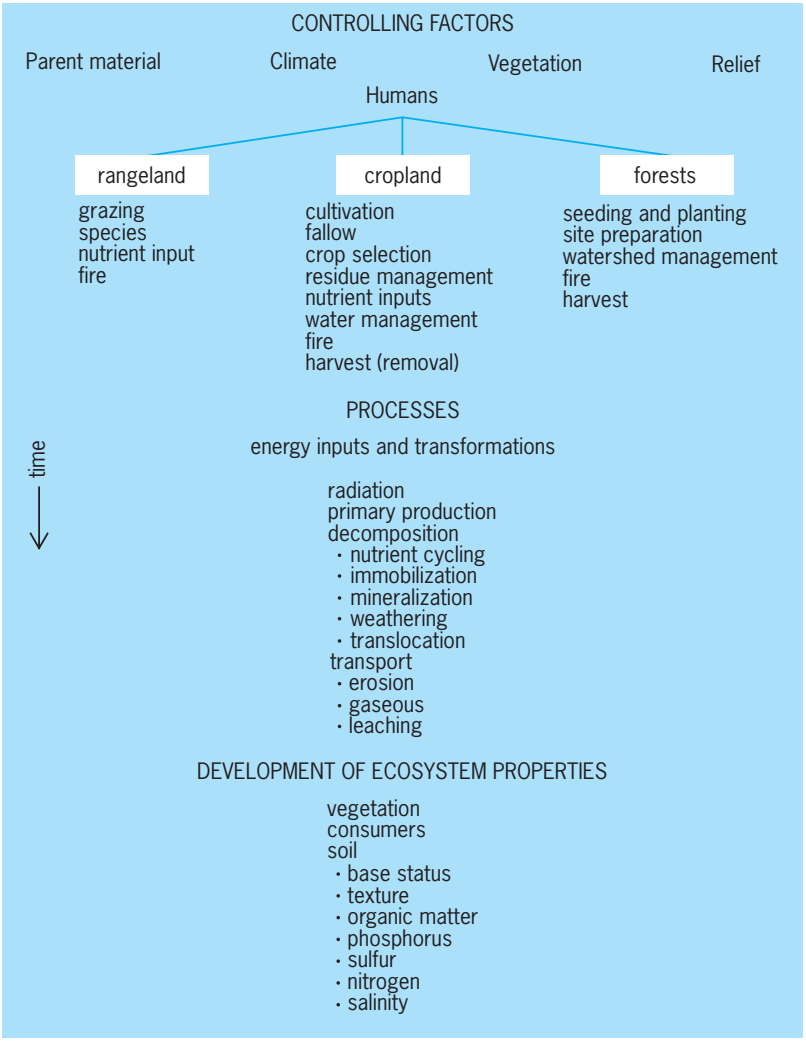


Fig. 1. Factors influencing soil development. Controlling factors affecting processes, over time, influence ecosystem properties. (*From D. C. Coleman, C. P. P. Reid, and C. V. Cole, Biological strategies of nutrient cycling in soil systems, Adv. Ecol. Res., 13:1–55, 1983*)

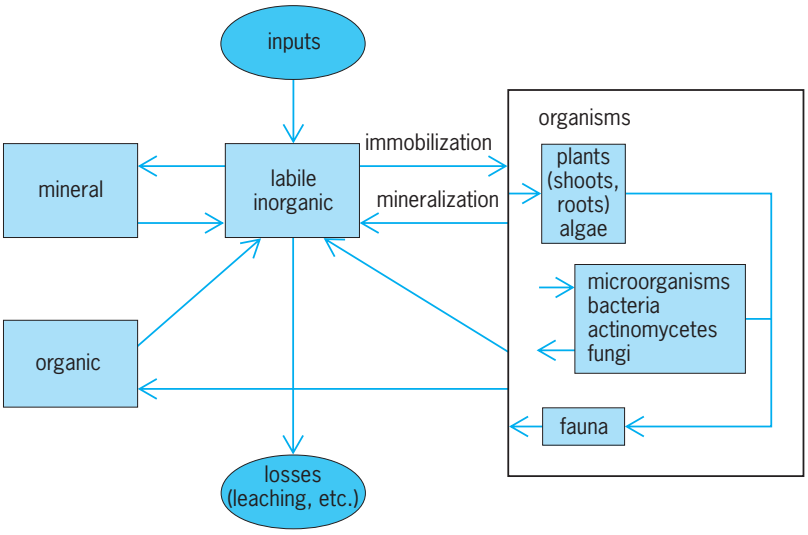


Fig. 2. Generalized nutrient-cycling scheme in soil. The biological activities of immobilization and mineralization are of major importance in ecosystem function. (*From D. C. Coleman, C. P. P. Reid, and C. V. Cole, Biological strategies of nutrient cycling in soil systems, Adv. Ecol. Res., 13:1–55, 1983*)

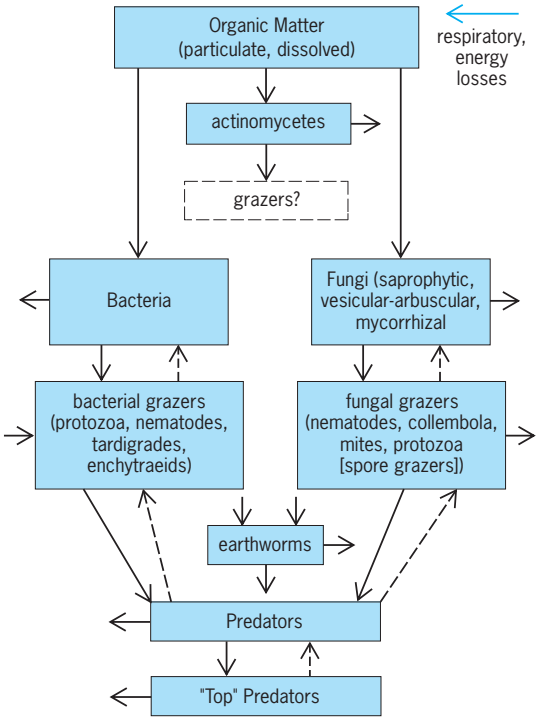


Fig. 3. Conceptual model of breakdown of particulate and soluble organic matter in agroecosystems. Vertical arrows show material flows; horizontal arrows show respiratory energy losses; and broken arrows show information feedback from lower trophic levels.

small bits of organic matter, including protozoa and nematodes. *See* OLIGOCHAETA.

Microorganism–fauna interactions. A number of studies have demonstrated the marked effects of protozoa on both bacterial numbers and nutrient dynamics. Protozoa, among the smallest and most abundant soil fauna, feed upon, and promote turnover of, microorganism populations. This has been shown in various natural ecosystems (forest, grassland, marine) as well as cropped-land ecosystems.

Soil incubations (microcosms) have been used to examine the role of some of the soil fauna in decomposition. The microfauna (protozoa) may be as abundant as 10,000–100,000 per gram of soil. Soil mesofauna, such as nematodes, are larger (about 100 times larger than protozoa), and still numerous (100–1000 per gram of soil). Both groups of organisms are voracious feeders on bacteria and fungi, and account for significant effects on nitrogen and phosphorus mineralization in soil (Figs. 2 and 3). Impacts on the primary decomposer populations vary, as noted below.

Several investigators have cultured fungus-feeding nematodes on various species of fungi and have monitored the fungal and nematode populations. Typically, they observed some inhibition of mycelial growth or respiration in cultures with nematodes versus those without nematodes. Grazing at very low nematode population levels actually stimulates the growth of fungi slightly, which is called a grazing-optimization response. The degree of such a response in natural soil detritus food chains is not known. In several studies, there have been complete shifts of fungal species abundances due to preferential feeding by Collembola (an order of primitive wingless insects) on certain fungal species in aspen woodland and in pine forests. *See* COLLEMBOLA.

Spatially, there are three regions where increased populations of microorganisms and fauna occur: roots, litter, and aggregates. These regions (or microsites) exist at interfaces, or zones where substrates accumulate. Roots produce exudates such as carbohydrates, amino acids, exfoliates (sloughed cells), and mucigel (in conjunction with microorganism activity) as the roots grow through soil. Microorganisms are more abundant in the rhizosphere (the volume 0.08–0.16 in. or 2–4 mm out from the root surface), and microorganism grazer populations (protozoa, nematodes) also are larger in the rhizosphere region. *See* RHIZOSPHERE; ROOT (BOTANY).

Several research groups in North America and Europe have investigated the effects of feeding by detrital food-chain organisms on plant growth. By using either protozoa or nematodes in experimental microcosms with growing plant seedlings, significant increases were observed in both plant shoot growth and nitrogen content when bacteria-feeding

Numbers of soil organisms in various ecosystems at four different sites*

Trophic group	Aspen poplar, Alberta, Canada	Lodgepole pine, Alberta, Canada	Dryland wheat, Nebraska	Mixed-grass prairies, Nebraska
Bacteria per gram soil	—	—	1.3×10^9	1.2×10^9
Fungi, meters per gram soil†	—	—	34	78
Protozoa per gram soil	—	—	37,000	43,000
Microbial biomass carbon, micrograms per gram litter or soil	21,000 (litter)	16,500 (litter)	966 (soil)	1400 (soil)
Nematodes (microbivorous‡) per gram soil	—	—	52	74
Collembola (fungal-feeding) per square meter	71,000	22,000	1300	2000
Mites (fungal-feeding and predatory) per square meter	158,000	135,000	900	1200

*1 g = 0.035 oz. 1 m^2 = 10.7 ft^2. 1 m = 3.28 ft.
† That is, hyphal length is measured in meters.
‡ Bacterial and fungal feeding.

fauna were present. The fungus–fungus-feeder food chain, however, showed no significant increases in plant growth when compared with the appropriate control treatments, probably because the fungi are very efficient mineralizers of nitrogen.

Selective removal or inhibition of certain target organisms such as fungi, nematodes, and arthropods has been done by using biocides in the field to determine microorganism–fauna interactions. Fungicides, insecticides, and nematicides have been used singly or in combination in field plots in various field sites, and then decomposition has been measured. Compensatory responses of certain decomposers have been observed; with an inhibition of fungi, bacteria become more abundant and active. The bacteria-feeding fauna then multiply, and predators upon them in turn become more active and enhance mineralization processes. *See* SOIL MICROBIOLOGY.

Mutualistic interactions in soils. There are several root–microorganism interactions in soil which are mutualistic, that is, beneficial for both organisms. Prominent among these are the mycorrhizal associations between certain fungi and plant roots. Unlike the saprophytic fungi mentioned earlier, these fungi are dependent on the host plant for energy-containing carbon compounds. When the thread-like filaments (hyphae) of these beneficial fungi encounter the growing roots, they penetrate the cortical cells, or the spaces between the cells, and establish metabolic connections with the plant. The fungus receives energy compounds from the host, and there is a beneficial exchange of mineral nutrients. The hyphae grow through the soil, outside the rhizosphere zone that has been depleted of nutrients (primarily inorganic phosphorus and nitrogen), and transport them into the host plant. This mutualism is so beneficial to the plant that mycorrhizal plants can sometimes grow 20 or 30 times more rapidly than otherwise comparable nonmycorrhizal plants. In addition, they may be more resistant to drought and better able to repel disease organisms than nonmycorrhizal plants. *See* MYCORRHIZAE.

Within certain groups of plants, such as the legume family, including various species of beans, clover, and alfalfa, there is another important group of symbionts. When a legume root grows through soil near the symbiotic bacteria *Rhizobium*, the two future partners initiate production of several compounds, including complex carbohydrate recognition chemicals, called lectins, that allow the root to recognize the bacteria as beneficial rather than pathogenic. The bacteria then invade the root tissue and form a nodule, a swollen structure on the root surface that houses an active population of bacteria, in which nitrogen fixation takes place. By using photosynthetic energy from the plant and a specialized array of enzymes (nitrogenases), the *Rhizobium* bacteria can incorporate (fix) gaseous molecular nitrogen, N_2, breaking its triple covalent bonds and incorporating it into organic compounds, which are then mineralized and made available to the plant. This biological source of nitrogen frees legumes from a dependence on soil-borne sources and allows them to grow vigorously where many other kinds of plants could not thrive without fertilization. *See* BIOGEOCHEMISTRY; LECTINS; NITROGEN CYCLE.

Agroecosystems versus natural ecosystems. There have been several studies of minimum tillage or no-tillage in agricultural systems and how it affects microorganism populations and associated nutrient dynamics.

Soil microorganism and biochemical changes associated with reduced tillage in soybeans and corn have been studied. Differences were found in microorganism populations which convert nitrogenous compounds to nitrate, or produce gaseous nitrous oxide (N_2O) or N_2, related principally to changes in soil water, organic carbon and nitrogen, and pH when the reduced tillage regimes were imposed. There was generally a higher carbon, nitrogen, and water content in the surface soil (0–7.5 cm or 0–3 in.) under no-till, reflected by higher microorganism populations and enzyme activities. These relations were reversed from 7.5–15 cm (3–6 in.) depth in no-till versus conventional tillage, probably because plowing places crop residues at lower depths.

Interestingly, increased numbers of microarthropods and earthworms were found in zero-tilled versus conventional-tilled fields in several locations, such as Georgia and Kentucky, and in Europe. These increases in some faunal groups are probably due more to the effects of tilling on abiotic physical parameters than to increased food availability in no-till.

Systems studies. It seems important to use some integrative properties of ecosystems, particularly those relating to functional groups or guilds, so that groups of organisms that feed on similar types of substrates can be considered. Nematodes and protozoa feeding on bacteria may have a more related role in ecosystem function than various fungus-feeding nematodes and fungus-feeding mites and Collembola, which, in turn, form their own guilds (Fig. 3). A functional understanding of feeding guilds will enable better understanding of the diverse processes that function in the general framework of decomposition, immobilization, and mineralization. These processes are key focal points in determining dynamics of terrestrial ecosystems. *See* AGRICULTURAL SOIL AND CROP PRACTICES; ECOLOGY; GUILD; POPULATION ECOLOGY; SOIL.

David C. Coleman

Bibliography. D. C. Coleman et al., Interactions of organisms at root/soil and litter/soil interfaces in terrestrial ecosystems, *Agric. Ecosys. Environ.*, 24:117–134, 1988; D. C. Coleman, C. P. P. Reid, and C. V. Cole, Biological strategies of nutrient cycling in soil systems, *Adv. Ecol. Res.*, 13:1–55, 1983; E. T. Elliott and D. C. Coleman, Let the soil work for us, *Ecol. Bull.*, 39:23–32, 1988; E. T. Elliott, H. W. Hunt, and D. E. Walter, Detrital foodweb interactions in North American grassland ecosystems, *Agric. Ecosys. Environ.*, 24:41–56, 1988; K. Killham, *Soil Ecology*, 1994; D. Parkinson, Linkages between resource availability, microorganisms and soil invertebrates, *Agric. Ecosys. Environ.*, 24:21–32, 1988; R. E. Phillips, *No-Tillage Agriculture Principles*, 1984.

Soil fertility

The ability of a soil to supply plant nutrients. Sixteen chemical elements are required for the growth of all plants: carbon, oxygen, and hydrogen (these three are obtained from carbon dioxide and water), plus nitrogen, phosphorus, potassium, calcium, magnesium, sulfur, iron, manganese, zinc, copper, boron, molybdenum, and chlorine. Some plant species also require one or more of the elements cobalt, sodium, vanadium, and silicon. *See* PLANT GROWTH.

Nutrient forms in soil. While carbon and oxygen are supplied to plants from carbon dioxide in the air, the other essential elements are supplied primarily by the soil. Of the latter, all except hydrogen from water are called mineral nutrients.

Only part of the 13 essential mineral nutrients in soil are in a chemical form that can be immediately used by plants. The unusable (unavailable) parts, which eventually do become available to plants, are of two kinds: they may be in organic combination (such as nitrogen in soil humus) or in solid inorganic soil particles (such as potassium in soil clays). The time for complete decomposition and dissolution of these compounds varies widely, from days to hundreds of years.

The quantity of the immediately available nutrient is often small. For example, it is common for a soil to contain only 30 lb of available nitrate nitrogen per acre (33 kg/ha) in the top foot (1 ft = 0.3 m) of soil. The same quantity of soil might contain 4000 lb (1800 kg) of unavailable nitrogen in organic compounds (soil humus). As another example, a soil might contain 25,000 lb of total potassium per acre (28,000 kg/ha), with only 300 lb/acre (336 kg/ha) immediately available to crops.

Limits to plant growth. Soils exhibit a variable ability to supply the mineral nutrients needed by plants. This characteristic allows soils to be classified according to their level of fertility. Fertility levels can vary from a deficiency to a sufficiency, or even toxicity (too much), of one or more nutrients. A serious deficiency of only one essential nutrient can still greatly reduce crop yields. For example, a soil considered fertile in all other ways could have levels of available potassium too low for optimum plant growth. Control of crop growth by the deficient nutrient is known as Liebig's law of the minimum. This law states that the plant growth rate is controlled by that mineral nutrient present in the most limiting quantity in the soil (**Fig. 1**). Modern chemical fertilizers applied to the soil reduce or eliminate nutrient deficiency and ensure that fertility does not limit a soil's potential to produce crops. *See* FERTILIZER.

Although fertilizers have dramatically increased the soil's potential to produce crops, other variables often control crop production. Both climatic and other soil factors can affect crop yields, as can crop variety, diseases, and insects. Elements of the climate such as insufficient water, temperature stresses (either too hot or too cold), and natural disasters such as wind and hail storms can all lower crop production regardless of a soil's inherent or cultivated fertility. Other undesirable soil conditions such as compacted soil layers, soil salinity, or acid soil layers that restrict root growth can also reduce crop yield.

Influence of crop type. The type of crop grown on a particular soil determines the required soil fertility level. Crops differ greatly in their need for nutrients and their ability to extract them from soils. Also, even if two crops extract a nutrient equally, one crop may use that nutrient more efficiently than the other.

Perhaps the most extreme example of the crop-type influence would be to compare a soil's nitrogen fertility for a legume such as alfalfa with its nitrogen fertility for a grass such as maize. Since alfalfa, through its symbiotic association with *Rhizobium* bacteria, obtains nitrogen from the dinitrogen (N_2) in the air, a severely nitrogen-deficient soil would not limit the growth of alfalfa. The growth of maize would be slowed by insufficient nitrogen, however, since maize depends directly on the uptake of available nitrogen in the soil. *See* NITROGEN CYCLE; SOIL MICROBIOLOGY.

Cation exchange and inherent fertility. Several soil properties are important in determining a soil's inherent fertility. One property is the adsorption and storage of nutrients on the surfaces of soil particles. Such adsorption of a number of nutrients is caused by an attraction of positively charged nutrients to negatively charged soil particles (similar to the attraction of the positive and negative poles of two magnets). This adsorption is called cation exchange (adsorbed cations can be exchanged with other cations in solution), and the quantity of nutrient cations a soil can adsorb is called its cation-exchange capacity. A cation is a single atom or a group of two or more atoms combined that has lost one or more of its electrons. Since atoms are made of equal numbers of protons (each has one positive charge) and electrons (each with one negative charge), they have zero charge. Some atoms and molecules lose one or more electrons, resulting in an excess of protons, thereby causing a

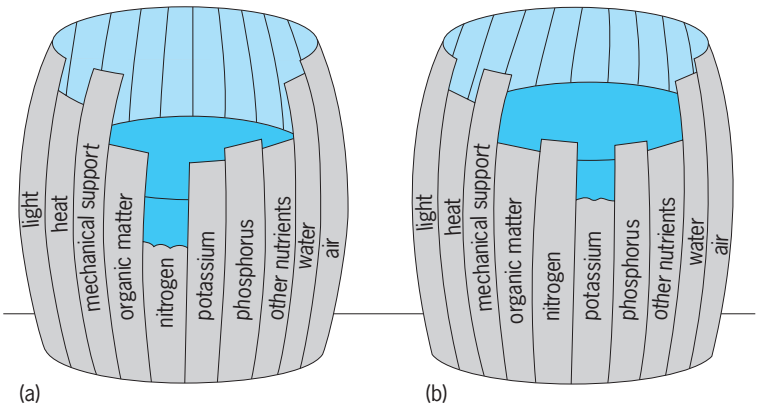


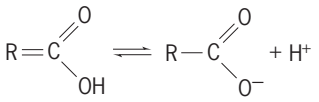
Fig. 1. An illustration of the principle of limiting factors. The level of water in the barrels above represents the level of crop production. (*a*) Nitrogen is represented as being the factor that is most limiting. Even though the other elements are present in more adequate amounts, crop production can be no higher than that allowed by the nitrogen. (*b*) When nitrogen is added, the level of crop production is raised until it is controlled by the next most limiting factor, in this case, potassium. (*After N. C. Brady, The Nature and Properties of Soils, 9th ed., Macmillan, 1984*)

positive charge. The plant nutrient calcium is a cation with two more protons than electrons, giving it two positive charges (Ca^{2+}). Atoms or groups of atoms having more electrons than protons are negatively charged ions known as anions. *See* ADSORPTION; ION EXCHANGE; SOIL CHEMISTRY.

The negative charge in soils is associated with clay particles. This negative charge results largely from an imbalance of positive and negative charges from cations and anions within their crystalline structure. The negatively charged clay particle adsorbs enough cations to maintain an overall zero charge balance. The cations adsorbed by clay particles are largely the plant nutrients calcium (Ca^{2+}), magnesium (Mg^{2+}), and potassium (K^{+}) with lesser amounts of the essential micronutrients manganese (Mn^{2+}), zinc (Zn^{2+}), iron (Fe^{2+} and Fe^{3+}), and copper (Cu^{2+}). Sodium (Na^{+}) can also be adsorbed in significant quantities, especially in soils of arid and semiarid regions. The cations Ca^{2+}, Mg^{2+}, K^{+}, and Na^{+} are often referred to as a soil's exchangeable bases. The adsorbed cations are in equilibrium with the same cations in the soil water (soil solution). Plant roots absorb most nutrient ions directly from the soil solution. As nutrient cations are taken from the soil solution, exchange reactions resupply the solution with that ion from the soil clays to maintain a nearly constant concentration.

In addition to the exchangeable bases and other nutrient cations, acid cations such as trivalent aluminum, hydroxy-aluminum, and the hydrogen ion are also adsorbed by the soil's negative charges. *See* CLAY MINERALS.

Organic matter and cation exchange. Some of the soil's cation-exchange capacity may also arise from organic matter (humus) in the soil. Negative charges of organic matter arise largely from carboxylic and phenolic acid functional groups. Since these functional groups are weak acids, the negative charge from organic matter increases as the soil pH increases, as shown below in equilibrium for carboxylic acid. As soil pH increases, hydrogen ion removal from the right side of the equilibrium causes the formation of additional negatively charged carboxyl groups. *See* CHEMICAL EQUILIBRIUM; HUMUS.



The negative charges of soil organic matter can adsorb the same cations as described for the soil clays. The proportion of cation-exchange capacity arising from mineral clays and from organic matter depends on the proportions of each in the soil and on the kinds of clays. In most mineral soils, the soil clays constitute the greater proportion of cation-exchange capacity. Within the class of mineral soils, those soils with more clay and less sand and silt have the greatest cation-exchange capacity.

The soil cation-exchange sites provide the immediate store of available nutrient cations that can be used by plants. Because of this store of available nutrients, those soils with greater cation-exchange capacities generally provide a greater supply of available nutrients for crop production. However, the fertility level of each plant nutrient must be balanced to the needs of the crop to achieve high productivity.

Soil acidity. The amounts and kinds of acid cations adsorbed on the cation-exchange sites can have a substantial influence on a soil's perceived fertility. Soils with a pH of 7 have their exchange sites nearly 100% saturated with the bases Ca^{2+}, Mg^{2+}, K^{+}, and Na^{+}. As soil pH's drop below 7, the proportion of the exchange complex saturated with acid cations begins to increase. Between pH 7 and 5.5, the predominant acid cation may be hydroxy-aluminum, which is tightly adsorbed to the clays and causes no reduction in plant growth. At soil pH's around 5.5 and below, the acid cation trivalent aluminum (Al^{3+}), which is toxic to plants, begins to appear on the cation-exchange sites. It and excess manganese (Mn^{+2}) together are the greatest hazards to the growth of healthy crops at low soil pH's. In some cases of extremely low soil pH, insufficient calcium may also limit crop growth, but such cases are rare. *See* PH.

Two factors cause soils to become acid: when crops are harvested, exchangeable bases are removed as part of the crop; and exchangeable bases move with drainage water below the crop's root zone (leaching). Leaching of bases is enhanced when acids are released to the soil from plant roots, from microbial respiration, and from other microbially mediated processes, especially nitrification, the conversion of ammonium nitrogen to nitrate nitrogen. For each ammonium ion converted to nitrate nitrogen, two hydrogen ions are released. Hydrogen ions or aluminum cations resulting from acid decomposition of clays can exchange some of the bases from cation-exchange sites to the soil solution to allow their loss by leaching. Since much of the nitrogen fertilizer supplied to crops contains ammonium nitrogen (this is true of both manure and chemical fertilizers), the addition of high rates of nitrogen enhances soil acidification. Finely ground limestone (calcium carbonate) can be applied to acid soils when needed to neutralize soil acids and to add Ca^{2+} and Mg^{2+}, thereby raising soil pH back to the desirable range for optimum crop production. *See* ACID AND BASE.

Nutrients in soil minerals. While the supply of some plant nutrient cations (especially calcium, magnesium, and potassium) is related to their amounts on a soil's cation-exchange sites, there is often a greater supply of these nutrients within the soil minerals themselves. However, nutrients from soil minerals become available to crops only after the minerals partially dissolve and release the individual ions to the soil solution. Most soil clays contain some potassium and magnesium within their crystalline structures. Feldspars in some soils contain potassium, while other minerals contain supplies of magnesium and calcium. Calcite in some high-pH soils constitutes a large supply of calcium. Other minerals supply many of the micronutrients in soil.

Minerals containing some micronutrients and, especially, the major nutrient phosphorus furnish

plant-available forms by slow dissolution. As plant roots absorb these nutrients, more of the soil mineral dissolves in the soil solution. Many micronutrients also have some immediately available forms adsorbed onto the cation-exchange sites, but phosphorus does not since its plant-available forms ($H_2PO_4^-$ and HPO_4^{2-}) are both anions.

Organic matter and nutrient supply. Most of the nitrogen and sulfur in soils are a part of soil organic matter. Soil organic matter can also contain phosphorus and micronutrients, although the major portions of these nutrients are components of soil minerals. Soils in the midwestern United States typically contain around 3000 lb of organic nitrogen per acre (3300 kg/ha) in their plow layer. During 1 year, less than 0.5% of the organic nitrogen might be converted to the plant-available ammonium and nitrate forms by soil microorganisms. The amount of total organic sulfur in soil and the amount converted to plant-available sulfur would both be approximately 10% of the nitrogen values. The amounts of mineralization (conversion from organic to available mineral forms) per year are only approximate, since mineralization depends on many factors, including soil properties, soil temperature, soil moisture, and the quality and quantity of decomposing crop residue. These factors can vary considerably from one location to another and from year to year at a given location.

Soil organic matter, especially the small proportion of soluble organic matter, can increase the supply of micronutrients available to plants by acting as a natural complexing agent. Complexing agents are large organic molecules that can encapsulate some micronutrients. This encapsulation maintains higher levels of that nutrient in the soil solution, increasing its availability for plant uptake. *See* COORDINATION COMPLEXES; MOLECULAR RECOGNITION.

Mineralization versus fertilizer use. Those nutrients that are stored primarily in soil organic matter (nitrogen and sulfur) may be mineralized in insufficient quantity for maximum production of organically grown crops. The amounts of nitrogen and sulfur required by crops are proportional to the yield of crop. For example, the adequate nutrition of corn generally requires 1.6 lb (0.7 kg) of nitrogen in the entire corn crop (roots, stalks, and grain) for each bushel of corn produced. A high-yielding corn crop of 200 bu/acre (80 bu/ha) would therefore require 320 lb (145 kg) of nitrogen per acre (200 bu/acre × 1.6 lb N/bu). Wheat requires about 2.5 lb (1.1 kg) of nitrogen in the entire crop for each bushel produced. A wheat crop of 70 bu/acre (28 bu/ha) would therefore require 175 lb N/acre (193 kg/ha). Other crops have their own specific nitrogen requirements that are proportional to their expected yield. The crop requirements for sulfur can be calculated similarly, except that the amounts required are about 6% of those for nitrogen. The expected yield would depend on other plant growth factors such as the anticipated rainfall or irrigation, the availability of other soil nutrients, or expected losses from pests.

The amount of nitrogen mineralized from soil organic matter to supply the crop's requirements for plant-available nitrogen depends on soil properties such as organic matter and clay concentrations and on environmental conditions such as temperature and the availability of water. The amounts of available nitrogen released from soil organic matter during the cropping season under typical agronomic practice of continuous grain crop production have been carefully measured in a limited number of studies and have been found to vary from 15 to 130 lb/acre (16.5 to 143 kg/ha), which is insufficient for high yields of corn or wheat. Following a nitrogen-fixing legume crop such as alfalfa, or immature legume or grass cover crops (often called green manure crops) that provide crop residues rich in nitrogen, the total amounts of nitrogen mineralized may be in the range of 115 to 230 lb/acre (126.5 to 253 kg/ha). Crop residues from green manure crops that are added to soil will decompose, and some of the green manure crop's nitrogen will mineralize, adding to the available nitrogen released from the soil organic matter alone. In these cases, no additional nitrogen fertilizer is usually needed when corn is grown following a good stand of alfalfa or a rich green manure crop, whereas nitrogen fertilizer is needed when corn follows another crop of corn. Therefore, when using organic crop production practices, it may be beneficial to include legumes or other green manure crops in rotation, or to add animal manures to enhance nitrogen availability and to ensure maximum crop production.

David E. Kissel

Soil testing. Soil testing is the use of rapid chemical analyses to determine the fertility status of a soil. It is used to identify those nutrients or substances that are present in either insufficient or excessive quantities for optimum plant growth. Analyses are also used to monitor increases or decreases in soil fertility over time. Soil test data are critical for determining rates of commercial fertilizer, animal wastes, and other nutrient sources that maximize profit to farmers that are environmentally responsible. Soil testing is also used by commercial greenhouse operators and by private individuals for lawn and garden management.

The first step for the user is collecting a representative sample from the field or area of interest. This process is of vital importance for the resulting soil tests to be meaningful. Attention must be given to the number of cores collected, the core locations, depth of sampling, time of sampling, and how the samples are handled after collection. The samples are usually submitted to a soil testing laboratory that will perform the tests and normally will provide an interpretation of the results, including fertilizer recommendations.

Another approach to soil sampling is zone or grid sampling. The objective is to generate a map of the levels of the soil parameter of interest rather than just an average for the entire field. Computerized fertilizer spreaders that obtain positioning information from satellites and can vary rates as they move across the field are used to apply fertilizer according to the mapped soil test levels.

Most of the laboratory procedures used for evaluating soil fertility involve use of specific chemical solutions to extract a portion of one or more nutrients from the soil. The solutions and procedures are

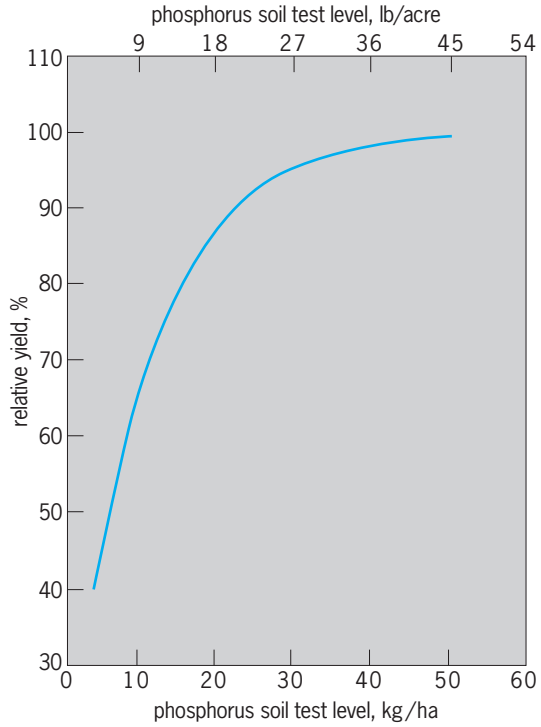


Fig. 2. Relationship between phosphorus soil test level and average relative corn yield or percent sufficiency in Illinois. (*After T. R. Peck et al., eds., Soil Testing: Correlating and Interpreting the Analytical Result, Amer. Soc. Agron. Spec. Publ. 29, 1977*)

selected from a database of correlation studies that have been conducted previously. These studies determine the correlation of the chemically extracted nutrient with the soil's ability to supply the nutrient to plants. The amount of the element present in the soil extract is determined by analytical instruments, which may be as basic as a colorimeter or specific ion electrode or as sophisticated as a plasma spectrophotometer capable of simultaneous multielement analysis.

Due to great diversity in soil chemical properties, several different laboratory procedures have been developed for each nutrient. The procedure must be appropriate for the soils in question or test results will be erroneous. For example, acid-based phosphorus extractants such as weak mixtures of hydrochloric acid and ammonium fluoride work well on acid or neutral soils. However, when used on alkaline soils containing calcium carbonate, the acid is often neutralized before it can extract any phosphorus, creating a false low soil test reading. Alkaline extractants, such as those containing sodium bicarbonate, were developed specifically for soils containing calcium carbonate. Soil testing laboratories determine which procedure to run, based on the properties of the soils being analyzed.

Indices of the soil nutrient availability are developed from the soil test values obtained from these determinations; they have practical significance only when the interpretation is based on extensive field calibration. In soil test calibration, researchers conduct many experiments across the geographic region of interest and across a range of soil test values. Each experiment includes rates of the nutrient in question such that the optimum rate for the site can be determined. Once a sufficient number of experiments have been performed, the probability of getting a response to application of a nutrient and the average percent sufficiency at a given soil test level can be determined (**Fig. 2**). The same studies are used to estimate the amount of a particular nutrient to apply to give optimum yield at a given soil test level.

The most subjective part of soil testing is the process of deriving specific fertilizer recommendations from soil test values. A multitude of factors can influence this process, including land tenure, supply of capital, fertilizer and crop prices, crop quality effects, crop rotation, and fertilizer placement. Therefore, the recommendations developed are dependent on the assumptions made with respect to these factors. Ideally they are fine-tuned to fit each individual situation. *See* AGRICULTURAL SOIL AND CROP PRACTICES; ANALYTICAL CHEMISTRY; PLANT MINERAL NUTRITION; SOIL.

Paul Fixen

Bibliography. C. A. Black, *Soil Fertility Evaluation and Control*, CRC Press, Boca Raton, FL, 1993; N. C. Brady, *The Nature and Properties of Soils*, 10th ed., Macmillan, New York, 1990; J. R. Brown (ed.), *Recommended Chemical Soil Test Procedures for the North Central Region*, Missouri Agricultural Experiment Station, Columbia, 1998; M. R. Carter (ed.), *Soil Sampling and Methods of Analysis*, CRC Press, Boca Raton, FL, 1993; R. H. Follett, L. S. Murphy, and R. L. Donahue, *Fertilizers and Soil Amendments*, Prentice Hall, Englewood Cliffs, NJ, 1981; J. L. Havlin et al., *Soil Fertility and Fertilizers: An Introduction to Nutrient Management*, Prentice Hall, Upper Saddle River, NJ, 1998; J. L. Havlin and J. S. Jacobsen (eds.), *Soil Testing: Prospects for Improving Nutrient Recommendations*, Soil Science Society of America, Madison, WI, 1994; T. R. Peck (ed.), *Soil Testing: Correlating and Interpreting the Analytical Results*, American Society of Agronomy, Madison, WI, 1977; R. L. Westerman (ed.), *Soil Testing and Plant Analysis*, rev. ed., Soil Science Society of America, Madison, WI, 1990.

Soil mechanics

The science that applies laws and principles of physics, especially continuum and particulate mechanics and hydraulics of porous media, to solve problems encountered in civil engineering that are related to force-deformation behavior of soil and the flow of fluids through soil. Civil engineers use the term soil for a naturally occurring multiphase material consisting of gas or liquid and solid particles, whether or not the material supports plant life. The term geotechnics or geotechnical engineering refers to a subspecialty of civil engineering that combines the disciplines traditionally known as soil mechanics, soil dynamics, rock mechanics, earthwork engineering, and foundation engineering for the design or analysis of natural or constructed facilities of, in, or on rock or soil. *See* ROCK MECHANICS.

Fundamental principles. Soil mechanics rests on five fundamental relationships. The effective stress

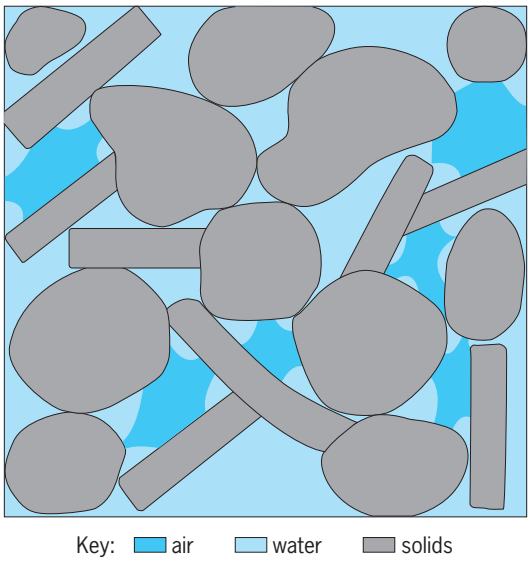


Fig. 1. Diagram showing the three phases in a thin section of a soil element.

principle formulated by K. Terzaghi in the early 1920s in relation to work on the consolidation of clays, is given by Eq. (1), where σ = total normal

$$\sigma = \overline{\sigma} + u \tag{1}$$

stress, $\overline{\sigma}$ = effective normal stress, and u = pore water pressure. The effective stress principle unifies and extends the applicability of other fundamental relationships such as Darcy's law, Eq. (2),

$$Q = kiA \tag{2}$$

where Q is flow rate, k permeability, i hydraulic gradient, and A cross-sectional area; Coulomb friction, Eq. (3), where τ is shear stress, σ_n normal stress,

$$\tau = \sigma_n \tan \phi \tag{3}$$

and ϕ friction angle; Newton's second law of motion, Eq. (4), where F is force, m mass, and a acceleration; and Hooke's law, Eq. (5), where σ is stress,

$$F = ma \tag{4}$$

$$\sigma = E\epsilon \tag{5}$$

E modulus of elasticity, and ϵ strain. Most of the advances that have been made in soil mechanics use these fundamental principles as their starting point. *See* FRICTION; HOOKE'S LAW; NEWTON'S LAWS OF MOTION.

Nature of soil. Soil consists of a multiphase assemblage of mineral or organic particles, as shown in the diagram in **Fig. 1**. The voids within this solid soil skeleton contain gas (generally air), liquid (generally water), or both. A saturated soil has its voids completely filled with water. Various mathematical equations are used to determine the relationships that exist among the three phases within a soil element, an arbitrarily small volume within a soil mass.

When an element of dry soil experiences a load (force F in **Fig. 2**), normal (N) and tangential (T) forces develop immediately at the contact points of the soil particles. The soil mass deforms as the air in the voids compresses or escapes. If the soil voids contain water (Fig. 2*b*), an essentially incompressible liquid, the spring–dashpot analogy helps visualize the load transfer process. Upon application of the load, the pore fluid (water in chamber) cannot compress, and the small size of the soil pores (valve orifice) restricts the escape of water. The pressure in the pore water increases to counteract the applied force F. Immediately after time $t = 0$, water slowly flows out through the tortuous path formed by the microscopic pores. The soil skeleton (spring) begins to feel an increased load, deforming as a load transfer from the pore water to the soil skeleton occurs. After sufficient time has elapsed, pressure within the pore water returns to zero, flow through the soil pores ceases, and the system reaches an equilibrium position, with a compressed soil skeleton resisting the applied force. The applied force imposes a stress F/A (force F per unit area A) on the system, and a strain d/L in the vertical direction (deformation d per unit

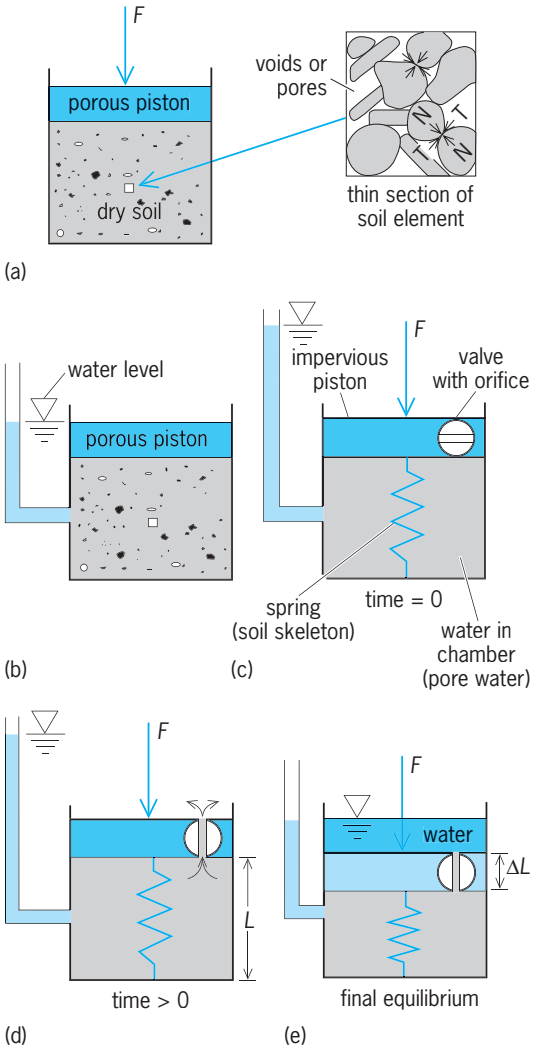


Fig. 2. Load transfer and consolidation; the effects of the application of a load (F) to soil. (*a*) Dry soil (N = normal force; T = tangential force). (*b–e*) Soil containing water and the corresponding spring and dashpot analogy from time = 0 until establishment of equilibrium.

length L in a given direction) in the equilibrium position. Geotechnical engineers consider stresses and strains positive when compressive. A system with large pores reaches equilibrium quickly, while a system with small pores requires a long time to reach equilibrium.

The spring–dashpot system exhibits linearly elastic behavior. Doubling the force F doubles the deformation. Upon removing the force F from the spring–dashpot system, the compressed spring pushes the piston back to its original position as water flows back into the chamber. In soils, crushing and slippage of particles results in nonlinear behavior and unrecoverable deformations that prevent a total rebound of the system to its original position. This nonlinear inelastic behavior means that soils do not obey Hooke's law relating stress to strain.

Since the early days of modern soil mechanics, Hooke's law served as a benchmark to evaluate how far soil behavior deviated from its simple generalization. Complex nonlinear stress–strain relationships attempt to model more closely the deformation characteristics of soil, but linear elasticity continues to provide a useful and sufficiently accurate simplification for the solution of many engineering problems.

Soil classification and index properties. The discipline of soil mechanics relies on the quantification of observed soil behavior. The geotechnical engineer utilizes soil classification to organize soils that exhibit similar behavior into groups. The **table** shows a simplified soil classification based on the widely accepted Unified Soil Classification System (USCS). Simple index tests such as grain size and plasticity help classify soils.

Among the most common index tests, the Atterberg limits that use the water content at the boundary between states of consistency to correlate soils. The water content of a remolded soil at which a $^1/_2$-in. (13-mm) section of a $^5/_{64}$-in.-wide (2-mm) groove closes from 25 successive shocks in a standard device corresponds to the liquid limit. In effect, the test measures the water content required for an arbitrary but constant shear strength of the remolded soil sample. Therefore, all soils have approximately the same shear strength at the liquid limit. The plastic limit corresponds to the water content at which a thread of soil with a diameter of $^1/_8$ in. (3.2 mm) crumbles when pressed together and rerolled. The shrinkage limit corresponds to the minimum water content required to completely fill all the voids of a dried soil pat (or similarly, the water content at which further drying will not cause additional shrinkage of the sample).

Correlations between fundamental soil properties (such as strength, permeability, compressibility, and expansivity) and index tests help the engineer obtain approximate solutions. The actual soil properties measured in laboratory or field tests can vary greatly from the values suggested by the index property correlations.

Pore pressures and heads. Water in the soil pores exists at a measurable pressure—positive, zero, or negative. Zero pore pressure corresponds to atmospheric pressure in the pore water. **Figure 3** shows the relationship between pore pressure and head. An instrument known as a piezometer measures total head near its sensor. The elevation of the piezometer sensor equals the elevation head, h_e. The dimensions of a piezometer sensor generally do not exceed 4 in. (10 cm) in diameter and 3.3 ft (1 m) in length. The total head, h_t, equals the elevation to which water rises in a piezometer. The height of water in the piezometer equals the pressure head, h_p. The pore pressure, u, is given by Eq. (6), where γ_w equals the

$$u = h_p \cdot \gamma_w \tag{6}$$

unit weight of water. Wells can have long collection

A simplified soil classification*

Material	Symbol	Grain size	USCA symbol	Name
Coarse-grained				
Boulder	None	12 in. + (30 cm +)		
Cobbles	None	3–12 in. (7.6–30 cm)		
Gravel	G		GW	Well-graded gravel
Coarse		$^3/_4$–3 in. (1.9–7.6 cm)	GP	Poorly graded gravel
Fine		No. 4, $^3/_4$ in. (1.9 cm)	GM	Silty gravel
			GC	Clayey gravel
Sand	S		SW	Well-graded sand
Coarse		No. 10–No. 4	SP	Poorly graded sand
Medium		No. 40–No. 10	SM	Silty sand
Fine		No. 200–No. 40	SC	Clayey sand
Fine-grained		Passing No. 200		
Silt	M		ML	Silt
			MH	Elastic silt
Clay	C		CL	Low-plasticity clay
			CH	High-plasticity clay
Organic	O		OL	Organic silt
			OH	Organic clay
			PT	Peat

*U.S. Standard Sieve Sizes (partial listing): 12 in. (300 mm); 3 in. (75 mm); No. 4: 0.2 in. (4.75 mm); No. 10: 0.08 in. (2.00 mm); No. 40: 0.02 in. (0.425 mm); No. 200: 0.003 in. (0.075 mm).

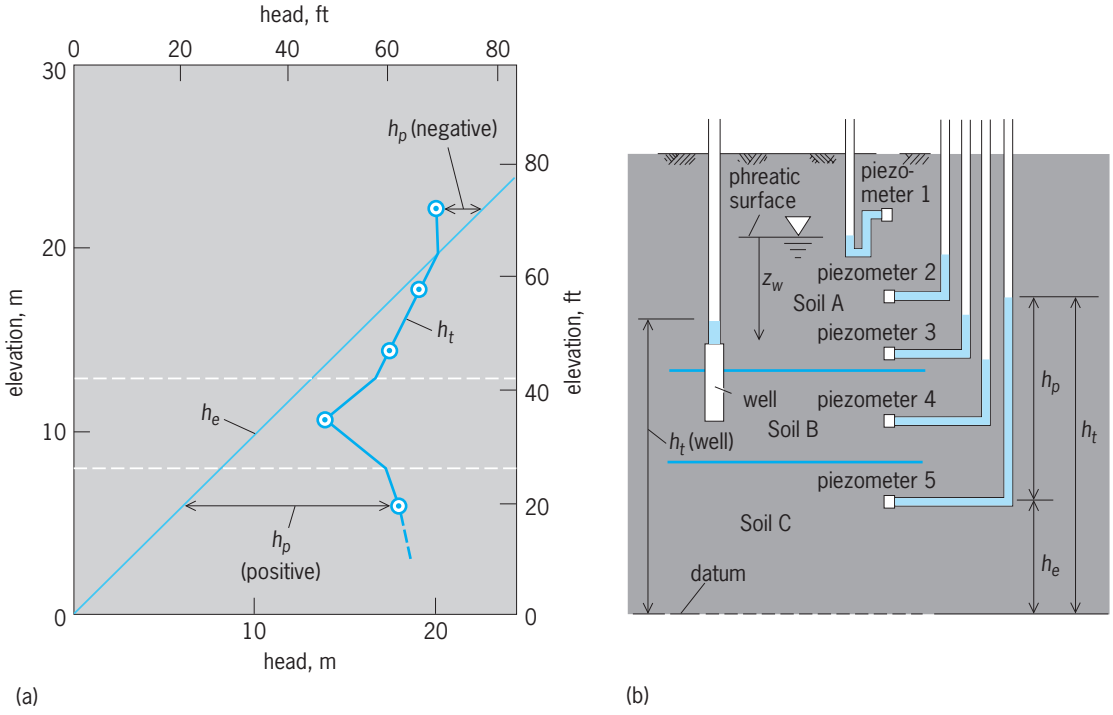


Fig. 3. Relationship between pore water and heads. (*a*) Relationship between heads in multilayered soil. (*b*) Placement of the piezometers and well [h_e = elevation head; h_p = pressure head (pore pressure); h_t = total head; z_w = depth below the phreatic surface (water table)].

zones that penetrate different soil layers. This characteristic makes total head data from wells difficult to interpret and use, as shown in Fig. 3*b*.

Positive pore pressures decrease effective stresses, which results in increased deformations and decreased soil shearing resistance. In general, the higher the pore pressure, the higher the possibility of encountering unsatisfactory performance.

For hydrostatic conditions, the total head remains constant everywhere, no flow occurs and the pore pressure, u, is given by Eq. (7), where z_w equals

$$u = z_w \cdot \gamma_w \tag{7}$$

the depth below the phreatic surface, the imaginary line where $h_p = 0$. Most soil deposits do not exist under hydrostatic conditions. A steady flow of water through the soil results from a constant pattern of pore pressures, known as the steady-state pore pressures or heads. Pore pressures different from steady-state values, known as excess pore pressures, result from loading, unloading, rainfall, increase in river or reservoir levels, leaks from pipelines or tanks, and drying of the soil, among other causes.

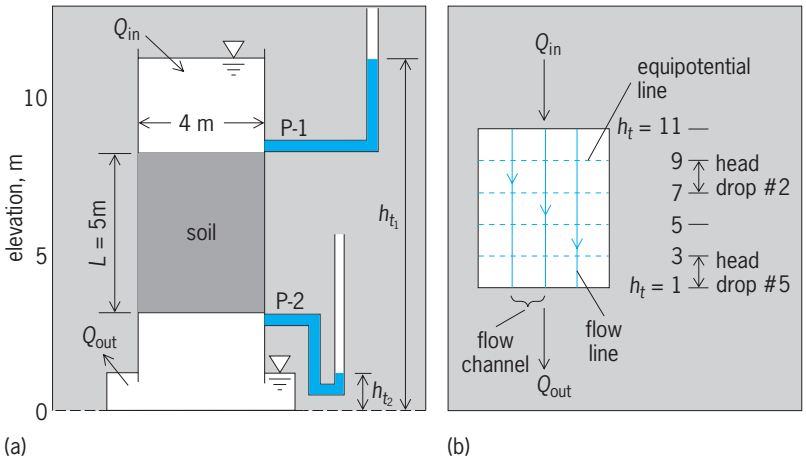


Fig. 4. Diagrams of flow through soil. (*a*) Flow through a permeameter. (*b*) Simple flow net.

Flow. Total heads determine the direction of flow in soil. Water flows from a zone of higher total head toward one of lower total head. Darcy's law [Eq. (2)] describes accurately the flow of water through the soil pores. **Figure** 4*a* illustrates flow through soil under conditions similar to those used by Darcy for his experiments. Rewriting Darcy's law as Eq. (8),

$$Q = kiA = k\frac{h_{t_1} - h_{t_2}}{L} \tag{8}$$

yields a flow rate equal to 8.5×10^{-3} gal/s (3.2×10^{-5} m^3/s). [Here Q is flow rate, k permeability or hydraulic conductivity, i hydraulic gradient, h_{t_1} total head at entrance end of soil, h_{t_2} total head at exit end of soil, L length of soil sample, and A cross-sectional area of sample.] Darcy's coefficient of permeability varies more than any other engineering soil property, as much as 10 orders of magnitude (or 10^9 times) between a clean gravel and a plastic clay. Furthermore, discontinuities such as cracks or very pervious thin layers can greatly affect the permeability of a soil deposit.

A flow net provides a graphical representation of flow through soil, particularly useful for evaluating complex conditions. For two-dimensional flow through isotropic soil, the flow net consists of a set of flow lines and orthogonal equipotential lines that divide the soil into curvilinear squares. A flow line traces the flow path through the soil for a drop of

water. While an infinite number of flow lines exist, a flow net utilizes an arbitrary selection to divide the soil into a convenient number of flow channels, n_f. Equipotential lines, drawn through points of equal total head, divide the flow net into a number of uniform head drops, n_d. The flow per unit length, l, perpendicular to the page is given by Eq. (9),

$$\frac{Q}{l} = k \cdot \Delta h \cdot \oint \tag{9}$$

where Δh = total head drop = $h_{t_1} - h_{t_2}$, and

$$\oint = \text{shape factor} = n_f / n_d$$

The flow net for the Darcy permeameter (Fig. 4*b*) has a shape factor equal to 4/5, which yields a flow equal to that computed by using Darcy's equation. The gradient between two points in the net equals the head drop divided by the distance between them. Since the net in Fig. 4 has equally spaced equipotentials, the gradient in the direction of flow remains constant.

Figure 5*a* shows a flow net for an earth dam with the following characteristics: Δh_t = 88 ft (27 m), n_f = 9.9, n_d = 18,

$$\oint = 0.55$$

and Q/l = 0.014 ft²/day (0.0013 m²/day) for $k = 4 \times 10^{-8}$ in./s (1×10^{-7} cm/s). The gradient in the direction of flow varies throughout the flow region, increasing as the flow approaches the drain. (Note that the distance between equipotentials decreases in Fig. 5*a*.) Areas of uncontrolled high gradients can experience instability problems from low effective stress and internal erosion. The graded filter shown on the upstream side of the drain ensures that the fine particles of the compacted fill do not get washed into the drain. Filters generally consist of layers of selected soil with particle sizes chosen to prevent migration of fines. Geosynthetic filters, rapidly growing in acceptance, consist of one or more layers of fabric engineered to retain soil fines. *See* DAM; GEOSYNTHETIC.

A flow net provides information on the pore pressure throughout the flow region. Rather than using the entire flow net for some analyses, engineers can obtain the information applicable to a surface of interest (for example, a potential shear surface) and portray the heads obtained from the net in terms of a piezometric line. Figure 5*b* shows that the piezometric line does not necessarily coincide with the phreatic surface. Under conditions of horizontal flow or no flow, the piezometric line and phreatic surface coincide.

Figure 5*c* shows a simplification that permits evaluation of the flow into a fully penetrating well using Darcy's law. The Dupuit assumption to Darcy's law states that the hydraulic gradient remains constant from top to bottom in the pervious formation and

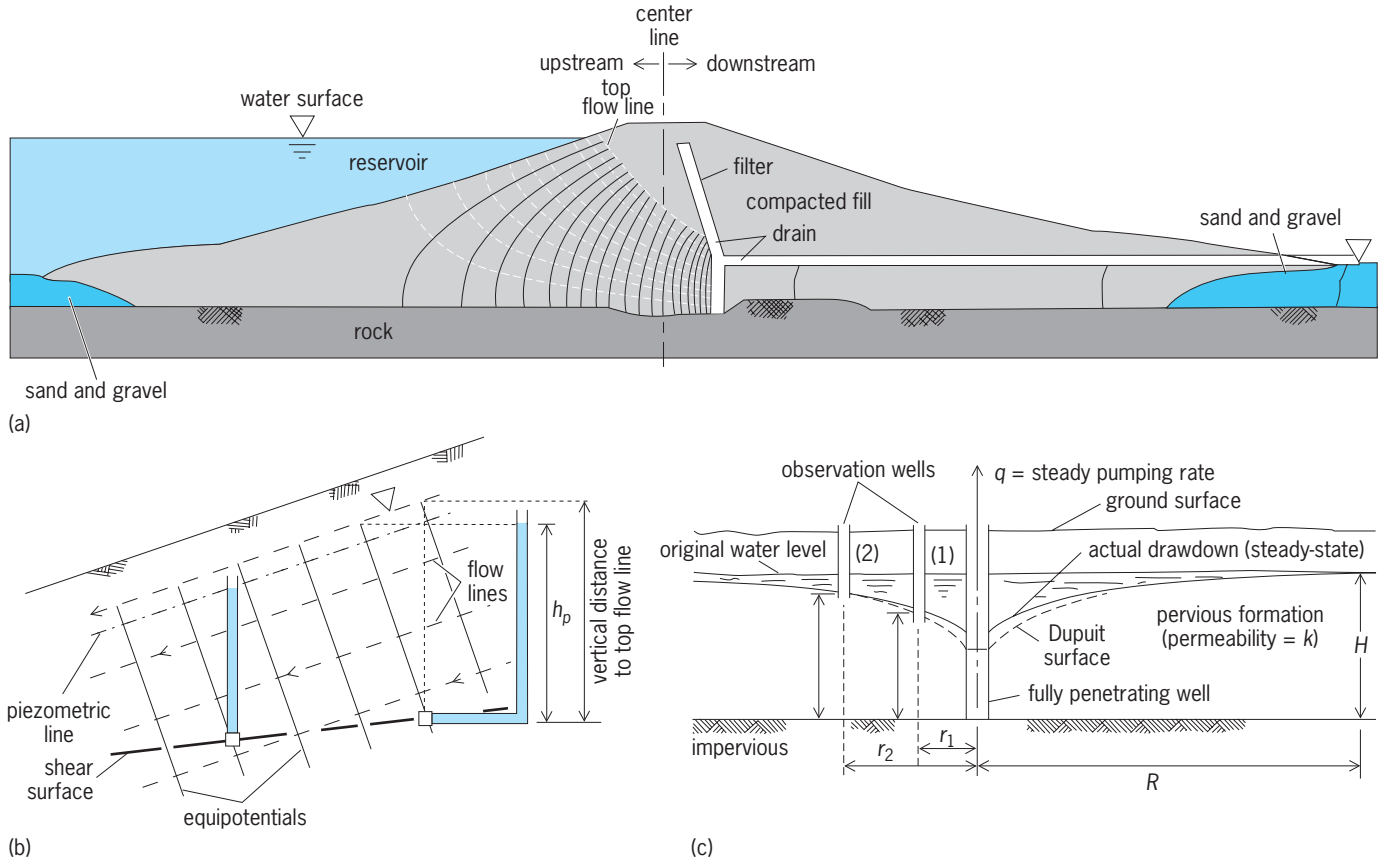


Fig. 5. Analysis of flow through soil of an earth dam. (*a*) Profile of flow through the dam. (*b*) Flow parallel to the slope. Piezometric line is a line located at a vertical distance h_p from a surface of interest; uppermost flow line represents the phreatic surface (water table). (*c*) Flow toward a fully penetrating well.

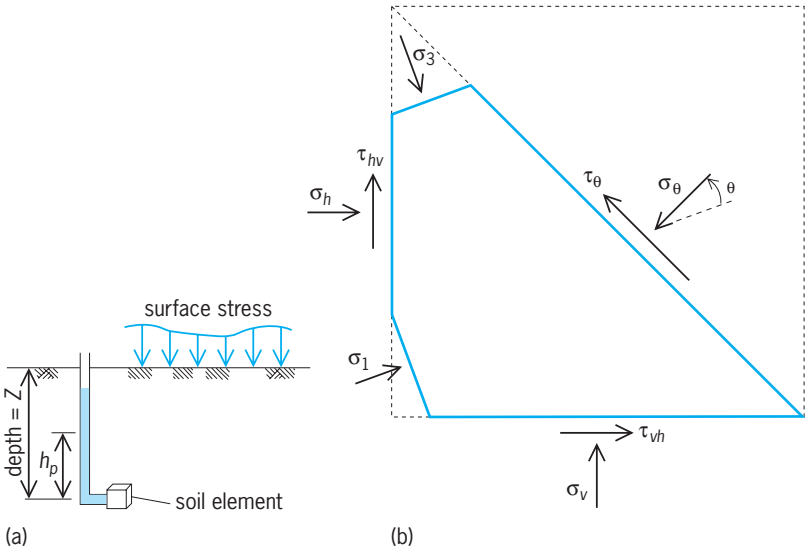


Fig. 6. Stress in soil represented by (*a*) soil element beneath loaded surface and (*b*) free-body diagram with stress vectors. τ = shear stress. Friction angle (θ) is measured counterclockwise from direction of normal stress (σ_1).

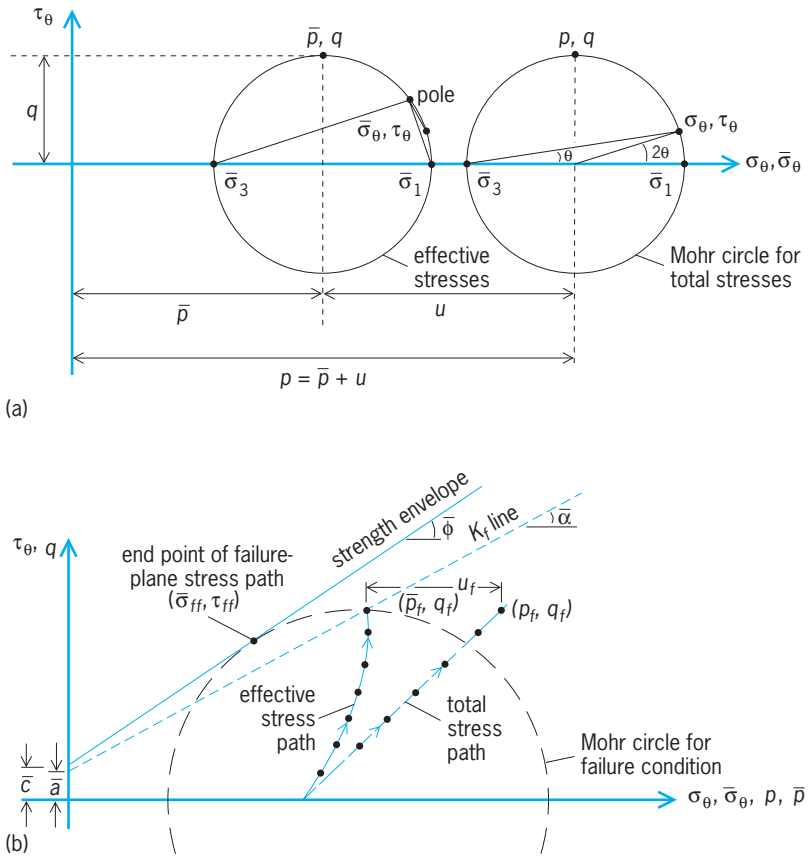


Fig. 7. Mohr's circle (*a*) for state of stress at a point and (*b*) stress path to failure for the soil element shown in Fig. 6.

equal to the slope of the water surface, as in Eq. (10).

$$i = dh/dr \tag{10}$$

Then, the flow rate, q, out of the well using any set of consistent units is given by formula (11), where q

$$q = \frac{\pi k(b_2^2 - b_1^2)}{\ln(r_2/r_1)} \tag{11}$$

is water quantity per unit time (Q/t). The values for b_1, b_2, r_1, and r_2 are measured (Fig. 5*c*) at observation wells 1 and 2, respectively. Pumped wells have many uses in projects involving civil engineering, including water supply, reduction of pore pressures to increase effective stresses and thus strength, and control of contaminants in the groundwater.

Stress in soil. **Figure 6** illustrates the concept of stress in soil. Geotechnical engineers frequently work with normal stresses acting perpendicular to a surface of interest, shear stresses acting parallel to a surface of interest, and pore pressures (sometimes known as neutral stresses) acting equally in all directions, expressed as stress vectors (Fig. 6*b*). *See* SHEAR.

The effective stress principle [Eq. (1)] states that the total normal stress experienced by an element of soil consists of the effective normal stress, which controls strength and deformation, and the pore water pressure. Since water can not sustain shear, pore water pressures do not affect shear stresses. A bar over the symbol identifies a stress as effective or a property as related to effective stresses (for example $\overline{\sigma}$, $\overline{\phi}$, $\overline{c}$.

The stress history of a soil greatly influences its behavior. Two powerful mathematical tools used to understand and make use of stress history are Mohr's circle of stress and stress path (**Fig. 7**). Mohr's circle provides a graphical representation of the stresses in any direction acting at a point. A distance equal to the magnitude of the pore pressure separates the circle for total stresses from that representing effective stresses. Each point along the circle represents the normal and shear stresses acting on a particular plane. The pole of the Mohr circle relates the stresses to the physical planes on which they act. A line through the pole and point σ_θ, τ_θ will lie parallel to the plane on which σ_θ and τ_θ act.

Two particular points prove very helpful in portraying stress history—the point for stresses on the failure plane, often taken as the point of maximum obliquity (maximum value of $\tau/\overline{\sigma}$), and that for maximum shear stress τ (identified by coordinates p and q). The total stress path (TSP) and effective stress path (ESP) provide a locus of stress points representing successive states of stress in the history of the soil. For clarity, Mohr circles usually do not appear in stress path plots, although one could reconstruct the circle at any point along the stress path. Failure occurs when the effective failure-plane stress path reaches the strength envelope (also known as yield surface) or when the $\overline{p}$-q stress path reaches the K_f line as in Fig. 7.

The stress path can reach the strength envelope because of a loading, an unloading, an increase in pore pressure, or a combination of these. Stress paths help engineers identify mechanisms and select methods and soil parameters for solving a problem. *See* STRESS AND STRAIN.

Strength. Soil behaves as a frictional material; that is, it resists shear stresses by the frictional action between particles. **Figure 8** shows a strength test where a sample under a constant normal force gets sheared along a predetermined horizontal plane. The displacement-versus-shear stress plot (Fig. 8*b*)

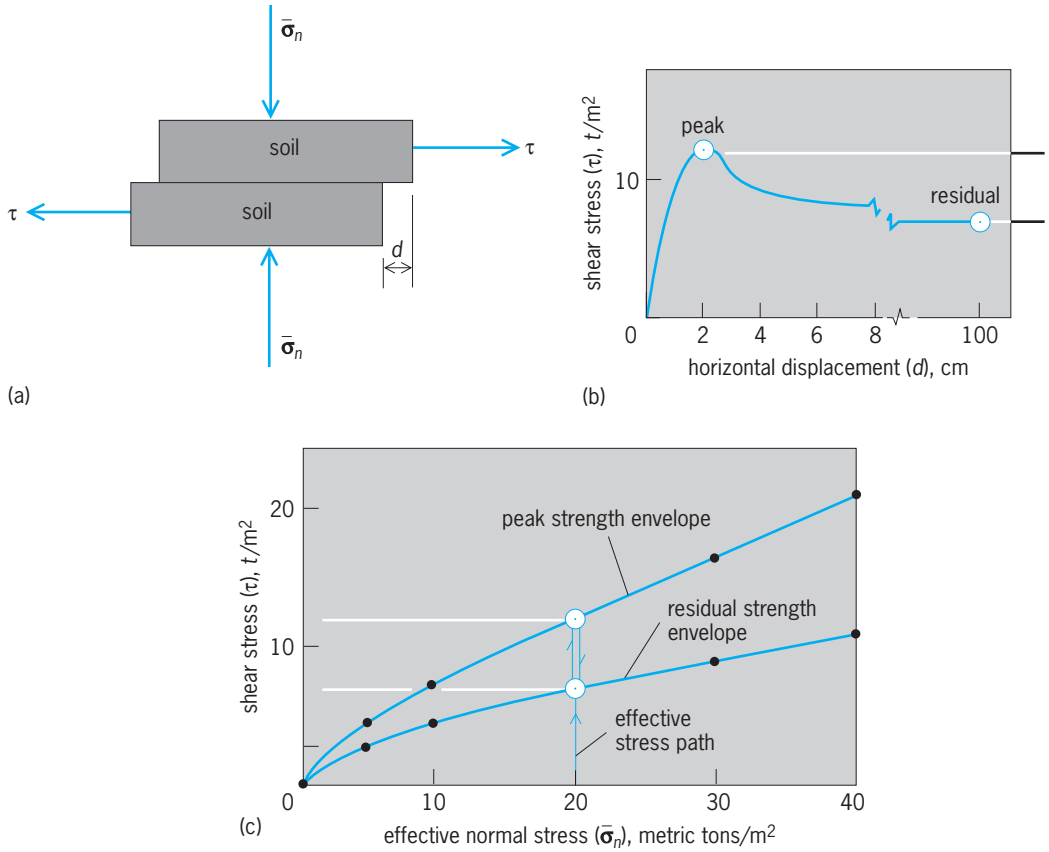


Fig. 8. Determination of soil strength. (*a*) Schematic diagram of a direct shear test. (*b*) Shear stress as a function of horizontal displacement. (*c*) Effective failure-plane stress path and strength envelopes. τ = shear stress; σ_n = effective normal stress. 1 m² = 10.8 ft²; 1 cm = 0.4 in.; 1 metric ton = 1.1 ton.

shows two important characteristics—a peak shear strength and a residual value obtained after large deformations. A series of these tests at different values of normal stress yields the peak and residual strength envelopes shown. These strength envelopes, unique for practical purposes, remain valid for a wide range of effective stress paths to failure. Some soils exhibit a monotonically ascending displacement-versus-shear stress curve where the peak strength equals that at large deformations.

Soils generally do not have true cohesion, that is, when $\overline{\sigma} = 0$, strength = 0. Since strength envelopes can be curved, use of the linear Coulomb friction equation (3) requires care in the selection of parameters that are valid for the stress range of interest.

Consolidation and swelling. Consolidation refers to a process during which a saturated soil experiences a decrease in water content due to a decrease in the volume of voids. The opposite process, swelling (heave) results in an increase in water content due to an increase in the volume of voids. Consolidation and heave involve flow of water into or out of the soil, and a corresponding change in volume. Changed pore pressures (for example, from a change in load or heads) establish a hydraulic gradient, since the pore pressure near the drainage boundary remains essentially unchanged. In soils with low permeability, the water flows slowly, and a time lag exists between the event that changed the pore pressures and the completion of consolidation. The compression of a partially saturated soil involves compression, solution, or expulsion of air until the soil becomes nearly saturated. The spring–dashpot example in Fig. 2 illustrates the mechanism of consolidation.

Figure 9*a* shows a layer of clay with double drainage subjected to a uniform load. After application of the load, pore pressures near the center of the layer increase by an amount equal to the magnitude of the load. The high permeability of the sand permits rapid dissipation of excess pore pressures, and deformations occur almost instantaneously. The total head difference between the center of the clay and the sand layers creates a hydraulic gradient, which produces a flow from the center outward in two directions. Figure 9*b* shows a consolidation test (known as an oedometer test) that models the one-dimensional compression experienced by the clay layer. The expression for final settlement shown in Fig. 9*d* remains valid for fine- as well as coarse-grained soils.

The oedometer tests provide information to predict both the magnitude and rate of deformation. In coarse-grained soils, deformations occur very rapidly. In fine-grained soils, the theory of consolidation helps predict the rate of deformation using the coefficient of consolidation, a parameter influenced mostly by the permeability of the soil. The theory of consolidation permits computation of more detailed results, including distribution of pore pressures and deformations throughout the deposit at any time.

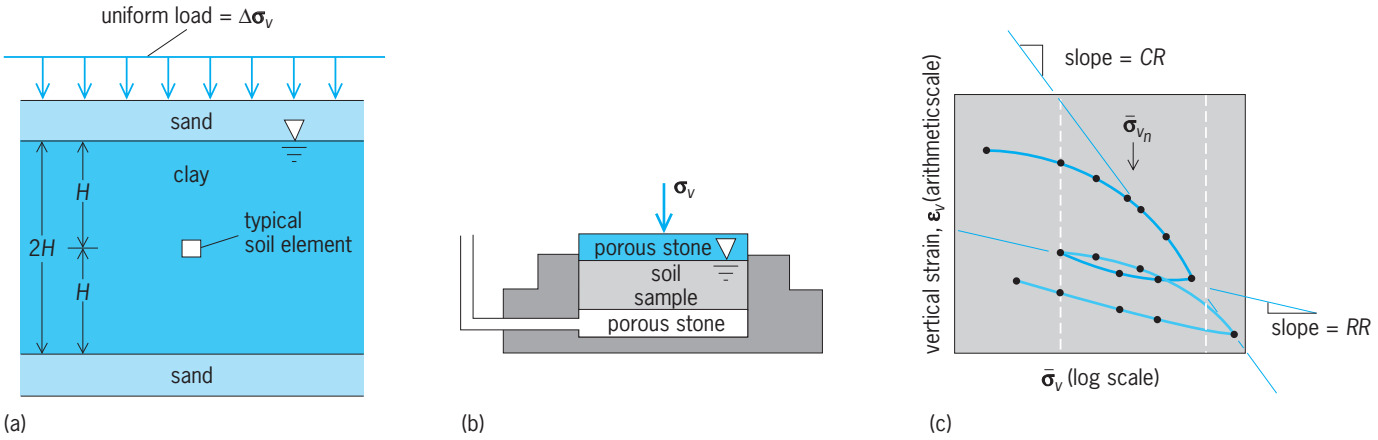


Fig. 9. Consolidation settlement. (*a*) Clay layer with double drainage, no lateral strains. (*b*) Oedometer test to obtain soil parameters. (*c*) Oedometer test results; each data point represents the end of a load increment; and left broken line represents 10 metric tons/m^2 and right broken line represents 100 metric tons/m^2. 1 metric ton = 1.1 ton; 1 m^2 = 10.8 ft^2.

Soil improvement. For many centuries, builders have tried to make soil stronger, denser, lighter, less compressible, less pervious, more pervious, or more resistant to erosion. One of the most common soil improvement techniques compacts the soil with a heavy roller or tamper. Compaction of fine-grained soils generally takes place in layers no more than 12 in. (30 cm) thick. Compaction densifies the soil, and it can increase its strength and decrease its compressibility and permeability. The key to obtaining the desired dry density lies in achieving a preselected water content in the soil undergoing compaction. Compaction at the optimum water content yields the highest dry density for a given compactive effort. Soil compacted at or near the optimum water content has negative pore pressures.

Other techniques use to improve engineering properties of soil include preloading; dewatering; stabilization with chemical admixtures; and reinforcement with rods, strips, grids, and fabrics made of natural of synthetic materials.

Slope stability. In earth slopes, the force of gravity pulls the soil mass downward. The internal friction between soil particles resists this force. As long as the resisting forces exceed the driving forces, the slope remains standing. A stability assessment of a slope involves a comparison of the stresses acting on the sliding mass and the strength of the soil along which sliding occurs. This assessment requires evaluation of the geometry and pore pressures, which together determine the stresses, and the strength of soils involved in a potential slide.

Geotechnical engineers use the term factor of safety (FS) to indicate the level of stability of a slope. Using forces, FS is the ratio of the shear resistance on a shear surface to the net shear force along a shear surface; or, using stresses, FS is the ratio of the shear strength along a shear surface to the average shear stress on a shear surface. When FS decreases to 1, failure occurs. **Figure 10** shows the forces acting on a wedge of soil. To increase FS would require an increase in shear resistance, S, or a decrease in shear force, T.

The following actions would increase S: (1) Increase W. (2) Increase $\bar{c}$ (densify, replace soil chemicals, temperature). (3) Increase $\bar{\phi}$. (4) Decrease U_s (drain).

The following actions would decrease T: (1) Decrease U_L (drain). (2) Decrease W (excavate, replace with lighter soil). (3) Decrease P_L (unload above slide). (4) Increase U_R (raise reservoir level). (5) Increase R (install tiebacks, piles, berm retaining wall).

Slope stability analyses generally use the principles of force and moment equilibrium, a technique known as limit equilibrium. In general, a free body bounded by the slope surface and a potential failure surface is isolated, and the stresses are determined by solving the static equilibrium equations. The ratio of the determined stresses and the shear strength of the soil yields the factor of safety. Repetition of the analysis with other possible failure surfaces yields different factors of safety because both the driving and resisting forces can vary from surface to surface. The most critical surface has the lowest factor of safety. Computer programs allow the rapid investigation of hundreds of potential failure surfaces along cross sections with different soil layers. *See* STATICS.

More sophisticated numerical techniques consider the stress-strain properties of the soil, and allow the analysis of slopes composed of soils with complicated stress-strain-strength properties. Finite-element stress-deformation analyses can

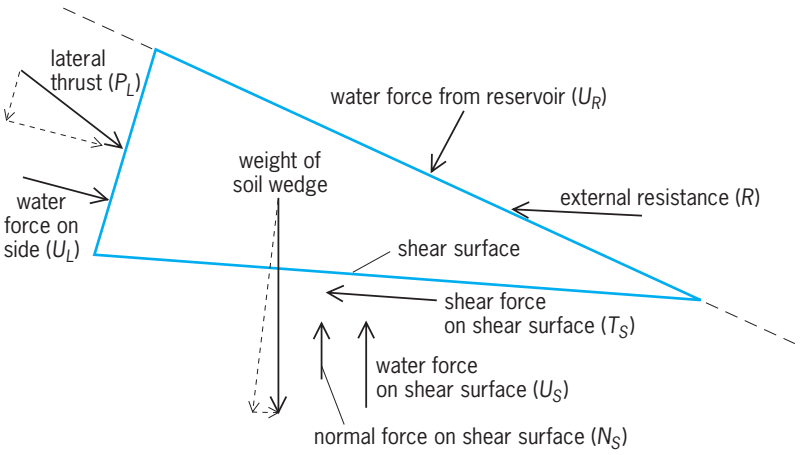


Fig. 10. Free-body diagram showing forces on a slope wedge.

provide a detailed description of the stresses acting throughout the slope. The engineer calculates the factor of safety by integrating the acting shear stress along a potential failure surface and comparing the result to the available shear strength along the surface. *See* FINITE ELEMENT METHOD.

The principles of stability analysis outlined above apply for natural slopes as well as for earth structures such as earth dams and highway embankments. With constructed facilities, the engineer has more detailed knowledge and some degree of control over the material properties. An earth structure must satisfy all the criteria needed to perform its intended function. These may include overall stability, limited deformation, weight limitations, specific flow characteristics, or particular dimensions. *See* HIGHWAY ENGINEERING.

Foundations. Nearly every large structure derives its support from contact with the earth. The foundation, whether of a building, tower, highway, retaining wall, bridge, or storage tank, transfers live and dead loads from the structure to the ground. Besides providing support, or adequate bearing capacity, the foundation must keep settlements within an acceptable range for the structure. Foundations vary from the simple preparation of a patch of ground to elaborate arrays of long steel piles driven to great depths. Other foundation types include rafts, footings, caissons, many types of piles, stone columns, grouted soil, frozen soil, mechanically reinforced soil, and hybrid combinations of these.

The bearing capacity of a foundation element depends on its ability to resist a shear failure, that is, a loss of stability. The principles of static equilibrium that govern stability of slopes also control foundation stability. The sum of the resisting forces must exceed the sum of the driving forces. *See* FOUNDATIONS.

Shallow. In practice, engineers design most shallow foundations by using semiempirical formulas that combine the concepts embodied in the theories with the experience gained by measuring the performance of actual foundations. Most shallow foundations have a large factor of safety against a general shear failure (3 or above) because restrictions on allowable settlement usually prove much more stringent and result in low allowable bearing stresses.

For either drained or undrained conditions, elastic theory predicts the stress change at any depth below a loaded area. The change in stress produces a strain in accordance with Hooke's law. The sum of the individual strains from the footing base to some depth where the stress change becomes negligible equals the settlement of the footing. In the case of shallow foundations on fine-grained soils, additional settlement will occur because of time-dependent consolidation under the applied load. Difficulties arise in selecting the appropriate elastic parameters, Young's modulus and Poisson's ratio, for the anticipated drainage and stress conditions. The stress path method, which attempts to simulate the anticipated changes in the state of stress with soil samples in the laboratory, can aid the engineer in selecting correct parameters. *See* ELASTICITY; YOUNG'S MODULUS.

Deep. Where soils of adequate strength do not lie sufficiently close to the ground surface, deep foundations can transfer the loads to stronger soil strata or distribute the loads throughout a greater mass of soil. Deep foundation elements, such as caissons or piles, can develop support along their entire length as well as at their base or tip. *See* CAISSON FOUNDATION.

Friction piles develop resistance primarily through shear resistance (skin friction) along their length. End-bearing piles develop resistance primarily at their tip by bearing on firmer soil or rock. Most piles combine the characteristics of the two pile types. Pile equipment has evolved that can drive the piles into the ground with powerful hammers, vibrate the piles through soil, or place the piles in predrilled holes. Pile widths or diameters range from about 4 in. (10 cm) for minipiles to several feet for offshore oil platforms, with lengths up to several hundred feet. Common pile materials include steel, concrete, and wood.

Caissons or drilled piers usually have greater diameters than piles and rely on end-bearing support. Their construction involves excavation instead of driving, and it often allows visual inspection of the ultimate bearing surface.

Tests and properties. Field and laboratory tests are used to determine soil properties. Those conducted in the field measure density, permeability, shear strength, penetration resistance, compressibility, and deposit boundaries. Those done in the laboratory measure water content, expansion index, grain size distribution, permeability, shear strength, dynamic stress, resonant frequency, and mineral composition.

The broad range of geotechnical instrumentation includes piezometers for measuring ground-water pressure, earth pressure cells for measuring total stress in soil, settlement plates for measuring vertical displacements, extensometers for determining deformation, tiltmeters for measuring inclination, inclinometers for measuring lateral deformation versus depth, piezoelectric transducers for measuring acoustic emissions, load cells for measuring load in a structural member, strain gages for measuring strain in a structural member, and various devices for measuring temperature.

Other soil engineering problems. A naturally occurring material such as soil presents geotechnical engineers with several particularly difficult problems. In addition to the complexities of working with a nonhomogeneous material with nonlinear anisotropic stress–strain properties, the strength varies from point to point, may vary with time or deformation, and remains difficult to measure precisely. The simple examples that have been discussed illustrate the fundamentals of soil mechanics. Engineers routinely extend and apply these fundamentals to the complicated conditions encountered in the field. Some of these applications include two- and three-dimensional problems; dynamic loading of soil; construction of earthquake-resistant facilities; earth pressures and retaining walls, tunnels, and excavations; subsidence; problems involving expansive soils; field exploration; probability and reliability;

and the use of synthetic materials such as fabrics, membranes, and superlight fills. *See* CLAY; CONSTRUCTION METHODS; ENGINEERING GEOLOGY; RETAINING WALL; SAND; SOIL; TUNNEL.

Francisco Silva-Tulla

Bibliography. American Society for Testing and Materials, *Annual Book of ASTM Standards*, sect. 4: *Construction*, vol. 04.08: *Soil and Rock; Building Stones, Geotextiles*, 1989; H. R. Cedergren, *Seepage, Drainage and Flow Nets*, 3d ed., 1997; R. F. Craig, *Soil Mechanics*, 6th ed., 1997; B. M. Das, *Fundamentals of Soil Dynamics*, 1983; J. Dunnicliff, *Geotechnical Instrumentation for Monitoring Field Performance*, 1988; T. W. Lambe and R. V. Whitman, *Soil Mechanics*, 1969; J. K. Mitchell, *Fundamentals of Soil Behavior*, 2d ed., 1993; K. Terzaghi and R. B. Peck, *Soil Mechanics in Engineering Practice*, 3d ed., 1996; U.S. Navy, *Soil Mechanics*, Design Man. 7.01, 1986.

Soil microbiology

The study of biota that inhabit the soil and the processes that they mediate. The soil is a complex environment colonized by an immense diversity of microorganisms. Soil microbiology focuses on the soil viruses, bacteria, actinomycetes, fungi, and protozoa, but it has traditionally also included investigations of the soil animals such as the nematodes, mites, and other microarthropods. These organisms, collectively referred to as the soil biota, function in a belowground ecosystem based on plant roots and litter as food sources.

Historically, soil microbiology has been concerned with the role of microorganisms in maintaining the soil's productivity for plant growth. Plant growth is often limited by the availability of nitrogen, and many of the early investigations focused on microorganisms that transform nitrogen into forms usable by plants. S. Winogradsky first isolated the bacteria responsible for the conversion of ammonium (NH_4^+) to nitrate (NO_3^-) in 1890. Since Winogradsky also advanced the understanding of microbial transformations of sulfur and iron and developed important concepts in classifying soil organisms, he is regarded as the founder of soil microbiology. Nitrogen fixation is a microbial process whereby atmospheric nitrogen (N_2) is converted to a form available for plants; it is the key to the nitrogen supply of leguminous plants (that is, clover, alfalfa, and soybean) and some woody species (such as alder and *Ceanothus*). Soil microorganisms have also been studied because of their abilities to produce medically useful antibiotics and to degrade potentially toxic pollutants. Modern soil microbiology represents an integration of microbiology with the concepts of soil science, chemistry, and ecology to understand the functions of microorganisms in the soil environment.

Microorganisms. The surface layers of soil contain the highest numbers and variety of microorganisms, because these layers receive the largest amounts of potential food sources from plants and animals. The soil biota form a belowground system based on the energy and nutrients that they receive from the decomposition of plant and animal tissues. The primary decomposers are the bacteria and fungi. Bacteria are single-celled prokaryotes and are the most numerous of the microorganisms in soil. Typically, there are about 1 billion bacteria representing thousands of different bacterial types in a single gram of soil. Bacteria are classified according to their sources of energy and carbon. Bacteria that decompose organic materials for carbon and energy are known as organotrophs or heterotrophs. Chemolithotrophic bacteria use alternative chemical sources of energy and inorganic carbon, and they are important in the transformations of nutrients and metals.

Fungi are the eukaryotic organisms commonly known as molds, mildew, yeasts, and mushrooms. Soil fungi exist primarily as thin, long filaments called hyphae. Many forest soils contain more than a kilometer of fungal hyphae in a single gram of soil. The mushrooms and fungi visible on the surface are the reproductive structures produced by the vast network of hyphae in the soil or woody material. The protozoa are single-celled, mobile eukaryotes such as the ciliate paramecia and the amebas. Nematodes (round worms) and the microarthropods are representatives of the soil animals. The protozoa and the soil animals graze or prey on the smaller soil organisms or on organic materials. Algae and lichens exist near the soil surface because they require light for photosynthesis. The distribution of soil organisms often depends on the size of the pore spaces between soil particles and the availability of water. *See* FOREST SOIL.

Ecological roles. Microorganisms, especially algae and lichen, are pioneering colonizers of barren rock surfaces. Colonization by these organisms begins the process of soil formation necessary for the growth of higher plants. After plants have been established, decomposition by microorganisms recycles the energy, carbon, and nutrients in dead plant and animal tissues into forms usable by plants. Therefore, microorganisms have a key role in the processing of materials that maintain life on the Earth. The transformations of elements between forms are described conceptually as the elemental cycles.

In the carbon cycle, microorganisms transform plant and animal residues into carbon dioxide and the soil organic matter known as humus. Humus improves the water-holding capacity of soil, supplies plant nutrients, and contributes to soil aggregation. Microorganisms may also directly affect soil aggregation. For example, fungal hyphae can physically bind microaggregates together, while bacteria produce polysaccharides that can bind soil particles into microaggregates. The extent of soil aggregation determines the workability or tilth of the soil. A soil with good tilth is suitable for plant growth because it is permeable to water, air, and roots. *See* HUMUS.

Soil microorganisms play key roles in the nitrogen cycle. The atmosphere is approximately 80% nitrogen gas (N_2), a form of nitrogen that is available to plants only when it is transformed to ammonia (NH_3) by either soil bacteria (N_2 fixation) or by

humans (manufacture of fertilizers). The N_2-fixing symbiotic bacteria, *Rhizobium*, form root nodules on legumes such as soybean, peas, beans, and alfalfa. Nitrogen-fixing crops are a key component of agricultural productivity, and their use decreases reliance on chemical fertilizers. Certain chemolithotrophic bacteria transform NH_4^+ to NO_3^- in the process called nitrification. Nitrate is very mobile in the soil and is readily available for plant assimilation. However, NO_3^- may also be lost from the soil through the processes of denitrification and leaching. Soil bacteria also mediate denitrification, which returns nitrogen to the atmosphere by transforming NO_3^- to N_2 or nitrous oxide (N_2O) gas. Microorganisms are crucial to the cycling of sulfur, phosphorus, iron, and many micronutrient trace elements.

In addition to the elemental cycles, there are several interactions between plants and microbes which are detrimental or beneficial to plant growth. Some soil microorganisms are pathogenic to plants and cause plant diseases such as root rots and wilts. Many plants form symbiotic relationships with fungi called mycorrhizae (literally fungus-root). Mycorrhizae increase the ability of plants to take up nutrients and water. The region of soil surrounding plant roots, the rhizosphere, may contain beneficial microorganisms which protect the plant root from pathogens or supply stimulating growth factors. The interactions between plant roots and soil microorganisms is an area of active research in soil microbiology. *See* BIOGEOCHEMISTRY; MYCORRHIZAE; NITROGEN CYCLE; NITROGEN FIXATION; NITROGEN OXIDES; RHIZOSPHERE.

Biotechnology and soil microorganisms. The incredible diversity of soil microorganisms is a vast reserve of potentially useful organisms. Antibiotics are compounds produced by microorganisms which can kill or inhibit the growth of other microorganisms. Many of the medically important antibiotics are produced by filamentous bacteria known as actinomycetes. The genus *Streptomyces* alone produces more than 3000 antibiotics. Fungi such as *Penicillum* also produce a wide variety of antimicrobial compounds. The soil is the largest reservoir of these medically important microorganisms. *See* ANTIBIOTIC.

The numerous natural substances that are used by microorganisms indicate that soil microorganisms have diverse mechanisms for degrading a variety of compounds. Human activity has polluted the environment with a wide variety of synthetic or processed compounds. Many of these hazardous or toxic substances can be degraded by soil microorganisms. This is the basis for the treatment of contaminated soils by bioremediation, the use of microorganisms or microbial processes to detoxify and degrade environmental contaminants. Often the degradation of contaminants by microorganisms can be stimulated by manipulating the environmental conditions, for example, by aeration. In some cases, adding specific degrading microorganisms can accelerate the breakdown of the pollutants. Recalcitrant compounds are extremely resistant to microbial degradation and may take many decades to degrade into harmless byproducts. Soil microbiologists study the microorganisms, the metabolic pathways, and the controlling environmental conditions that can be used to eliminate pollutants from the soil environment. *See* HAZARDOUS WASTE.

Methods for study. Microbiologists traditionally isolate pure strains of microorganisms by using culture methods. However, generally less than 10% of the soil microorganisms are recoverable by culture methods, and many soil microorganisms remain poorly characterized or unknown. Methods that do not rely on culturing microorganisms include microscopic observation and biochemical or genetic analysis of specific cell constituents. Microscopic methods generally use fluorescent stains so that microbial cells are visible against the opaque soil particles. Alternatively, cells can be separated from soil particles and then viewed. Fluorescent stains can be linked to antibodies that react only to a specific soil microorganism, a technique known as immunofluorescence, to obtain counts of a specific microorganisms without culturing.

Specific cell constituents, such as the fatty acids present in the cell membranes, can be extracted for biochemical analysis. Sequences of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) can also be used to identify specific microorganisms or groups of microorganisms. A known DNA or RNA sequence can be used as a probe to determine the amount of a specific gene in a soil population. The gene probe is labeled chemically or radioactively and incubated with the target DNA or RNA extracted from the soil. After the gene probe binds to identical sequences, genes can be detected and quantified. These techniques have recently been used to characterize the diversity of soil microorganisms that are not cultured with standard methods. *See* DEOXYRIBONUCLEIC ACID (DNA); RIBONUCLEIC ACID (RNA).

The rates or controlling factors for microbial processes are studied by using methods from chemistry, biology, and ecology. Typically, these studies involve measuring the rate of production and consumption of a compound of interest. Many investigators use stable or radioactive isotopes to trace the transformations of the compound. The results of these studies are commonly analyzed by using mathematical models. Models allow the information from one system to be generalized for different environmental conditions. *See* FLUORESCENCE MICROSCOPE; IMMUNOFLUORESCENCE; MICROBIOLOGY; MOLECULAR BIOLOGY; SOIL; SOIL CHEMISTRY; SOIL ECOLOGY.

Jeanette M. Norton

Bibliography. M. Alexander, *Soil Microbiology*, 2d ed., 1977; R. M. Atlas and R. Bartha, *Microbial Ecology*, 4th ed., 1997; D. P. Labeda, *Isolation of Biotechnological Organisms from Nature*, 1990; F. B. Metting, *Soil Microbial Ecology*, 1993; E. A. Paul and F. E. Clark, *Soil Microbiology and Biochemistry*, 2d ed., 1996; R. L. Tate, *Soil Microbiology*, 2d ed., 2000.

Soil sterilization

A chemical or physical process that results in the death of soil organisms. This control method affects many organisms, even though the elimination of only

specific weeds, fungi, bacteria, viruses, nematodes, or pests is desirable. Even if complete sterilization is achieved, it is short lived since organisms will recolonize this biological vacuum quite rapidly. Soil sterilization can be achieved through both physical and chemical means. Physical control measures include steam and solar energy. Chemical control methods include herbicides and fumigants. Dielectric heating and gamma irradiation are used less frequently as soil sterilization methods. Composting can be used to sterilize organic materials mixed with soil, but it is not used for the sterilization of soil alone. Soil sterilization is used in greenhouse operations, the production of high-value or specialty crops, and the control of weeds.

Steam. Heating the soil with steam to 180°F (82°C) for 30 min will kill most organisms except a few heat-tolerant viruses and weed seeds. These heat-tolerant organisms can be difficult to eliminate even at 212°F (100°C). However, heating soil to temperatures greater than 176°F (80°C) has disadvantages. These high temperatures require more energy, create a biological vacuum, and leave the soil toxic to plant growth because of excessive exchangeable manganese, nitrates, and ammonia. These toxic materials must be leached from soil with water, which can result in environmental pollution problems. An alternative to sterilization is pasteurization. The pasteurization of soil with aerated steam at 158°F (70°C) kills most plant pathogens without leaving a biological vacuum. The saprophytic microflora that survive pasteurization leave the soil more suppressive to plant pathogens. Pasteurization also avoids toxicity problems. Steam can be applied to soil by releasing it directly under plastic tarpaulins supported by nylon netting, or through perforated pipes buried 18–23 in. (45–60 cm) deep in the soil. A common problem in steaming a large ground bed area is using only a single steam inlet. An area no longer than 66 ft (20 m) and no wider than 16 ft (5 m) should be sterilized with one inlet.

Negative-pressure steaming is an alternative to conventional steaming and requires less energy. In this method a fan is used to set up a negative pressure within buried pipes so that steam supplied at the soil surface is actively moved downward. This arrangement provides more uniform and deeper distribution of heat than sheet steaming. *See* SOIL CHEMISTRY.

Solarization. Solarization is a simple technique that captures radiant heat energy from the Sun. This energy causes physical, chemical, and biological changes in the soil. These changes pasteurize the soil, leading to control or suppression of soil-borne plant pathogens such as fungi, bacteria, nematodes, and pests, along with weed seed and seedlings. Solarization is accomplished by placing a clear plastic tarpaulin over the soil surface and anchoring the edges by burying them in a trench around the area to be treated. The Sun's radiant energy then heats the soil beneath the plastic. Solarization is most effective when done during June and July in western states such as California, when the day lengths are long, air temperatures are high, and skies are clear. Depending on geographic location, solarization can also be effective in May, August, and September. The plastic should be left in place for 4–6 weeks to allow heating to reach greater depths. The clear plastic tarpaulin can be construction-grade plastic and vary in thickness from 1 to 6 mils (0.001 + 0.006 in. or 0.0025 to 0.015 cm). The thinner plastic allows better soil heating, since it reflects less solar energy. For some applications, ultraviolet-resistant plastic should be considered. This plastic is more expensive but it can be used more than once.

Efficacy can be further improved by combining solarization with allelopathy, which involves incorporating green plant material from crucifers into the soil. The gases released from this plant material are trapped under the plastic, increasing the efficacy of the solarization treatment.

Solarization can be done on a small or large scale. Large areas can be solarized by connecting a series of tarpaulins with long-lasting, heat-resistant glue. This technique can even be applied postplant to orchards with heat-tolerant plants such as pistachio. When the treatment is complete, removal of the plastic can be troublesome, especially on large acreage; if the plastic is left in place too long, it becomes brittle and begins to tear. Solarization is a safe and effective means to control soil-borne plant pathogens and is especially well suited for use by home gardeners and organic growers. On a large scale, this technique is economical only on high-value crops.

Herbicides. In the early 1900s, scientists began testing materials for the purpose of killing perennial weeds and sterilizing the soil. Salt and copper sulfate were tested, but the arsenicals proved the most satisfactory. The organic compounds introduced in the mid-1940s brought an end to the use of nonorganic arsenicals, since they were safer to use, had less potential for long-term harm to the environment, and provided a broader spectrum of weed control.

As agriculture became more mechanized, chemical weed control began to play an integral part in large-scale crop production. The herbicides used in crop production tend to be selective herbicides (temporary soil sterilants) and not long-term broad-spectrum sterilants. The broad-spectrum sterilants tend to be used more for weed control along railroads, highways, driveways, irrigation ditches, pipelines, power lines, and industrial sites. *See* HERBICIDE.

Fumigation. Sterilization by fumigation depends on infiltrating the soil with the vapor of a volatile chemical. Some of the chemicals have a broad spectrum of activity against soil microflora, fauna, and weeds, while others are toxic to fewer organisms. Like pasteurization with steam, fumigation usually leaves a fumigant-tolerant microflora, particularly at depth and at the edges of the treated area. Soil in a confined area (a garbage can or a plastic-covered area on a cement pad) can be fumigated much more thoroughly. Most fumigants require 2–3 days to kill organisms in the soil, depending on soil type and temperature. In order to keep the fumigants from

escaping, they are injected or incorporated into the soil, which is immediately covered with plastic or sealed by rolling. Some fumigants such as metham sodium can be applied in irrigation water through a sprinkler system. After 7–14 days the soil can be worked to allow the gas to escape. Some fumigants require a waiting period of several weeks before planting or seeding, to ensure that the fumigant has escaped or degraded. Nitrogen fertilization may also have to be adjusted since nitrifying bacteria are readily killed by fumigation. Leaching of the soil may also be required with some fumigants such as methyl bromide.

Fumigants are available as a gas, gas compressed to liquid, liquids alone or in mixtures, or granules. The combination of methyl bromide plus chloropicrin has strong activity against bacteria, fungi, nematodes, insects, and weeds. The other fumigants have a more limited range of activity. Some fumigants such as methyl bromide may move 20–32 in. (50–80 cm) in the soil from the point of injection, while other fumigants such as metham sodium may move only 4 in. (10 cm). The distance that vapors move in the soil is determined not only by the chemical applied but also by the size and continuity of soil pores, soil temperature, soil moisture, and soil type. The use of fumigants is expensive and generally limited to use on high-value crops such as potatoes, strawberries, bedding plants, and tobacco. Some fumigants are extremely toxic to humans and animals and must be handled with great caution. *See* FUMIGANT.

Soil sterilants and other pesticides have been closely regulated in the United States by the Environmental Protection Agency (EPA) under the Federal Insecticide, Fungicide, and Rodenticide Act and the Federal Food, Drug, and Cosmetic Act. More recently, the Montreal Protocol and the Food Quality Protection Act (FQPA) have created considerable concern for pesticide users since they will potentially eliminate many pesticides, including some soil sterilants. The pesticide methyl bromide has been included in the Montreal Protocol, an international environmental treaty, as an ozone depleter. Thus the EPA, based on the Clean Air Act, has frozen production levels of methyl bromide and set a schedule to phase it out of production by the year 2005. The loss of methyl bromide will be disruptive since the production, storage, shipping, or processing of nearly 100 crops depends on this pesticide. Methyl bromide is an effective herbicide, nematicide, insecticide, and fungicide and was used commercially in the United States for soil fumigation and quarantine purposes for most of the twentieth century. The pending elimination of methyl bromide has stimulated creative research that should lead to viable alternatives. However, finding a single alternative pesticide to replace the many uses of methyl bromide is very unlikely. Research will probably pursue ecologically based alternatives such as organic crop production, improved cultivars, organic soil amendments, biological controls, better crop rotations, soil solarization, better pesticide application methods, and the judicious use of low-risk chemical alternatives. Similar ecologically based alternatives will be needed for other pesticides when the EPA finishes implementing the FQPA and evaluating commercial products, particularly those classified as carbamates, organophosphates, and group B2 carcinogens (for example, iprodione). *See* INSECTICIDE; NEMATICIDE; PESTICIDE.

General considerations. The soil sterilization methods are improved by the presence of good soil moisture (not saturated soil). Good moisture enhances the probability that microorganisms will be in a metabolically active state, making them more susceptible to control measures. Sterilization methods are also improved by good soil tilth and the elimination of large pieces of plant debris. Some chemicals used for sterilization have soil temperature requirements, especially fumigants.

Soil sterilization through the use of chemicals is frequently the most practical method of weed control. However, soil sterilization for the control of soil-borne plant pathogens is an expensive control measure that affects many nontarget organisms. For control of soil-borne plant pathogens, consideration should be given to using safer and more cost-effective measures such as resistant plants, certified disease-free planting material, crop rotation, sanitation, seed treatments, fungicide soil drenches, and cultural practices (proper tillage, plant spacing, fertilization, and weed control). Prior to initiating control, a trained individual such as a county agent or university specialist should be consulted, to identify the target organism, recommend the safest and most cost-effective control measure, and help optimize effectiveness. *See* AGRICULTURAL CHEMISTRY; AGRICULTURAL SCIENCE (PLANT); FUNGISTAT AND FUNGICIDE.

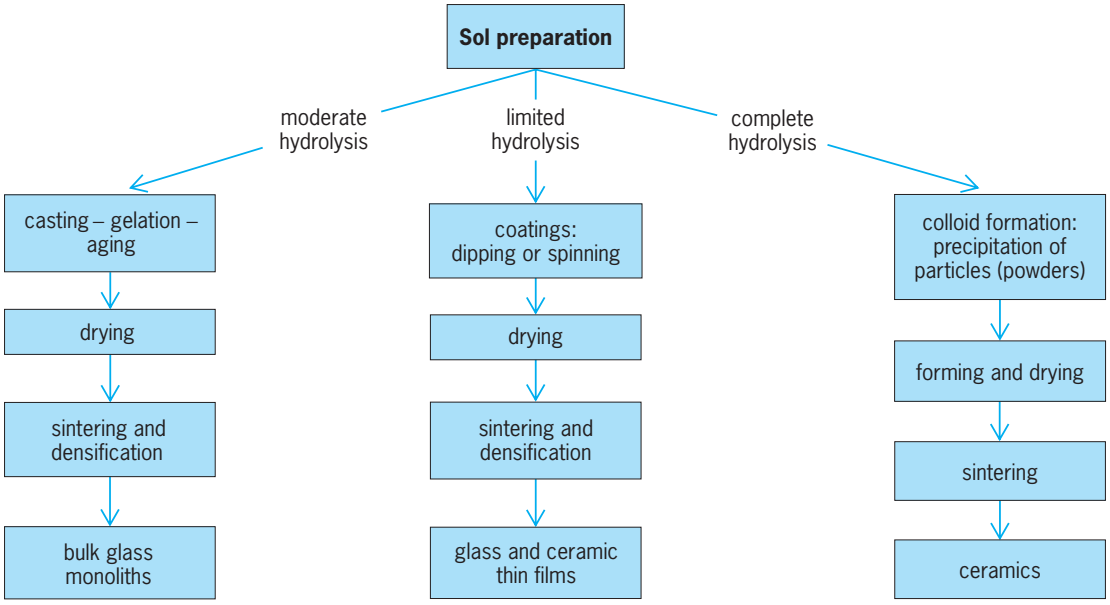
Carl A. Strausbaugh

Bibliography. A. S. Crafts, *Modern Weed Control*, 1975; W. R. Jarvis, *Managing Diseases in Greenhouse Crops*, 1992; J. B. Ristaino and W. Thomas, Agriculture, methyl bromide, and the ozone hole: Can we fill the gaps?, *Plant Dis.*, 81:964–977, 1997.

Sol-gel process

A chemical synthesis technique for preparing gels, glasses, and ceramic powders. The synthesis of materials by the sol-gel process generally involves the use of metal alkoxides, which undergo hydrolysis and condensation polymerization reactions to give gels (see **illustration**).

The production of glasses by the sol-gel method is an area that has important scientific and technological implications. For example, the sol-gel approach permits preparation of glasses at far lower temperatures than is possible by using conventional melting. It also makes possible synthesis of compositions that are difficult to obtain by conventional means because of problems associated with volatilization, high melting temperatures, or crystallization. In addition, the sol-gel approach is a high-purity process that leads to excellent homogeneity. The sol-gel approach is adaptable to producing films and fibers as well as bulk pieces, that is, monoliths (solid



Flow chart showing the sol-gel methods used to prepare bulk glasses, thin films of glass or ceramics, and powders; the powders are then sintered to produce ceramics.

materials of macroscopic dimensions, at least a few millimeters on a side).

Glass formation. Formation of silica-based materials is the most widely studied system. An enormous range of multicomponent silicate or nonsilicate compositions have also been prepared.

The sol-gel process can ordinarily be divided into the following steps: forming a solution, gelation, drying, and densification. The preparation of a silica glass begins with an appropriate alkoxide such as tetraethylorthosilicate [$Si(OC_2H_5)_4$; TEOS], which is mixed with water (H_2O) and a mutual solvent, such as ethanol (C_2H_5OH), to form a solution. Tetramethylorthosilicate [$Si(OCH_3)_4$; TMOS] is also widely used as a precursor alkoxide. Hydrolysis leads to the formation of silanol groups (Si—OH). These species are only intermediates. Subsequent condensation reactions produce siloxane bonds (Si—O—Si) plus the by-product alcohol (ROH) or water. An example of the formation of a silica gel by the organometallic route is shown in reactions (1)–(3) [R is an alkyl group: $R = C_xH_{2x+1}$].

Hydrolysis and the formation of silanol groups:

$$\equiv Si—OR + H_2O \rightarrow \equiv Si—OH + ROH \qquad (1)$$

Alcohol condensation and the formation of siloxane groups:

$$\equiv Si—OR + HO—Si\equiv \rightarrow \equiv Si—O—Si\equiv + ROH \quad (2)$$

Water condensation and the formation of siloxane groups:

$$\equiv Si—OH + HO—Si\equiv \rightarrow \equiv Si—O—Si\equiv + H_2O \quad (3)$$

The silica gel formed by this process leads to a rigid, interconnected three-dimensional network consisting of submicrometer pores and polymeric chains. During the drying process (at ambient pressure), the solvent liquid is removed and substantial shrinkage occurs. The resulting material is known as a xerogel. When solvent removal occurs under hypercritical (supercritical) conditions, the network does not shrink and a highly porous, low-density material known as an aerogel is produced. Heat treatment of a xerogel at elevated temperature produces viscous sintering (shrinkage of the xerogel due to a small amount of viscous flow) and effectively transforms the porous gel into a dense glass.

Although the above discussion is based on silica, the sol-gel process is widely applicable to other inorganic oxides. By using various metal alkoxides $M(OR)_n$, where M is Al, Ti, V, Cr, Mo, W, Zr, and so on and OR is an alkoxy functionality as described previously, different oxides can be prepared. The resulting xerogels are either amorphous or nanocrystalline. Heat treatment of these materials leads to crystallization rather than the formation of a dense glass as occurs in the case of silica.

Hydrolysis and condensation. In general, the processes of hydrolysis and condensation polymerization are difficult to separate. The hydrolysis of the alkoxide need not be complete before condensation starts; and in partially condensed silica, hydrolysis can still occur at unhydrolyzed sites. Several parameters have been shown to influence the hydrolysis and condensation polymerization reactions, including the temperature, the solution pH, the particular alkoxide precursor, the solvent, and the relative concentrations of each constituent. In addition, acids and bases catalyze the hydrolysis and condensation polymerization reactions; therefore, they are added to help control the rate and extent of these reactions.

It is generally agreed that hydrolysis occurs by bimolecular nucleophilic displacement reactions. Under acidic conditions, an alkoxide group is protonated rapidly. As electron density is withdrawn from the silicon (Si), the silicon becomes more electrophilic and thus susceptible to attack by water.

Under basic conditions, water dissociates to rapidly form nucleophilic hydroxyl anions (OH^-). The hydroxyl anion then attacks the silicon atom. In one proposed mechanism for base hydrolysis, OH^- replaces OR^-, where OR^- is an alkoxide ion.

A number of factors influence hydrolysis. Under acidic conditions, such as in the presence of hydrochloric acid (HCl), the hydrolysis rate increases with alkyl substitution (electron-providing groups), while the reverse trend occurs with basic conditions, such as in the presence of ammonia (NH_3). The effect of the H_2O/Si ratio is to promote hydrolysis [reaction (1)] when the ratio is increased, while the alcohol-producing condensation mechanism [reaction (2)] is favored at less than stoichiometric amounts of water ($H_2O/Si < 4$). It should also be noted that the hydrolysis reaction may proceed in the reverse direction; in this reesterification, an alcohol displaces a hydroxyl group to produce water.

Condensation polymerization reactions also involve acid- or base-catalyzed bimolecular nucleophilic substitution reactions. The acid-catalyzed condensation mechanism involves protonated silanols, while the base-catalyzed mechanism involves deprotonated silanols. Electron-providing alkyl groups will reduce the acidity of the corresponding silanol and thus influence the pH dependence of the condensation mechanism. The nature of the solvent may also influence condensation. The ability of the solvent to encourage depolymerization has a significant effect on the morphology of the polymer structures produced. In addition, hydrogen bonding of the solvent (protic or aprotic) to the different silanols can either promote or retard condensation. *See* HYDROGEN BOND; HYDROLYSIS; POLYMERIZATION.

Microstructural development. The conditions under which hydrolysis and condensation occur have a profound effect on gel growth and morphology. These structural conditions greatly influence the processing of sol-gel glasses into various forms. It is well established that acid-catalyzed solutions with low water content (that is, conditions that produce linear polymers) offer the best type of solution for producing fibers. In the case of sol-gel films, control of precursor structure (that is, weakly branched to highly branched) is of vital importance in controlling the microstructure and the resulting properties of the film. Silicate gels prepared at low pH (<3) and low water content (<4 mol water per mol alkoxide) produce primarily linear polymers with low crosslink density. Additional crosslinks form during gelation, and the polymer chains become increasingly entangled. Silicate gels prepared under more basic conditions (pH 5–7) or higher water contents produce more highly branched clusters, which behave as discrete species. Gelation occurs by linking clusters together. At still higher pH and excess water content, colloidal silica is formed. *See* CATALYSIS; COLLOID; GEL.

Gelation and aging. As the hydrolysis and condensation polymerization reactions continue, viscosity increases until the solution ceases to flow. The time required for gelation to occur is an important characteristic that is sensitive to the chemistry of the solution and the nature of the polymeric species. This sol-to-gel transition is irreversible, and there is little if any change in volume. At this stage, the one-phase liquid is transformed into a two-phase system. The gels formed in alcohol are known as alcogels, while those formed in aqueous solutions are hydrogels. After gelation, gels are generally subjected to an aging process for a period of hours to days. The gels are kept in sealed containers, and very little solvent loss or shrinkage occurs. Syneresis (spontaneous separation of liquid from a gel) is frequently observed at this stage. During aging of silica gels, the conden\-sation polymerization reactions continue, increasing the degree of crosslinking in the network and producing an increase in gel viscosity. The mechanical strength and rigidity of the gel thus increase during aging.

Drying. The drying process involves the removal of the liquid phase; the gel transforms from an alcogel to a xerogel. Low-temperature evaporation is frequently employed, and there is considerable weight loss and shrinkage. The drying stage is a critical part of the sol-gel process. As evaporation occurs, drying stresses arise that can cause catastrophic cracking of bulk materials. The ability to dry without cracking serves to limit the sizes of monolithic pieces produced by the sol-gel method.

Much is known about the drying process, as it has been characterized by physical measurements, mechanical properties, and chemical methods. A comprehensive analytical model has been developed that has identified the various details related to the generation of drying stresses in terms of their origin, their spatial distribution, and how stresses lead to cracking and warping in gels. This model indicates that as the gel network becomes stiffer, shrinkage stops and drying stresses arise because of spatial variability in strain rate; that is, gel shrinkage is faster at the exterior surface than in the interior. It is at this point that maximum stress is developed at the surface of the gel and cracking is prone to occur. The analysis indicates that stresses may be minimized by a low rate of solvent evaporation or the use of additives to improve solvent permeability.

Densification. The final stage of the sol-gel process is densification, where the gel-to-glass conversion occurs and the gel achieves the properties of the glass. As the temperature increases, several processes occur, including elimination of residual water and organic substances, relaxation of the gel structure, and, ultimately, densification. At moderate temperatures (above 150°C or 302°F but below the glass transition temperature T_g), skeletal densification occurs from both condensation polymerization processes and structural relaxation. These processes account for considerable shrinkage and reduction in surface area, but the resulting material is still quite porous (>25% porosity). At higher temperatures (starting at temperatures close to T_g, viscous sintering becomes the predominant factor in removing porosity. In this way, fully dense sol-gel glasses can be produced without melting, that is, merely by heating to temperatures close to the glass transition temperature in order to obtain viscous sintering. The precise temperatures involved for producing dense

sol-gel glasses depend upon the specific system. In the case of silica materials, temperatures corresponding to a viscosity of 10^{13} poises (≈1200°C or 2200°F) are generally required in order to obtain dense sol-gel silica glass. *See* GLASS; SINTERING.

Applications. The sol-gel process offers advantages for a broad spectrum of materials applications. The types of materials go well beyond silica and include inorganic compositions that possess specific properties such as ferroelectricity, electrochromism, or superconductivity. The most successful applications utilize the composition control, microstructure control, purity, and uniformity of the method combined with the ability to form various shapes at low temperatures. Films and coatings were the first commercial applications of the sol-gel process. The development of sol-gel-based optical materials has also been quite successful, and applications include monoliths (lenses, prisms, lasers), fibers (waveguides), and a wide variety of optical films.

One application of sol-gel materials that has emerged in recent years is that of bioactive glasses for biosensors. Based on the guest-host concept, a wide range of biomolecules, including globular and membrane-bound proteins and enzymes, have been immobilized in the pores of sol-gel glasses, resulting in a material that is not only bioactive but mechanically, thermally, and chemically stable. Moreover, silica-based sol-gel materials have excellent optical transparency, making optical sensors possible. Owing to the nature of the sol-gel process, the amorphous solid network forms around the biomolecules, confining them within the pores of the solid-state matrix. Although the relatively large biomolecules are immobilized within the inorganic network, small ions and molecules are able to diffuse into and out of the network. The sol-gel encapsulated biomolecules retain their native biological activity and spectroscopic properties, and they can respond to chemical changes in their environment. These materials are being widely explored for long-term biomolecular storage and for a new generation of biomedical and chemical sensors. *See* BIOSENSOR; MICROSENSOR.

Aerogels are a class of sol-gel materials that possess very low density and high surface area. Aerogels are mesoporous solids in which nanometer-scale solid domains are networked with a high volume of continuous pores. The synthesis of aerogels is based on traditional sol-gel chemistry and the resulting solids offer unusual properties of very large surface areas (on the order of 1000 m^2/g) and high pore volumes (greater than 95%). Aerogel materials have been known since the 1930s and widely used as catalysts, catalyst supports, and thermal insulation materials. In recent times, the applications of aerogels have expanded to electrochemical devices, low dielectric constant materials for integrated circuits, and the capture of cosmic dust particles.

A key step in the synthesis of aerogels is the process of removing the liquid phase from the pores of the wet gel. If the pore liquid is allowed to evaporate, the volume of the material shrinks significantly due to the capillary pressure and accompanying collapse of the gel network, resulting in a xerogel. On the other hand, if the pore liquid is removed using methods where capillary stresses do not develop or are minimized, the porosity of the wet gel can be maintained with relatively little volume shrinkage, and an aerogel is produced. Traditionally, aerogels are prepared through supercritical drying; however, ambient-pressure drying using nonpolar solvents with low surface tension has also been widely used, with the resulting materials called ambigels.

Bruce Dunn; Esther H. Lan

Bibliography. C. J. Brinker, D. E. Clark, and D. R. Ulrich (eds.), *Better Ceramics Through Chemistry I-IV*, vol. 32, 1984, vol. 73, 1986, vol. 121, 1988, vol. 180, 1990; C. J. Brinker and G. Scherer, *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, 1990; L. C. Klein (ed.), *Sol-Gel Optics: Processing and Applications*, 1994; L. C. Klein (ed.), *Sol-Gel Technology for Thin Films, Preforms, Electronics and Specialty Shapes*, 1988; J. D. Mackenzie and D. R. Ulrich, *Ultrastructure Processing of Advanced Ceramics*, 1988; W. Yuming and A. Wagendristel, *An Introduction to the Physics and Technology of Thin Films*, 1994.

Solanales

An order of flowering plants, division Magnoliophyta, in the euasterid I group of the asterid eudicotyledons. The order consists of two large and three small families, of approximately 4275 species. Solanaceae and Convolvulaceae account for all but 25 of the species. Solanales are generally characterized by sympetalous flowers (in early developmental stages only in *Montinia*) with a superior ovary, and alternate leaves. *See* ASTERIDAE; FLOWER; FRUIT; LEAF; MAGNOLIOPHYTA; MAGNOLIOPSIDA.

Solanaceae, approximately 2600 species, are cosmopolitan herbs, shrubs, lianas, and trees, with branched hairs and often spines and alkaloids. The family is of great economic significance, yielding potatoes, tomatoes, and eggplant (*Solanum*), peppers (*Capsicum*), tobacco (*Nicotiana*) and many ornamentals. Deadly nightshade (*Atropa*), jimson weed (*Datura*), and henbane (*Hyoscyamus*) are well-known poisonous members of the family. *See* BELLADONNA; EGGPLANT; ORNAMENTAL PLANTS; PEPPER; POTATO, IRISH; SOLANALES; TOBACCO.

Convolvulaceae, approximately 1650 species, include herbaceous and woody members that are often climbers. The main economic crop is sweet potato (*Ipomoea*), but the family also includes ornamentals (morning glory, *Ipomoea*; and *Convolvulus*), noxious weeds (*Calystegia* and *Convolvulus*), and parasites (*Cuscuta*). *See* POTATO, SWEET; WEEDS.

Michael F. Fay

Solar cell

A semiconductor electrical junction device which directly and efficiently absorbs the radiant energy of sunlight and converts it into electrical energy. Solar cells may be used individually as light

detectors, for example in cameras, or connected in series and parallel to obtain the required values of current and voltage for electric power generation.

Most solar cells are made from single-crystal silicon and have been very expensive for generating electricity, but have found application in space satellites and remote areas where low-cost conventional power sources have been unavailable. Research has emphasized lowering solar cell cost by improving performance and by reducing materials and manufacturing costs. One approach is to use optical concentrators such as mirrors or Fresnel lenses to focus the sunlight onto smaller-area solar cells. Other approaches replace the high-cost single-crystal silicon with thin films of amorphous or polycrystalline silicon, gallium arsenide, cadmium sulfide, or other compounds.

Solar radiation. The intensity and quality of sunlight are dramatically different outside the Earth's atmosphere from that on the surface of the Earth. The number of photons at each energy is reduced upon entering the Earth's atmosphere due to reflection, to scattering, or to absorption by water vapor and other gases. Thus, while the solar energy at normal incidence outside the Earth's atmosphere is 1.36 kW/m^2 (the solar constant), on the surface of the Earth at noontime on a clear day the intensity is about 1 kW/m^2.

On clear days the direct radiation is about 10 times greater than the diffuse radiation, but on overcast days the sunshine is entirely diffuse. The mean annual solar energy falling on the Earth's surface varies greatly from one location to another. The sunniest regions of the globe receive about 2500 kWh/m^2 per year of total sunshine on a horizontal surface. The Earth receives about 10^{18} kWh of solar energy each year. The worldwide annual energy consumption is about 80×10^{12} kWh, so that from a purely technical viewpoint, the world energy consumption corresponds to the sunlight received on about 0.008% of the surface of the Earth. *See* SOLAR RADIATION.

Principles of operation. The conversion of sunlight into electrical energy in a solar cell involves three major processes: absorption of the sunlight in the semiconductor material; generation and separation of free positive and negative charges to different regions of the solar cell, creating a voltage in the solar cell; and transfer of these separated charges through electrical terminals to the outside application in the form of electric current.

In the first step the absorption of sunlight by a solar cell depends on the intensity and quality of the sunlight, the amount of light reflected from the front surface of the solar cell, the semiconductor band-gap energy which is the minimum light (photon) energy the material absorbs, and the layer thickness. Some materials such as silicon require tens of micrometers' thickness to absorb most of the sunlight, while others such as gallium arsenide, cadmium telluride, and copper sulfide require only a few micrometers.

When light is absorbed in the semiconductor, a negatively charged electron and positively charged hole are created. The heart of the solar cell is the electrical junction which separates these electrons and holes from one another after they are created by the light. An electrical junction may be formed by the contact of: a metal to a semiconductor (this junction is called a Schottky barrier); a liquid to a semiconductor to form a photoelectrochemical cell; or two semiconductor regions (called a *pn* junction).

The fundamental principles of the electrical junction can be illustrated with the silicon *pn* junction. Pure silicon to which a trace amount of a fifth-column element such as phosphorus has been added is an *n*-type semiconductor, where electric current is carried by free electrons. Each phosphorus atom contributes one free electron, leaving behind the phosphorus atom bound to the crystal structure with a unit positive charge. Similarly, pure silicon to which a trace amount of a column-three element such as boron has been added is a *p*-type semiconductor, where the electric current is carried by free holes. Each boron atom contributes one hole, leaving behind the boron atom with a unit negative charge. The interface between the *p*- and *n*-type silicon is called the *pn* junction. The fixed charges at the interface due to the bound boron and phosphorus atoms create a permanent dipole charge layer with a high

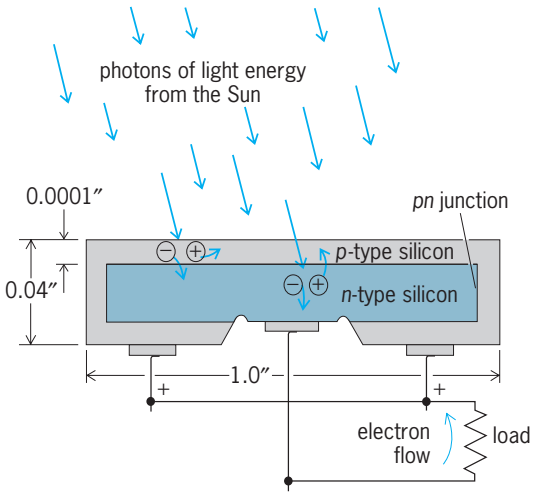


Fig. 1. Cross-sectional view of a silicon *pn* junction solar cell, illustrating the creation of electron pairs by photons of light energy from the Sun. 0.0001″ = 2.5 μm; 0.04″ = 1 mm; 1.0″ = 25 mm.

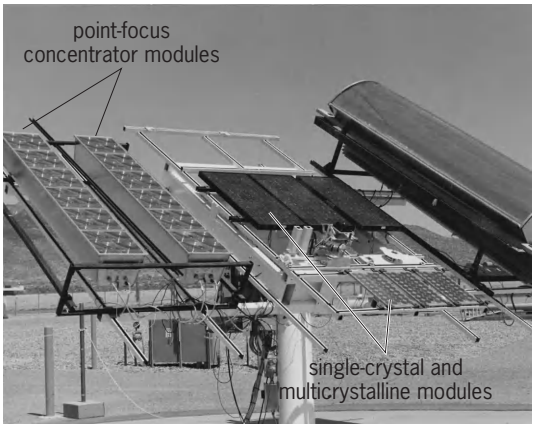


Fig. 2. Silicon solar cells assembled into modules and arrays to obtain higher voltage and current output. (*Sandia National Laboratories*)

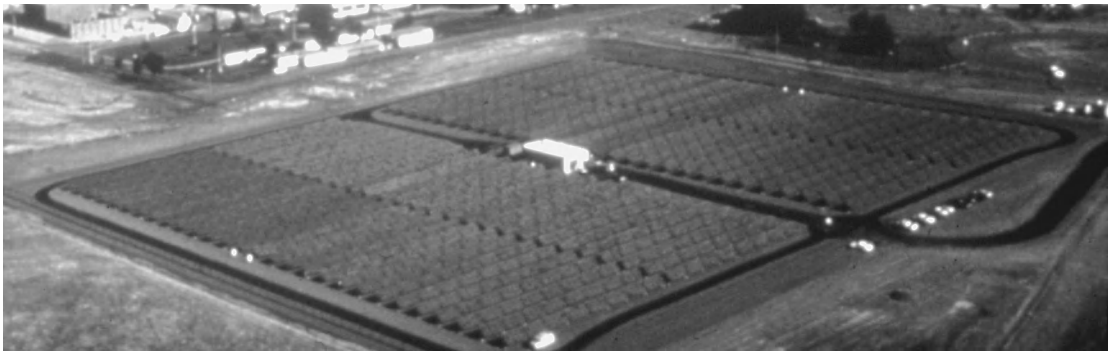


Fig. 3. A 2-MW photovoltaic array. (*Sacramento Municipal Utility District, California*)

electric field. When photons of light energy from the Sun produce electron-hole pairs near the junction, the built-in electric field forces the holes to the *p* side and the electrons to the *n* side (**Fig. 1**). This displacement of free charges results in a voltage difference between the two regions of the crystal, the *p* region being plus and the *n* region minus. When a load is connected at the terminals, electron current flows in the direction of the arrow, and electrical power is available at the load. *See* PHOTOVOLTAIC CELL; PHOTOVOLTAIC EFFECT; SEMICONDUCTOR; SEMICONDUCTOR DIODE.

Characteristics. The short-circuit current of a typical silicon *pn*-junction solar cell is directly proportional to light intensity and amounts to 28 mA/cm^2 at full sunlight (1000 W/m^2). The open-circuit voltage rises sharply under weak light and saturates at about 0.6 V for radiation between 200 and 1000 W/m^2. The maximum power output from the solar cell irradiated by full sunlight is about 11 mW/cm^2 at an output voltage of 0.45 V.

Under these operating conditions the overall conversion efficiency from solar to electrical energy is 11%. The output power as well as the output current is of course proportional to the irradiated surface area, whereas the output voltage can be increased by connecting cells in series just as in an ordinary chemical storage battery. Experimental samples of silicon solar cells have been produced which operate at efficiencies up to 18%, but commercial cell efficiency is around 10–12% under normal operating conditions.

Using optical concentration to intensify the light incident on the solar cell, efficiencies above 20% have been achieved with silicon cells and above 25% with gallium arsenide cells. The concept of splitting the solar spectrum and illuminating two optimized solar cells of different band gaps has been used to achieve efficiencies above 28%, with expected efficiencies of 35%. Thin-film solar cells have achieved between 4 and 9% efficiency and are expected in low-cost arrays to be above 10%. K. W. Mitchell

Arrays. Individual silicon solar cells or photovoltaic cells are about 100 cm^2 (15.5 in.^{2}) of surface area in size. At a 15% conversion efficiency, such a cell can deliver about 1.4 W at 0.5 V when in full sunlight. To obtain higher power and higher voltage, a number of cells must be assembled in panels or arrays (**Fig. 2**). Modules are of two basic types: flat plates (crystalline silicon or thin film) and concentrators. Cells may be connected in series to multiply their output voltage and in parallel to multiply their output current. Cells operated in series must be closely matched in short-circuit current since the overall performance of a solar cell array is limited by the cells having the lowest current.

Applications. Although the photovoltaic effect was discovered by A. C. Becquerel in 1839, practical solar cells made of silicon crystals were not developed until 1955. Beginning with *Vanguard 1*, launched in 1958, silicon solar cell arrays have become the almost exclusive power source for satellites. *See* SPACE POWER SYSTEMS.

Solar cell arrays have been used primarily to power small remote electrical loads that would otherwise be impractical or uneconomical to power by conventional means such as storage batteries or motor-generator sets. Solar cell arrays are sold worldwide to power such equipment as remote radio repeaters, navigational aids, consumer products, railroad signals, cathodic protection devices, and water pumps. Since most of the aforementioned uses require powering the load at times even when the Sun is not shining, electrical storage batteries are typically used in conjunction with solar cell arrays to provide reliable, continuous power availability.

Terrestrial uses take advantage of the ease of installation of photovoltaic arrays and the ease of adding or removing modules to meet changing electrical demands. These uses include agricultural, residential, commercial, and industrial applications (**Fig. 3**). When powering loads which require ac voltage, a static inverter is used to convert the dc voltage from the solar cell array into usable ac power. *See* SOLAR ENERGY. Margorie L. Tatro

Bibliography. M. A. Green, *Solar Cells*, 1982; T. Markvart (ed.), *Solar Electricity*, 2d ed., 2000; R. C. Neville, *Solar Energy Conversion: The Solar Cell*, 2d ed., 1995; L. D. Partain (ed.), *Solar Cells and Their Applications*, 1995; H. S. Rauschenbach, *Solar Cell Array Design Handbook*, 1980.

Solar constant

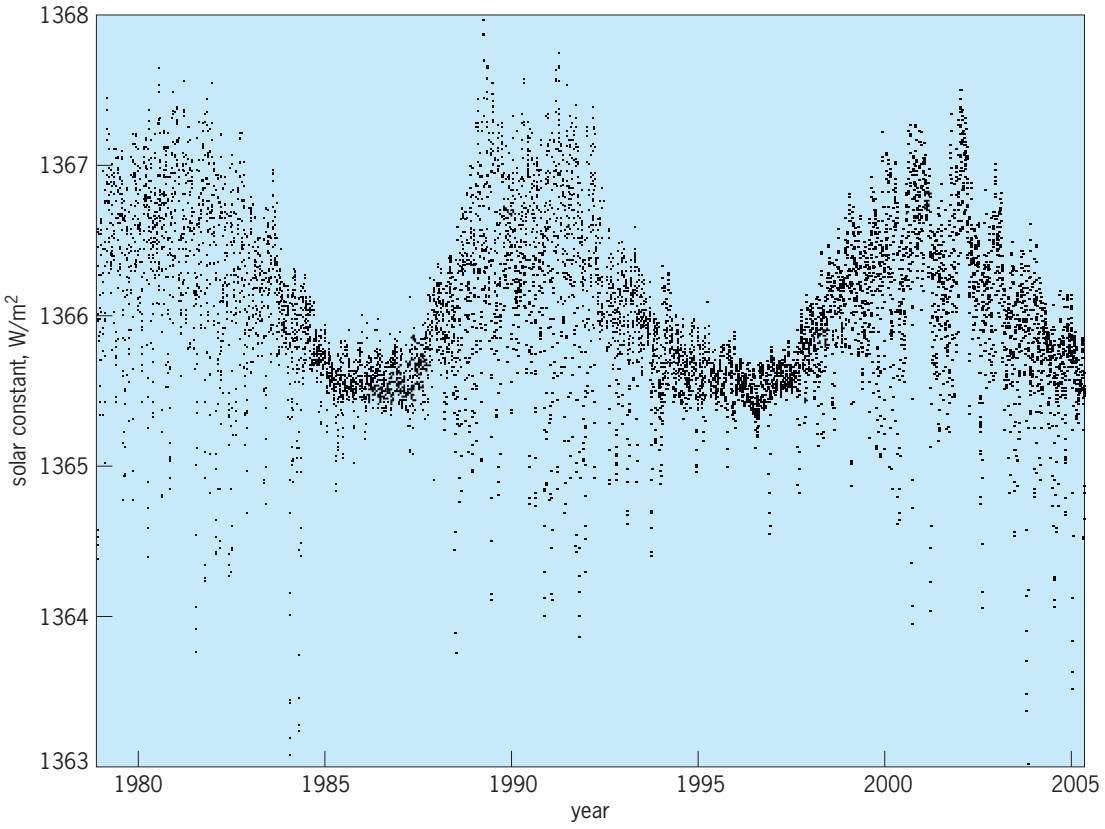
The total solar radiant energy flux (total irradiance) incident upon the top of the Earth's atmosphere at a standard distance (1 astronomical unit,

1.496×10^8 km or 9.3×10^7 mi) from the Sun. In 1980 it was discovered that the so-called solar constant actually varies with time, though only by small amounts, around a value of about 1366 $W \cdot m^{-2}$ (1.96 $cal \cdot cm^{-2} \cdot min^{-1}$). The solar radiant energy flux received at the Earth is more than 10^{10} times that of the next brightest star, and almost 10^6 times that of the full moon. This solar energy, averaged over its apparent variation (almost 7%) due to the Earth's orbital motions, maintains the climate on Earth. The solar constant represents a good measure of the solar luminosity (its total radiant energy), although anisotropic components (for example, a dark sunspot), which produce directional solar-constant variations, must be taken into account. *See* CLIMATOLOGY; EARTH ROTATION AND ORBITAL MOTION; PHOTOSPHERE; STAR.

An important test of the validity of any theory of stellar structure and evolution is the accuracy with which it predicts the steady luminosity of the Sun. However, the main astrophysical interest in variations of the solar constant concerns the new window it gives on transient processes in the solar interior. This new knowledge is welcome in view of the theoretical difficulties involved in calculating the solar neutrino flux successfully. *See* HELIOSEISMOLOGY; SOLAR NEUTRINOS; STELLAR EVOLUTION.

Causes of variability. Both expected and unexpected items contribute to the variability of the solar constant. Sunspots can produce deficits of up to a few tenths of 1% of the solar constant, on typical time scales of 1 week. Other surface manifestations of solar magnetic activity, faculae, contribute excesses rather than deficits. These features produce modulations at a roughly 27-day period due to solar rotation. Similar types of variability (due, for example, to so-called starspots) occur on other stars. These can have larger relative amplitudes. Convective motions and global oscillations of the solar interior, analogous to seismic waves on the Earth, produce variations of a few parts per million on time scales of a few hundred seconds; these waves are used to study solar interior structure and dynamics. Finally, and unexpectedly, there is an apparent 11-year sunspot cycle variation amounting to an approximately 0.1% increase of the solar constant during the sunspot maxima (see **illus.**). This long-term effect has the opposite dependence from that found for individual sunspots, which block the solar radiant energy and cause decreases rather than increases. If such small changes in solar energy flux affect the Earth's climate directly or indirectly, then this solar-cycle variation is consistent with an association noted some time ago: the so-called Maunder minimum period of greatly reduced sunspot activity in the seventeenth century corresponded with decades of cool climate in Europe, known as the Little Ice Age. *See* CLIMATE HISTORY; SUNSPOT.

Measurement. To measure the variability of the solar constant requires the deployment of a stable radiometer in space, above the Earth's atmosphere. The original high-quality observations came from



Composite time series of daily values of the solar constant as obtained from radiometers in space. (*Physikalisch-Meteorologisches Observatorium Davos/World Radiation Center; adapted from http://www.pmodwrc.ch/pmod.php?topic=tsi/composite/SolarConstant*)

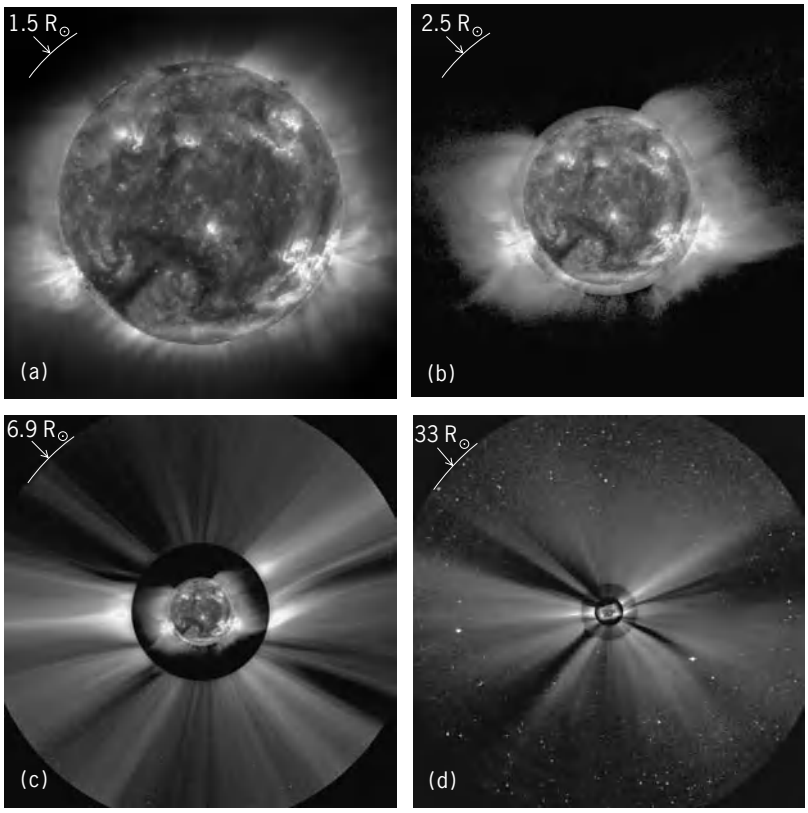
NASA's *Solar Maximum Mission* (1980–1989), and several other spacecraft have since contributed to the composite time series shown in the illustration. To measure the total solar irradiance (solar constant) typically requires an active cavity radiometer making an absolute, not relative, measurement. Because of the difficulty of this kind of measurement, it is not presently possible to infer the presence of any gradual variation extending to time scales exceeding a single sunspot cycle. On shorter time scales the measurements are very precise and can reveal the parts-per-million variations resulting from solar global oscillations. *See* RADIOMETRY; SUN.

Hugh S. Hudson

Bibliography. J. A. Eddy, R. L. Gilliland, and D. V. Hoyt, Changes in the solar constant and climatic effects, *Nature*, 300:689–693, 1982; C. Fröhlich, Long-term behaviour of space radiometers, *Metrologia*, 40:60–64, 2003; J. Pap and P. Fox (eds.), Solar variability and its effects on climate, *Geophys. Monogr. Series*, 141, 2004.

Solar corona

The outer atmosphere of the Sun. The corona is characterized by extremely high temperatures of several million kelvins, which cause it to extend far above the denser surface regions of the Sun. Coronal gas is constrained to follow the magnetic field of the Sun, forming it into the shapes seen during a solar eclipse, with a coronagraph, or in x-rays (see **illustration**). These shapes include long streamers that penetrate interplanetary space, loop-like tubes over the strongest fields, and vast regions of very low density called coronal holes. Near the north and south poles the coronal polar rays resemble the pattern formed by iron filings near a bar magnet, but at midlatitudes the more complicated fields of active regions dominate. The corona is hot enough to emit x-rays, and x-ray telescopes in space can form images of the corona. Such images display the magnetic loops connecting bright regions in the lower corona, while visible-light observations can detect the corona above the edge of the solar disk by carefully blocking out the strong photospheric light with an occulting disk (see illustration). *See* CORONAGRAPH; ECLIPSE; SOLAR MAGNETIC FIELD; X-RAY ASTRONOMY.



Global view of the solar corona, from observation by instruments on the *Solar and Heliospheric Observatory* (*SOHO*) on April 17, 1998. (*a*) On-disk images taken by the Extreme-Ultraviolet Imaging Telescope (EIT). (*b*,*c*,*d*) A set of nested visible-light images taken by the Large-Angle Spectroscopic Coronagraph (LASCO), showing the off-limb corona out to 30 solar radii ($R_\odot$). The active-region corona is shown in the on-disk images, while the streamer structure of the outer corona far above the solar surface is seen in the coronagraph images. (*Courtesy LASCO/EIT Consortium*)

Coronal holes are the source of solar matter streaming into interplanetary space. These streams often interact with the Earth's magnetic field to cause auroras and other geophysical phenomena. The interaction between the hot coronal plasma and the evolving magnetic fields in the corona sometimes leads to instabilities which eject large amounts of matter into space. These coronal mass ejections (CMEs) can have geophysical effects similar to those of the streams and ejecta. *See* AURORA; MAGNETOSPHERE; SOLAR WIND.

From a minimum in the lower atmosphere (approximately 4000 K or 7000°F), the temperature rises to values in the corona of $1–5 \times 10^6$ K ($2–9 \times 10^6$ °F) caused by dissipation of mechanical or magnetic energy produced by turbulent flows in the lower atmosphere. The mechanisms by which the corona is heated are uncertain, although results from the Transition Region and Coronal Explorer (TRACE) satellite indicate that the complex rearrangement of magnetic fields in the corona is largely responsible for coronal dynamics, via the rapid release of energy stored in the hot magnetized plasma. *See* SUN.

Leon Golub

Bibliography. A. H. Gabriel (ed.), *The Solar Corona*, Elsevier Science, 1994; L. Golub and J. M. Pasachoff, *Nearest Star*, Harvard University Press, 2001; L. Golub and J. M. Pasachoff, *The Solar Corona*, Cambridge University Press, 1997; International Astronomical Union, *IAU Colloquium on Coronal Structures*, 1993; A. I. Poland (ed.), *Coronal and Prominence Plasmas*, NASA Conf. Pub. 2442, 1986; J. B. Zirker, *Journey from the Center of the Sun*, Princeton University Press, 2001.

Solar energy

The energy transmitted from the Sun. The upper atmosphere of Earth receives about 1.5×10^{21} watt-hours (thermal) of solar radiation annually. This vast

amount of energy is more than 23,000 times that used by the human population of this planet, but it is only about one two-billionth of the Sun's massive outpouring—about 3.9×10^{20} MW. *See* SUN.

The power density of solar radiation measured just outside Earth's atmosphere and over the entire solar spectrum is called the solar constant. According to the World Meteorological Organization, the most reliable (1981) value for the solar constant is 1370 ± 6 W/m^2. Of this power, 8% is in the ultraviolet wavelengths, 47% in the visible spectrum, and 45% in the infrared region. The solar constant actually is not a true constant, but is subject to a small continuous variation due to the shape of Earth's orbit, amounting to −3.3% from average about July 5, when Earth is at its greatest distance from the Sun, and +3.4% about January 3, when Earth is closest to the Sun. *See* SOLAR CONSTANT.

Solar radiation is attenuated before reaching Earth's surface by an atmosphere that removes or alters part of the incident energy by reflection, scattering, and absorption. In particular, nearly all ultraviolet radiation and certain wavelengths in the infrared region are removed. However, the solar radiation striking Earth's surface each year is still more than 10,000 times the world's energy use. Radiation scattered by striking gas molecules, water vapor, or dust particles is known as diffuse radiation. Clouds are a particularly important scattering and reflecting agent, capable of reducing direct radiation by as much as 80 to 90%. Because cloud distributions and types are highly variable, these reductions are quite unpredictable. The radiation arriving at the ground directly from the Sun is called direct or beam radiation. Global radiation is all solar radiation incident on the surface, including direct and diffuse. **Figure 1** illustrates the events that affect solar radiation as it passes through Earth's atmosphere. *See* SOLAR RADIATION.

The amount of atmospheric absorption and scattering of solar radiation is a function of the effective distance (depending on atmospheric thickness and content) through which the radiation travels. Spectral irradiance is a function of air mass. The term air mass zero is used to represent zero mass of air through which the Sun's rays must pass (the solar intensity outside Earth's atmosphere). Radiation arriving at the surface of Earth can be measured against that reaching the fringes of the atmosphere, where there is no air. The light of the Sun directly overhead at a 90° solar altitude, or zenith, at sea level is said to pass through an air mass of 1, providing an average peak intensity of 1 kW/m^2. The solar declination, the angle between the Earth-Sun line and Earth's equatorial plane, varies as Earth circles the Sun, and this affects the solar radiation (insolation) reaching the ground. Intensity weakens at Sun angles approaching the horizon since the rays have more atmosphere to penetrate. The amount of solar energy incident on a horizontal surface at sea level ranges up to 2400 Btu/ft^2-day (7 kWh/m^2-day). The sites between latitudes 35°N and 35°S receive between 2000 and 3500 h of sunshine per year; higher latitudes receive less.

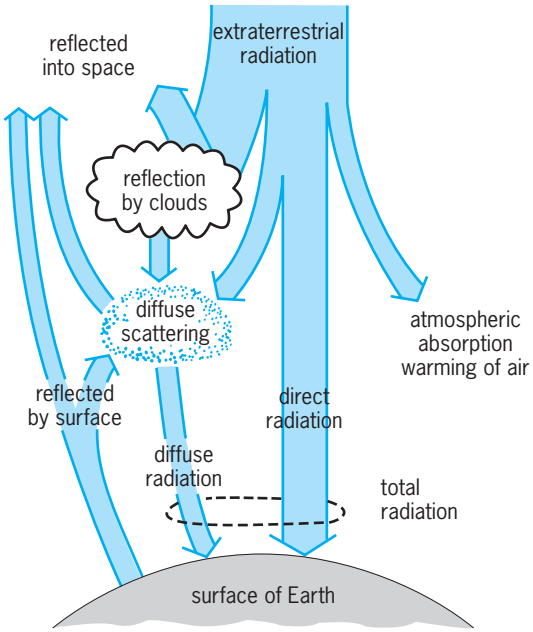


Fig. 1. Direct, diffuse, and total solar radiation. (*After W. C. Dickinson and P. N. Cheremisinoff, eds., Solar Energy Technology Handbook, pt. A: Engineering Fundamentals, Marcel Dekker, 1980*)

See ABSORPTION OF ELECTROMAGNETIC RADIATION; INSOLATION; SCATTERING OF ELECTROMAGNETIC RADIATION.

Solar research and technology development aim at finding the most efficient ways of capturing low-density solar energy and developing systems to convert captured energy to useful purposes. Also of significant potential as power sources are the indirect forms of solar energy: wind, biomass, hydropower, and the tropical ocean surfaces. With the exception of hydropower, these energy resources remain largely untapped. *See* ENERGY SOURCES.

Five major technologies using solar energy are being developed. (1) The heat content of solar radiation is used to provide moderate-temperature heat for space comfort conditioning of buildings, moderate- and high-temperature heat for industrial processes, and high-temperature heat for generating electricity. (2) Photovoltaics convert solar energy directly into electricity. (3) Biomass technologies exploit the chemical energy produced through photosynthesis (a reaction energized by solar radiation) to produce energy-rich fuels and chemicals and to provide direct heat for many uses. (4) Wind energy systems generate mechanical energy, primarily for conversion to electric power. (5) Finally, a number of ocean energy applications are being pursued; the most advanced is ocean thermal energy conversion, which uses temperature differences between warm ocean surface water and cooler deep water to produce electricity.

Solar Energy Systems in Buildings

Systems employing solar energy are classified as passive or active systems.

Passive systems. Passive solar energy systems use the building itself in combination with solar energy

as an energy-saving system. A passive building uses its environment, the features of its site, its structure, and its materials so that it requires much less fuel energy. Passive lighting, usually called daylighting, illuminates the interior of a building as a replacement for or supplement to electric lighting.

Heating systems. Most solar thermal systems consist of three basic parts: a collector, a storage medium, and a distribution system. In a passive solar heating system, the collector is usually the building itself. In direct systems, the Sun's rays enter the building, usually through windows, and heat the room or space in direct sunlight. Excess heat can be vented off or stored in the mass (walls and floor) of the building for later use. The entering solar energy is distributed through a combination of reradiation from the mass and natural convection of warm air between rooms. Indirect heating systems use one or more rooms as a collector, which is combined with a storage mass that separates that room from the rest of the building. The Sun's energy is reradiated and convected into the building after passing through the storage mass, which can store heat during the day for use at night. Popular passive heating systems include greenhouses and sunspaces.

Cooling systems. One type of passive solar cooling occurs when the cycle established in a passive heating system is reversed, allowing the building to radiate heat to the sky at night. As the mass is cooled, it lowers the temperature inside the house. Then, during the daytime, the cool mass helps maintain comfortable conditions in summer without using air conditioning. Sometimes, the radiant cooling process is assisted by placing a storage mass, such as bags of water, on top of the building. The storage mass is covered to prevent heating during the day but is exposed to the night sky.

Another form of passive cooling is natural ventilation, such as is caused by air currents entering through open windows or doors. Ventilation also occurs as heated air rises to the ceiling. If openings are placed near the ceiling of a room and at floor level, warm air will rise and exit through the top openings, drawing cooler air in through the lower openings.

Natural cooling also occurs when a building is shaded during the summer. If shade trees are deciduous (shed their leaves during the winter), they will not interfere with a passive heating system. The house can also be self-shading through the use of overhangs and extended walls.

Daylighting systems. In commercial, institutional, and industrial buildings, the largest energy use is often electric lighting. Daylighting can be used to replace or supplement electric lighting in buildings occupied primarily during the daytime. Windows and other types of apertures allow daylight to enter the building, and special control systems regulate the electric lighting to turn it off when there is sufficient daylight in the building. Unlike other passive systems, daylighting can occur on overcast days because the daylight in the building does not depend on the direct use of the Sun.

It is possible to use a variety of materials to effect passive solar performance, such as antireflective window coatings, insulating window shades and curtains, or modular water walls. Some window coatings carry an electric current that can change window transmittance in response to weather conditions. In addition, there are several chemical compounds that can be incorporated either into window glass or between panes to increase a window's storage or insulating capacity.

Active space conditioning and hot water systems. Solar active space conditioning and hot water systems use mechanical means to collect, store, and distribute solar energy to heat and cool buildings and to heat water for domestic or commercial use. In general, water heating, the simplest system, is characterized by the smallest collector array, the simplest design, and the lowest cost; retrofit can be accomplished easily. Space heating requires a large collector array, has the largest heat storage requirement, and is relatively expensive. Space heating and cooling is the most complex of the solar active systems, requiring the largest collector array, the highest temperatures, and a complicated mechanical process.

Active systems generally encompass four parts: collection (including storage), conversion, distribution, and control (**Fig. 2**).

Collection. Collectors can be flat-plate, evacuated-tube, or concentrating. In most active systems, flat-plate collectors are used in one or more modules. The Sun's radiation is absorbed by the collector and is converted to heat in a liquid (water or glycol) or gas (air) heat-transfer medium. The heat is transported by pumps or fans to the conversion and distribution systems for immediate use or is stored for later use. The heat-transfer medium in liquid systems usually is water, which must be protected from freezing in cold climates. An antifreeze can be added to the water, or the collectors can be drained when the outdoor temperature approaches freezing. If an antifreeze is added, a heat exchanger is used to isolate the water in the collectors from the internal building water.

Conversion. For both space heating and cooling and domestic hot water, a conversion system is required to convert the thermal energy to useful heat. The thermal energy may also be converted to mechanical work that can subsequently drive a conventional vapor compression cooling unit. Thermal energy can also be used directly, through absorption refrigeration, for space cooling.

Desiccant cooling for air conditioning uses recirculated, dehumidified air as well as evaporative cooling to produce cool, comfortable air for the building (**Fig. 3**). The desiccant material is regenerated and reused when water is driven off by solar-heated air.

Distribution. The distribution system transports the hot or cold fluid from the conversion system to the point of use in the building. Generally, one distribution system for both the solar system and auxiliary (backup) system is used to reduce duplication.

Control. The control system collects temperature information from inside and outside the building, processes the information, and sends commands as

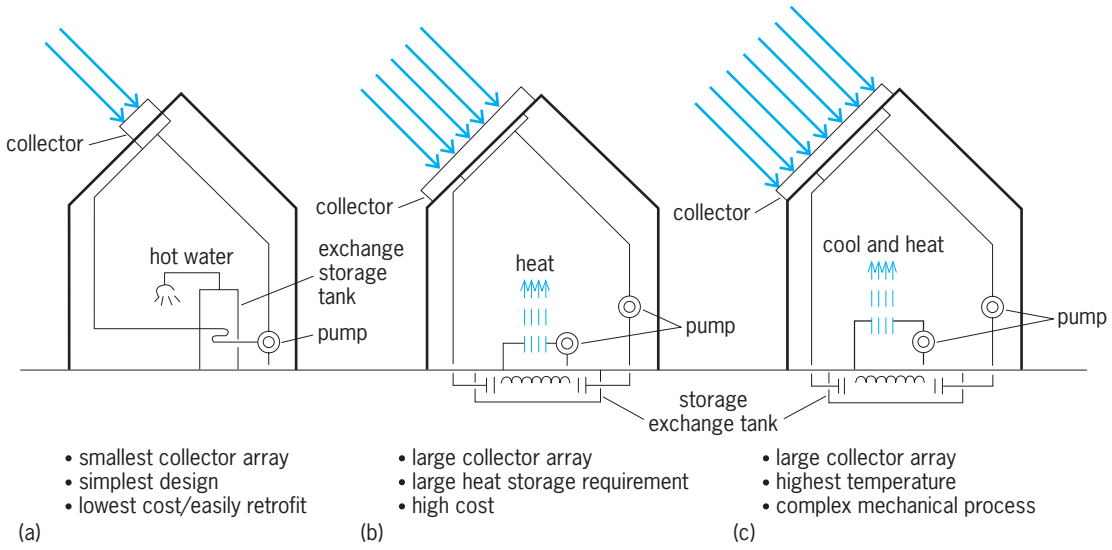


Fig. 2. Building space conditioning and hot water systems. (*a*) Water heating. (*b*) Space heating. (*c*) Space heating and cooling. (*After DOE 5-Year Research Plan for National Active Solar Heating and Cooling Program, U.S. Department of Energy, 1983*)

conditions dictate. Temperature sensors provide data to a network of relays or a microprocessor, which processes the data and sends commands to operate the collection, conversion, or distribution system.

Research has included efforts to improve the cost of collectors by testing lightweight flat-plate collectors using thin-film plastics rather than conventional metal-glass materials, identifying desiccants and other cooling system materials, and collecting data from a network of building sites on performance, reliability, and durability of active solar systems. *See* SOLAR HEATING AND COOLING.

Solar Thermal Technology

Solar energy can be converted to useful work or heat by using a collector to absorb solar radiation, allowing much of the Sun's radiant energy to be converted to heat. This heat can be used directly in residential, industrial, and agricultural operations; converted to mechanical or electrical power; or applied in chemical reactions for production of fuels and chemicals.

Collector systems. The collector system contains a concentrator and a receiver. The concentrator redirects and focuses sunlight on the receiver by using mirrors or lenses, and the receiver absorbs solar radiation and converts it to heat. Solar collectors are of two basic classes: nonfocusing and focusing. They are further distinguished on the basis of their concentrator optical properties (**Fig. 4**), and the operating temperature that can be obtained at the receiver (**Table 1**).

Nonfocusing collectors. Nonfocusing collectors are generally kept fixed (nontracking), yielding low-temperature outputs. Two types of this class of collectors are the solar pond and the evacuated-tube collector.

One type of solar collector that could have wide applicability is the salt-gradient solar pond (**Fig. 5**). A thin layer of fresh or low-salinity water covers a deeper layer of water in which a salinity gradient is created (that is, salt concentration increases with depth). The bottom layer is usually at or near salt saturation. Sunlight passing through the water is absorbed and heats the bottom layer. The salt concentration in the deeper water serves to increase the liquid density and thereby prevents the natural convection that would normally mix both the warm lower layers and the cool layers above. The heat

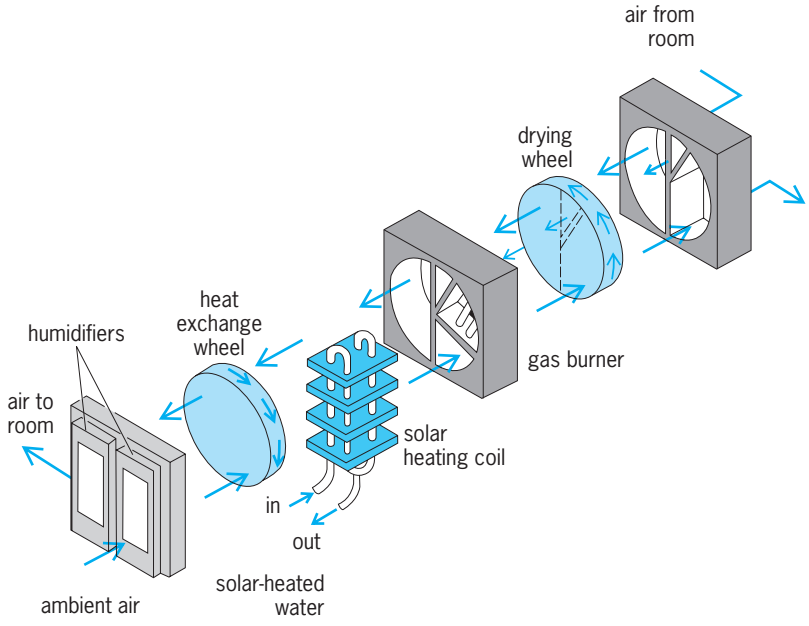


Fig. 3. Desiccant solar cooling system with gas backup in recirculation mode. (*After R. S. Barlow, Analysis of the Adsorption Process and of Desiccant Cooling Systems: A Pseudo-Steady Model for Coupled Heat and Mass Transfer, SERI/TR-631-1330, Solar Energy Research Institute, 1982*)

TABLE 1. Typical operating temperatures for solar thermal collectors

Type of collector	Operating temperature range, °F (°C)
Central receiver	400–2500 (200–1400)
Point focus (parabolic dish and Fresnel lens)	550–2500 (300–1400)
Line focus (parabolic trough and Fresnel lens, also multiple reflector)	160–600 (70–320)
Evacuated tube	120–350 (49–177)
Flat-plate	50–160 (10–71)
Solar pond	50–180 (10–82)

is thus trapped and stored at the bottom and can be extracted for use. Care must be taken to prevent the extraction process from upsetting the stability of the layers. Wind effects on the surface tend to cause mixing that is detrimental to pond performance, and floating or suspended debris will reduce the pond's transparency to sunlight.

Several solar ponds have been constructed and are operating around the world. The bottom hot brine, with temperatures near 180°F (82°C), can be used directly or via a heat exchanger to provide low-temperature heat for many industrial applications, for district space heating, or for electricity generation.

The evacuated-tube collector, sometimes incorporated with a reflective backing to provide greater solar concentration, can provide temperatures of up to 300°F (150°C), the temperature required to produce low-grade, industrial process steam.

Focusing collectors. This major class of collectors includes the distributed receiver system and the point-focus central receiver system.

Distributed receiver systems, such as line-focus (for example, parabolic trough) and point-focus (for example, parabolic dish), are commonly considered for remote community-power systems, military applications, individual factory or commercial building systems, or agricultural applications. These collectors must always point toward the Sun and cannot make use of diffuse and reflected light, which can account for a 10 to 20°F (5 to 10°C) temperature rise in flat-plate collectors.

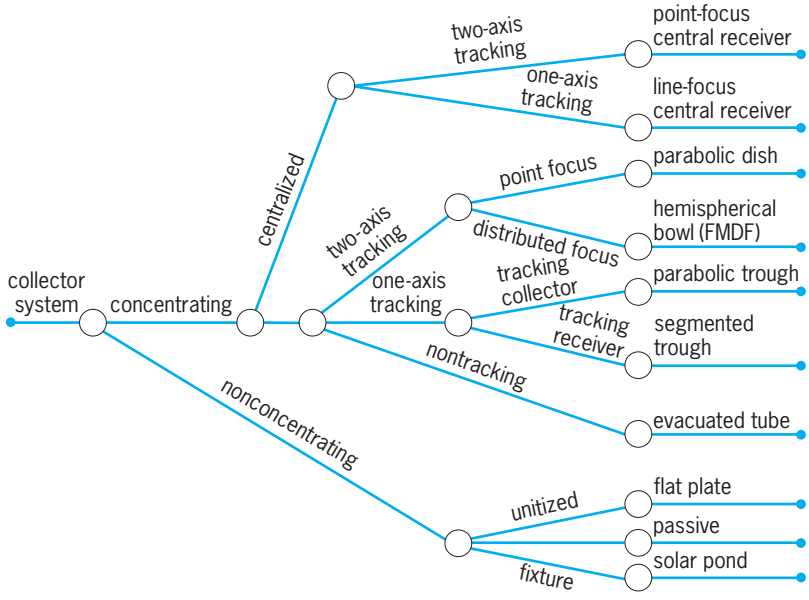


Fig. 4. Solar collector classification based on optical properties. (*After J. P. Thornton, A Comparative Ranking of 0.1-10 MWE Solar Thermal Electric Power Systems, SERI/TR-351-461, Solar Energy Research Institute, 1980*)

1. Line-focus. The line-focus collector allows the concentrating, mirrored, reflector surface to follow the Sun by rotating about one axis (either east-west or north-south). Radiation is reflected onto an absorber tube (the receiver). The parabolic trough collector (**Fig. 6**) is the most common type of line-focus collector. It offers a wide range of operating temperatures from approximately 160 to 600°F (70 to 320°C). Therefore, it has been used for water heating below 212°F (100°C), steam production at 212 to 600°F (100 to 320°C), and power production at up to 600°F (320°C) using heat-transfer coils.

2. Point-focus. This major class of collectors offers the highest optical and thermal performance of any collector type because of its ability to track the Sun over its full range of diurnal motion and because of the relatively small area of the absorber used in the receiver unit. Parabolic dish–shaped collectors concentrate direct solar radiation to a point at the focus of the parabola. Each dish is a complete power-producing unit (**Fig. 7**), which can function either as an independent system or as part of a group of modules linked to form a larger system. A single parabolic-dish module can achieve fluid temperatures from 550 to 2500°F (300 to 1400°C) and can efficiently produce up to 25 kW of electricity.

The largest subsystem of a parabolic-dish module is the concentrator, a shallow dish with a reflective surface that tracks the Sun to focus sunlight on a receiver. Tracking along two axes ensures maximum solar energy collection during the day. Because the receiver absorbs concentrated solar radiation at very high temperatures (up to 2500°F or 1400°C), the system can be used for high-temperature chemical reactions or the production of steam, electric power, or fuels and chemicals. In many system designs, the engine for thermal-to-mechanical energy conversion is mounted with the receiver at the focus.

3. Central receiver system. The central receiver system concept consists of flat or slightly curved mirrors (heliostats) that rotate about two axes to reflect direct solar radiation onto a tower-mounted receiver. The absorbing surface can be a cylinder (external receiver) or a planar surface contained within a cavity (cavity receiver). Temperatures in excess of 2500°F (1400°C) and more and pressures of 1015 lb/in.^{2} absolute (7 megapascals) can be generated at the receiver. The central receiver–collector concept appears to hold promise for producing solar thermal electric power in capacities from approximately 1 to 100 MWe. **Figure 8** shows the mode of operation of a typical point-focus central receiver system for power generation using conventional steam technology.

Conversion systems. Solar thermal heat can be converted directly to electricity, converted to mechanical and then electrical power, or used in the production of chemicals and fuels via an appropriate thermodynamic engine cycle. Three thermodynamic cycles have been considered for thermal-to-mechanical energy conversion: the Stirling cycle engine with closed-cycle operation; the Brayton cycle engine with either open- or closed-cycle operation; and the conventional Rankine cycle. Key concerns are the achievement of high engine performance and reliability and identification of operating requirements at the interface of the engine and receiver. Thermal energy conversion using parabolic-dish technology can be made at the point of heat generation, or at some distant location by piping heated fluid elsewhere for heat extraction and conversion. *See* BRAYTON CYCLE; RANKINE CYCLE; STIRLING ENGINE.

Storage. A solar energy system is normally designed to be able to deliver useful heat for 6 to 10 h a day, depending on the season and weather. Storage capacity in the solar thermal system is one way to increase a plant's operating capacity.

There are four primary ways to store solar thermal energy: (1) sensible-heat-storage systems, which store thermal energy in materials with good heat-retention qualities; (2) latent-heat-storage systems, which store solar thermal energy in the latent heat of fusion or vaporization of certain materials undergoing a change of phase; (3) chemical energy storage, which uses reversible reactions (for example, the dissociation-association reaction of sulfuric acid and water); and (4) electrical or mechanical storage, particularly through the use of storage batteries (electrical) or compressed air (mechanical). *See* ENERGY STORAGE.

Applications. Solar thermal systems that are used in electric power applications include the small, distributed power systems (typically 1000 kWe or less) and the large, central power systems of 10,000 kWe or more. Small-power uses include those for remote communities; military applications; and individual factory, commercial building, or agricultural applications. For these categories, the distributed receiver systems are most common. For generating electric power to distribute through a utility grid or to produce fuels and chemicals, the central receiver power and parabolic-dish systems are the most appropriate.

Industrial process heat is the thermal energy used directly in the preparation and treatment of materials and goods manufactured by industry. Hot water at temperatures between 120 and 200°F (50 and 100°C) can be supplied by directly heating water in the absorber tubes of evacuated-tube-type collectors or linear concentrating collectors. Hot air for industrial drying operations below about 350°F (180°C) can be supplied by collector systems designed to handle air as a circulating fluid or to circulate fluid pumped through an air-liquid heat exchanger. Concentrating solar thermal collector systems, both distributed and centralized, are commonly considered for steam generation at temperatures between 300 and 750°F (150 and 400°C) for use as direct heat in various manufacturing industries. Potential uses also include the production of ammonia in large agricultural fertilizer plants and enhanced oil recovery.

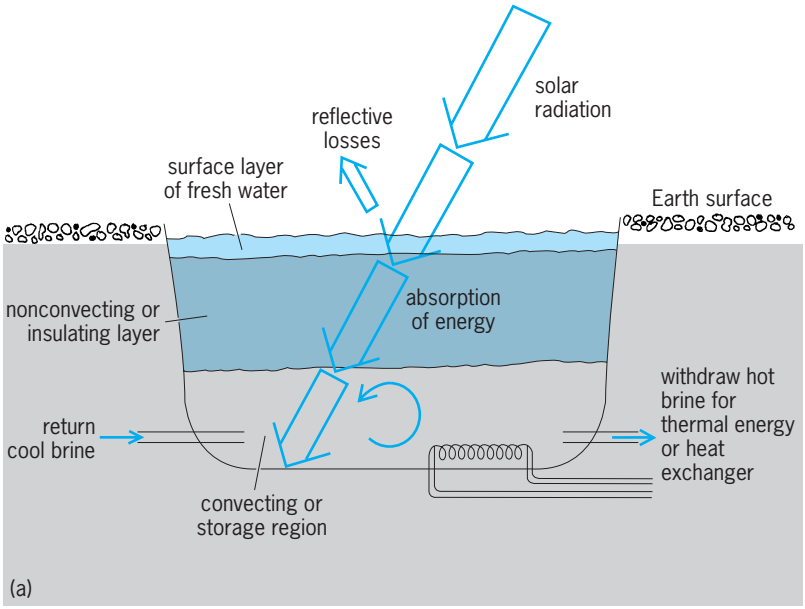


Fig. 5. Salt-gradient solar pond. (*a*) Schematic of the collection process (*after G. Franta et al., Solar Design Workbook, SERI/TR-62-308, Solar Energy Research Institute, 1981*). (*b*) Experimental pond (*Solar Energy Research Institute*).

Photovoltaics

Photovoltaic systems convert light energy directly to electrical energy. Using one of the most versatile solar technologies, photovoltaic systems can, because of their modularity, be designed for power needs ranging from milliwatts to megawatts. They can be used to provide power for applications as small as a wristwatch to as large as an entire community. They can be used in centralized systems, such as a generator in a power plant, or in dispersed applications, such as in remote areas not readily accessible to utility grid lines.

Historically, photovoltaic systems have been used

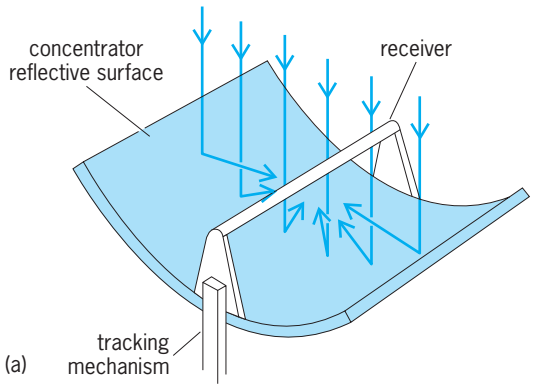


Fig. 6. Parabolic trough system. (*a*) Schematic of the components (*after Jet Propulsion Laboratory, Fact Sheet on Parabolic Trough Systems: A Solar Thermal Concentrating Collector Concept, 5106-22/3, 1982*). (*b*) Typical system (*Solar Energy Research Institute*).

where traditional electric power was unavailable or too expensive, providing power for satellites, remote cabins and houses, communication stations, observatories, water pumps, mobile military applications, and even for entire villages. Electricity has been brought to some areas of the world and has replaced fuel- and maintenance-reliant diesel generators. Photovoltaic systems are also used as power sources for hand-held calculators, watches, and portable television sets.

System components. The solar cell, composed principally of semiconductor material, is the basic component of the photovoltaic system. Though many materials and structures are being investigated, most commercial cells use a pure single-crystal silicon wafer that is doped *p*-type (that is, a small amount of an appropriate impurity is added, inducing extra unfilled bonds, or holes, in the wafer). Another impurity is diffused into a shallow region on the top of the wafer, promoting extra unbonded electrons, causing that portion to become *n*-type. A potential barrier, or electric field, is set up at the junction of the electrically different semiconductor layers. When photons strike the cell, they create ion pairs (negatively charged electrons and positively charged holes). These charges move around the silicon lattice structure until they recombine or cross the junction to be used in the photovoltaic process. The potential barrier segregates the ion pairs to opposite ends of the cell. The separated charges set up a potential (voltage), which is used to drive a current (composed of the light-generated electrons) through electrical contacts attached to both faces of the cell and through an external circuit (**Fig. 9**). *See* PHOTOVOLTAIC CELL; SEMICONDUCTOR.

The typical crystalline silicon cell produces a characteristic open-circuit voltage of about 0.5 V direct current (dc). The current generated depends on cell efficiency, cell area, and incident sunlight. A typical cell 4 in. (10 cm) in diameter can produce about 1 W of power under conditions where the Sun is directly overhead in a clear sky (one-sun conditions).

For larger voltages or currents, photovoltaic modules are formed by mounting groups of cells together on a rigid plate and interconnecting them in series or parallel. Stringing them in series increases the voltage; connecting them in parallel increases the current (**Fig. 10**). To further increase voltage or current output, modules may be grouped in parallel or strung in series to form photovoltaic arrays. The arrays, in turn, may be interconnected to form array fields. As

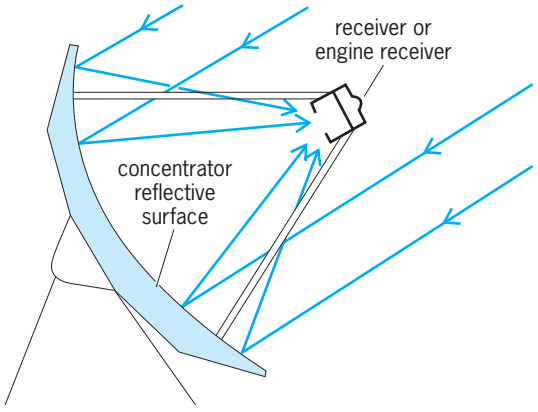


Fig. 7. Point-focus collector. (*a*) Schematic of collection process (*after Jet Propulsion Laboratory, Fact Sheet on Parabolic Dish Systems: A Solar Thermal Concentrating Collector Concept, 5106-22/2, 1982*). (*b*) Typical dish-receiver and engine module (*Jet Propulsion Laboratory*).

with cells, series combinations are voltage-additive, and parallel arrangements are current-additive.

Most commercial modules and arrays are based upon flat-plate collectors (which look like windows with cells behind them) that are able to use either direct or diffuse sunlight effectively. They can be mounted on a stationary surface, such as a roof top, and faced in a southerly direction (for systems in the Northern Hemisphere) at an appropriate angle. Concentrating systems, which use lenses or mirrors to focus sunlight on each cell, are also being used. Concentration increases the power produced by each cell. Since such systems require direct sunlight, they must use mechanisms to track the Sun. Also there must be provision for removing heat, since high temperatures can reduce cell output power. *See* PHOTOELECTRIC DEVICES.

Many electric power applications require alternating–current (ac) electricity. To obtain ac, a power-conditioning system is needed to convert the photovoltaic-generated dc to ac. A typical power-conditioning system consists of an inverter for converting the dc output to ac; filtering circuits to remove harmful or unwanted signals; a logic subsystem to control the on-off cycling of the photovoltaic system; a dc contactor to disconnect the dc source from ac lines during outages; and an isolation transformer to isolate the dc source voltage from the ac line voltage (**Fig. 11**).

When the Sun is not shining, at night or because of inclement weather, no power is generated by photovoltaic collectors. For this reason, and because the photovoltaic system must be able to start up and shut down smoothly and be able to satisfy user needs, a backup or storage system is often required. Systems installed on private residences usually can tie into the local utility grid, using it as an energy receptacle or resource, as conditions warrant. Other options include battery storage, pumped hydroelectric storage, and flywheel energy storage. A diesel engine-generator also can be used. *See* FLYWHEEL; GENERATOR.

Efficiencies. Under ideal conditions the theoretical maximum efficiency of a single-crystal silicon cell is about 25%, while the practical maximum efficiency hovers around 22%. Of the factors that combine to limit the efficiency, the most important are due to the inherent physics of the situation, the interaction between the incident photons and the silicon solar cell. A great percentage of the incident photon radiation is not energetic enough to be absorbed by the cell to create electron-hole pairs. An even greater percentage of the radiation is too energetic—the excess photon energy is not used in the conversion process.

By 1984, the efficiency of a typical commercial solar cell made from single-crystal silicon had increased by about a factor of two. Commercial cells became available that could convert from 13 to 15% of the incident radiant energy to electrical energy (at 77°F or 25°C and 1 kW/m^2 incident radiation). Single-crystal silicon cells have been produced in the laboratory with efficiencies approaching 19% under one-sun conditions, and further improvements are anticipated. Strides have also been made to increase the efficiencies of modules and arrays by decreasing intercell connection resistances and by employing protective circuitry that uses little power. This has resulted in array efficiencies of 12%, nearing the efficiency figures that some feel are necessary in order for photovoltaic systems to become competitive with conventional energy sources in the central utility market.

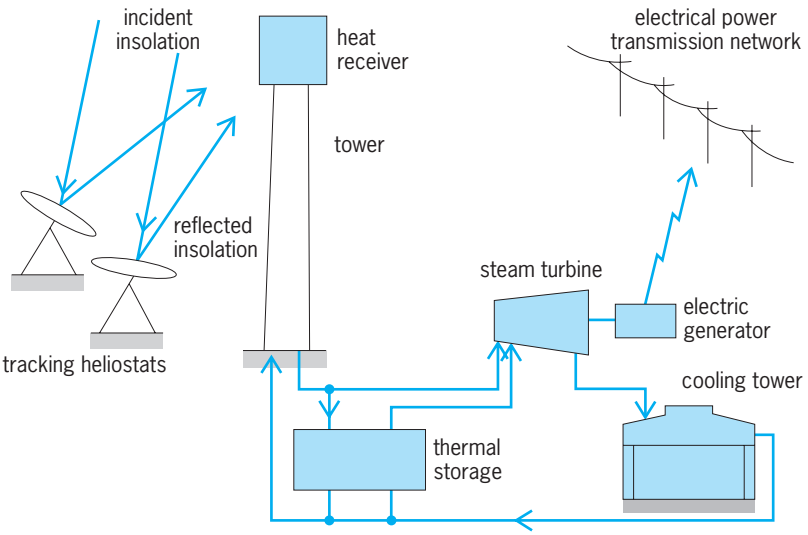


Fig. 8. Schematic of a typical central receiver collector for power generation using conventional steam technology. (*After F. Krawiec, J. Thorton, and M. Edesess, An Investigation of Learning and Experience Curves, SERI/TR-353-459, Solar Energy Research Institute, 1980*)

Advances have been made in power conditioning subsystems, another source of power loss for a photovoltaic system. Though commercially available converters in the early 1980s were only about 80% efficient over their broad range of operation, experimental systems have now reached 95% efficiency.

Other photovoltaic materials have energies that more suitably match the Sun's spectrum and thus make more efficient use of it. One of them, single-crystal gallium arsenide, has high efficiency, stability, and the ability to withstand higher operating

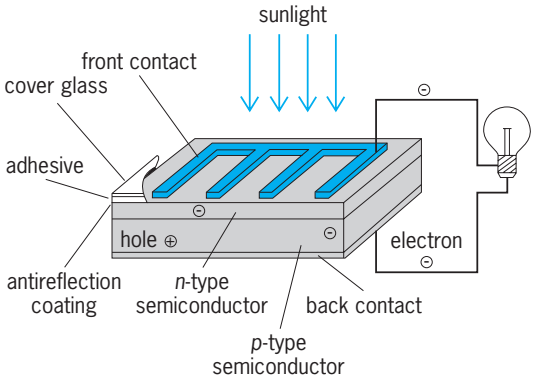


Fig. 9. Typical single-crystal silicon solar cell. (*After G. Cook, SERI Photovoltaic Advanced Research and Development: An Overview, SERI/SP-281-2235, Solar Energy Research Institute, 1984*)

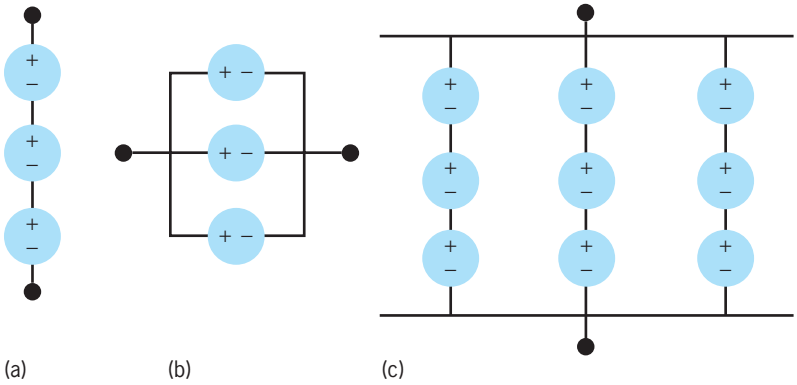


Fig. 10. Photovoltaic cell arrangements for building voltage and current. (*a*) String of photovoltaic cells (series). (*b*) Groups of cells (parallel). (*c*) Group of strings (combination). (*After Renewable Energy Technology Handbook for Military Engineers, SERI/SP-200-1413, Solar Energy Research Institute, 1982*)

temperatures than single-crystal silicon. For one-sun conditions, the highest theoretical efficiency of a single-crystal gallium arsenide cell is greater than 27%, and laboratory cells have reached efficiencies as high as 22%. Under sunlight concentrated 150 times, laboratory gallium arsenide cells have reached efficiencies nearing 24%.

For even higher efficiency, there is the multijunction (or stacked) device, a complex structure that uses several different semiconductor materials. Each layer has a junction sensitive to a different portion of the solar radiation spectrum, to make optimal use of all incident radiation. For example, an arrangement using three cells and three junctions—gallium aluminum arsenide, gallium arsenide, and silicon—has a theoretical maximum efficiency greater than 40% (**Fig. 12**).

Biomass

Biomass is plant and animal material. Biomass energy is solar energy stored in plant and animal matter. Through photosynthesis in plants, energy from the Sun transforms simple elements from air, water, and soil into complex carbohydrates. These carbohydrates can be used directly as fuel (for example, burning wood) or processed into liquids and gases (for example, ethanol or methane). Sources of biomass that can be converted to useful energy include agricultural crops, wastes, and residues; forest wood, waste, and residues; animal wastes; some municipal wastes; aquatic and desert plants; algae; and bacteria. Biomass is a renewable energy resource because it can be harvested periodically and converted to fuel. *See* BIOMASS; CARBOHYDRATE; PHOTOSYNTHESIS.

Feedstock production. Feedstock for biomass conversion can consist of wood, herbaceous plants, or aquatic plants.

Wood. Wood has been used as a fuel for tens of thousands of years. Woody biomass comes from a variety of sources, ranging from standing forests to wood processing operations. It includes mill and logging residues, timber thinnings, and harvests from well-managed natural stands or from short-rotation, fast-growing plantations. Mill residues are the by-products of processing operations that include the production of lumber, plywood, pulp and paper, and furniture. Dry mill residues from these operations are the most readily available of all tree biomass. Logging residues (residues left after tree harvest) include tree tops and branches, cull logs, standing dead trees, and stumps.

Wood has a heating value of about 8500 Btu/lb (20 megajoules/kg) when oven-dried compared to gasoline at about 21,000 Btu/lb (48 MJ/kg) and coal at about 13,500 Btu/lb (31 MJ/kg). Moisture dramatically affects the value of wood as a fuel. With a 30% moisture content, wood has a heating value of 5950 Btu/lb (14 MJ/kg); at 50% moisture the value is 4250 Btu/lb (10 MJ/kg). There are several methods of reducing moisture in wood to improve its performance as a fuel. Perhaps the simplest is drying, but some faster methods have been developed that apply residual combustion heat to dry the feedstock.

Improved silvicultural practices on harvested forests include overstocking (for example, 1000 stems per acre instead of 600) and improving regrowth from stumps following harvest. Another

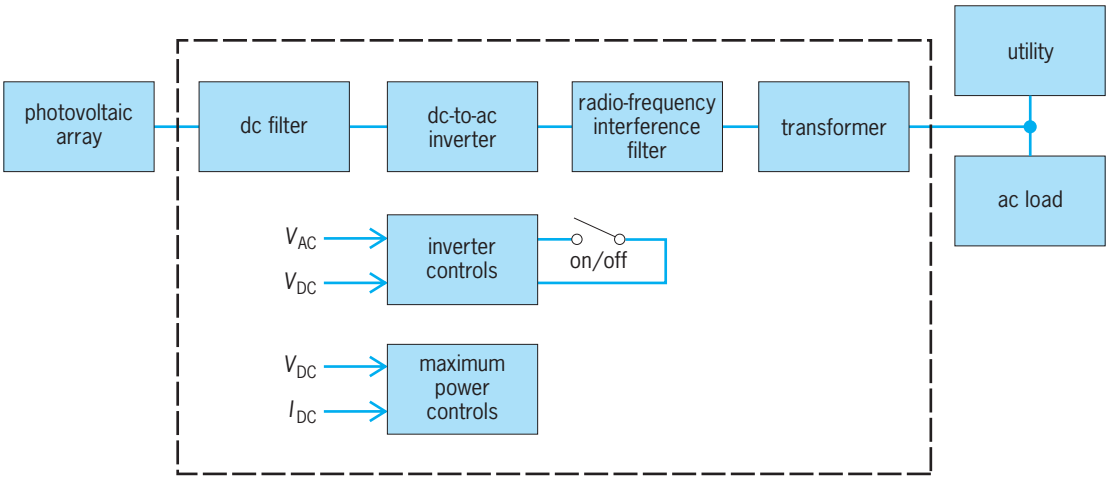


Fig. 11. Transformer isolation of dc from ac. (*After Renewable Energy Technology Handbook for Military Engineers, SERI/SP-200-1413, Solar Energy Research Institute, 1982*)

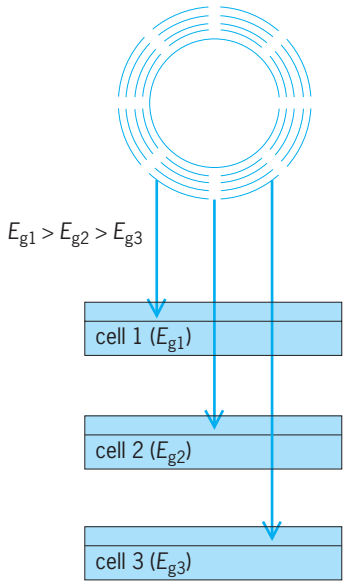


Fig. 12. Multiple-junction cell for highest efficiency. Gallium aluminum arsenide cell band gap (E_{g1}) = 1.82 eV. Gallium arsenide cell band gap (E_{g2}) = 1.45 eV. Silicon cell band gap (E_{g3}) = 1.1 eV. (*After G. Cook, SERI Photovoltaic Advanced Research and Development: An Overview, SERI/SP-281-2235, Solar Energy Research Institute, 1984*)

forestry practice improves production by cultivating fast-growing trees for energy conversion. These trees are selected for their rapid growth, ease of establishment, stump regeneration, and pest and disease resistance. Short-rotation plantations may require intensive cultivation practices similar to those used in agriculture such as irrigation, fertilization, and harvesting operations. *See* SILVICULTURE.

Herbaceous plants. Nonwoody (herbaceous) biomass feedstocks such as grasses and grains show promise as sources of fuel. Some herbaceous plants have both the potential for high yield and the ability to grow on arid and marginal land (**Fig. 13**). Herbaceous feedstock types can include plants intended as fuel for direct combustion to energy and plants grown for their carbohydrate or hydrocarbon content for conversion to liquid and gaseous fuel.

Grasses and legumes can be pressed into pellets for direct burning or gasification, or they can be converted via fermentation to alcohol. Starch and sugar crops also can be converted to energy via fermentation to alcohol. Agricultural residues not used for other purposes such as food or forage can be burned for heat or converted to alcohol.

Numerous plant and seed species are possible candidate crops for oil production. For example, about 2000 species of *Euphorbia* (for example, poinsettia, cassava, or the rubber tree) produce a latex rich in hydrocarbons that can be a substitute for petroleum in almost all its uses. Liquid hydrocarbons that are extracted from *Euphorbia* spp. may be processed into gasoline and similar fuels. *See* EUPHORBIALES; RUBBER TREE.

Aquatic plants. Certain aquatic plants can be cultivated on nonarable or nonproductive land. They have exceptional rates of biomass production and provide yields that surpass those of land plants by a factor of four or more. Microalgae have been studied as producers of gaseous (hydrogen) and liquid (oils) fuels. Macroalgae, such as the giant kelp, grow naturally in the open ocean and can be cultivated offshore or onshore in ponds. These species have high growth rates and produce significant amounts of

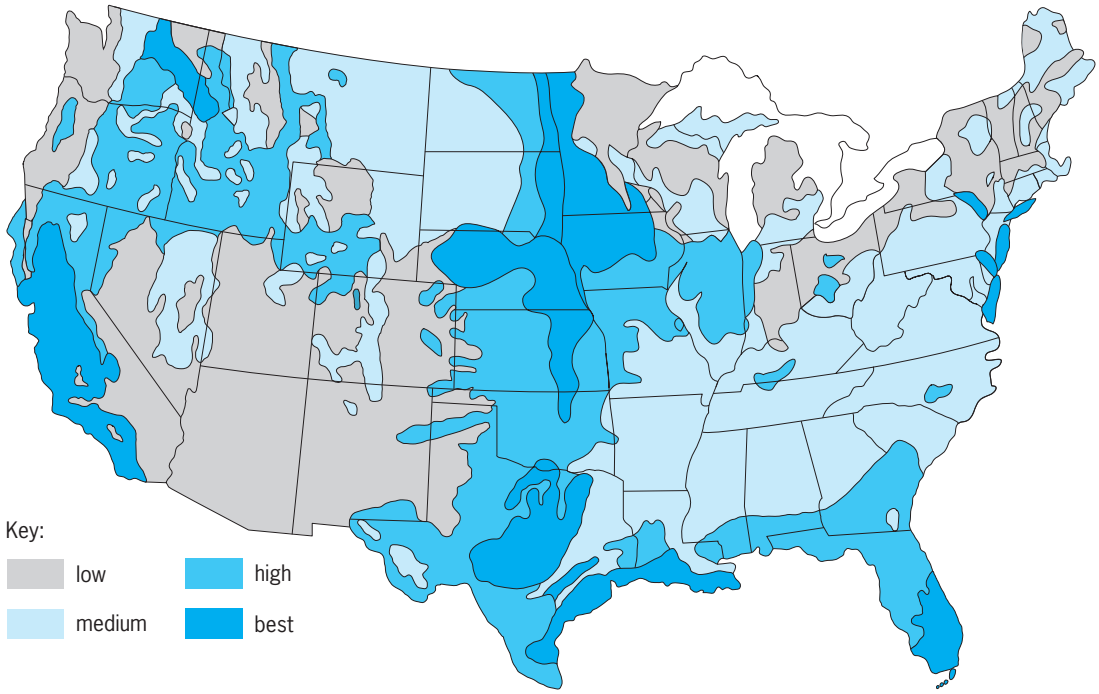


Fig. 13. Potential large-scale productivity of herbaceous biomass with all arable land considered. (*After Solar Energy Research Institute, FY 1982 Biomass Energy Technology Division Annual Technical Progress Report, vol. 1: Executive Summary, SERI/SP-281-2154, Solar Energy Research Institute, 1983*)

TABLE 2. Projected biomass resources for thermochemical conversion

Resource	Available quantity, 10^6 dry tons/year (10^6 dry metric tons/year)					
	1985		1990		2000	
	Maximum	Probable	Maximum	Probable	Maximum	Probable
Wood	464 (421)	167 (152)	429 (389)	154 (140)	549* (498)	71 (64)
Surplus agricultural crops†	45 (41)	34 (31)	36 (33)	31 (28)	33 (30)	39 (35)
Agricultural residues	220 (200)	115 (104)	240 (218)	123 (112)	278 (252)	110 (100)
Municipal solid waste	92 (83)	69 (63)	99 (90)	77 (70)	116 (105)	89 (81)
Other energy crops	8 (7)	104 (94)	69 (63)	203 (184)	172 (156)	247 (224)
Total	829 (752)	489 (444)	873 (792)	588 (533)	1148 (1041)	556 (504)

*Assumes wood from silvicultural energy farms starting in 1985.
†Assumes an aggressive program to establish sorghum as a cash crop that would divert land from corn in 1990 and 2000; data include surplus grain rejected for food processing.

carbohydrates that can be anaerobically digested to methane or fermented to alcohol. Emergent plants, such as cattails and bulrushes, and floating plants, such as water hyacinths, are also highly productive and represent potentially significant feedstock resources for conversion to methane or alcohol. *See* ALGAE.

Conversion to heat and fuel. Liquid and gaseous fuels usually can be transported more easily and efficiently than raw biomass. To be converted to these forms, most biomass must be treated either thermally, chemically, or biologically (**Table 2**). The oldest and simplest method for converting biomass is direct combustion. However, wood is 25 to 50% less efficient as a fuel source than fuel oil, natural gas, or coal. Even dried wood produces lower flame temperatures and requires larger fireboxes and heat transfer surfaces. Converting biomass to heat energy in large boilers is about 70 to 75% efficient, whereas wood can be burned in stoves in homes with efficiency ranges of 50 to 70%.

Gasification. Gasifiers can convert biomass as varied as hardwood chips or animal manure to combustible gases. Most gasifiers are supplied with air to produce a nitrogen-rich low-Btu gas. The oxygen in the air reacts with part of the biomass to produce heat. This heat is then used to produce gas from the remaining biomass. The low-Btu gas is usually used in nearby burners or engines. Conversion efficiencies are generally good, with 60 to 80% of the energy content of the dry biomass transferred to the gas. Use of biomass gasifiers offers flexibility in industrial plant design and provides a way of coupling biomass into existing plants with interchangeability between biomass fuel and conventional fuel (**Table 3**).

Another gasification process uses pure oxygen instead of air to produce synthesis gas (a mixture of carbon monoxide and hydrogen). Methanol, a liquid fuel, can be produced from the synthesis gas by one of several conventional processes. Efficiencies for the overall conversion to methanol are in the range of 40 to 50%. A process for the conversion of methanol to gasoline also is available. *See* ALCOHOL; METHANOL.

Pyrolysis. Producing charcoal by pyrolysis (breaking complex molecules into simpler ones by heat) has been done for centuries. Wood pyrolysis in general leaves a char that constitutes about 30% (by weight) of the wood and contains about 42% of its heat content. It also yields an oil that carries about 32% of the wood's heat content. The pyrolysis gases that are produced are usually burned to supply the processing plant's energy. *See* PYROLYSIS.

Fermentation. The conversion of sugars and starch by fermentation to ethanol is a well-established technology. Sugar from molasses and cheese whey, and starch from grains or tuberous plants are the most commonly used starting materials. However, most of the recent interest has been in fuel ethanol production involving the use of cornstarch, which is converted to glucose, which is then fermented to ethanol. Ethanol fermentation has traditionally been done with yeast in an 18–48-h batch process that produces 8 to 10% ethanol by volume. With modern

TABLE 3. Evaluation of biomass-gasification technologies

Technology	Advantages	Disadvantages
Pyrolysis	Low operating temperature; simple design; no other reactants required	Can get complex char, tar, and gas mixture; tar removal difficult; heat transfer rate may be limited in some cases
Air gasification	Simple design, only one reactor required; less tar than pyrolysis	Gas diluted with nitrogen (N_2) cannot be used for synthesis; low-Btu gas (LBG) may have a reduced combustion efficiency
Oxygen gasification	Medium-Btu gas (MBG) undiluted with N_2 is suitable for synthesis; can operate at high temperature and pressure if desired	Oxygen plant required, greatly increasing the capital cost at small scales
Steam gasification	MBG undiluted with N_2; no oxygen plant required; catalysts can adjust gas composition in gasifier	MBG is methane-rich; must be steam-reformed for methanol synthesis but is an advantage for synthetic natural gas (SNG) synthesis

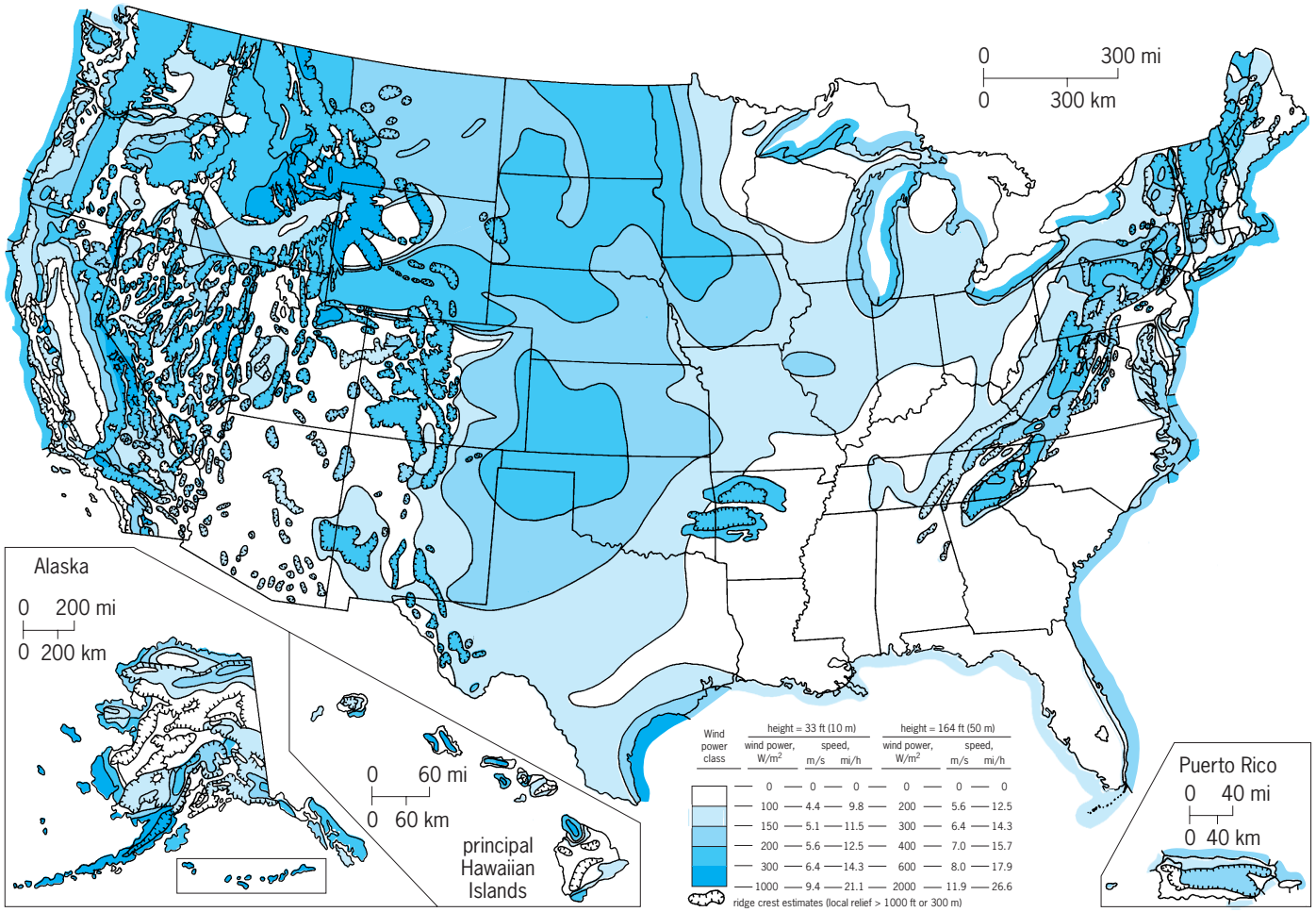


Fig. 14. Wind power in the United States (annual average). (*After Solar Energy Research Institute*)

techniques, the yield of ethanol is about 79 gallons per bushel of corn (8300 liters/m^3); 81% of the manufacturing cost of ethanol is the cost of the starting material. *See* ALCOHOL FUEL; ETHYL ALCOHOL.

Biomass with a high fiber content (woody biomass) is the largest renewable resource available for energy conversion. It is more complex in composition and structure than is starch, and is therefore more difficult to process into potential fuels. Theoretically, the carbohydrates in fibrous biomass can be converted to 100 gallons of ethanol per ton (420 liters/metric ton) of biomass at a 40% net energy conversion compared with 120 gallons/ton (500 liters/metric ton) for corn. One component of woody biomass, lignin, could be chemically removed and used since it has a heating value of about 15,000 Btu/lb (35 MJ/kg) to provide part of the heat needed to convert the carbohydrates to ethanol. *See* FERMENTATION; LIGNIN.

Anaerobic digestion. Anaerobic digestion by bacteria to make methane can use almost any biomass. The technology can be used in a digester of any size, from a small backyard fermenter that uses human waste to the large fermenters that use feedlot waste. Manure is particularly amenable to digestion because of its abundant nutrients, high moisture content, and lack of inhibitors of methane-producing bacteria. The gas produced during digestion is 50 to 70% methane and can be purified, burned as it is, or fed to a natural gas pipeline.

Photosynthetic microorganisms. Some types of microalgae and bacteria have the ability to manufacture energy-rich substances photosynthetically. For example, under appropriate conditions some microalgae will produce increased amounts of oils and lipids that can be used directly as fuel, and some bacteria will evolve hydrogen that also can be used directly as a fuel. Research concentrates on genetic engineering, artificially maintained production systems, and the basic chemistry and physics of photosynthesis.

Wind Energy Conversion Systems

Energy from the wind has been used for centuries to propel ships, to grind grain, and to lift water. Wind turbines extract energy from the wind to perform mechanical work or to generate electricity. Applications to generate electricity are recent and date from about 1930. This discussion concentrates on the production of electricity from wind. *See* WIND.

Wind resource. Wind is a source of energy derived primarily from unequal heating of Earth's surface by

the Sun. Wind speeds are altered by local terrain and vary significantly by season and day.

Figure 14 shows wind resources as annual average wind power classes based on average wind power density. Average wind power density rather than wind speed is used because it incorporates the combined effect of wind speed, wind speed distribution, and air density.

Since the average wind power density is proportional to the cube of wind speed, location or siting is very important. Wind power of at least 200 W/m^2 or wind speed 12 mi/h (19 km/h) is considered prerequisite for additional site evaluation and data gathering. The annual average wind speed is only one consideration. The distribution of wind speed by month, day, and hour is also important, especially in matching power output to power requirements.

The variation of wind speed with height is another important consideration. Generally, wind speed is greater above ground level than it is at the surface due to the frictional effects of Earth's surface.

The output power from a wind turbine can be defined by a power coefficient, C_p, which is a measure of the efficiency with which the rotor extracts power from the wind. The maximum theoretical efficiency, or Betz limit, for a propeller-type rotor is 0.593. Wind turbines available today generally have a C_p between 0.30 and 0.45. *See* WIND POWER.

Wind turbine system. The main components of a wind turbine system (**Fig. 15**) are the rotor, generator, tower, and storage devices (or utility interconnection). The rotor consists of blades and a hub that connects the rotor to the turbine system. The blades can spin around a horizontal axis parallel to the ground or around a vertical axis perpendicular to the ground. Both designs are used in commercial and residential applications. Most commercial models are horizontal-axis wind turbines (HAWTs). They range in output from less than 5 kW to 100 kW. The U.S. Department of Energy has tested a large, 2500-kW prototype HAWT, called the MOD-2, for utility applications; and design of a 3200-kW machine, the MOD-5B, has begun. A wind farm is a system in which many turbines are clustered at especially windy sites to generate megawatts of electricity to sell to utilities.

There are three common configurations of vertical-axis wind turbines (VAWTs); the most developed is the Darrieus design (**Fig. 16**). They, unlike HAWTs, are omnidirectional—they need not be turned as the wind direction changes. Rotors on HAWTs have many blade configurations (**Fig. 17**). In the familiar farm windmill, the blades constitute most of the area swept by the rotor as it spins about its axis. The farm windmill and the Dutch windmill combine drag and lift for inducing rotation. The rotors on these machines have low speed, but have high torque and are well suited to lifting water or grinding grain.

The blades of high-speed turbines are designed as airfoils. In these, the tip-speed ratio—the ratio of the speed of the blade tips to the wind speed—can be as high as 15. A high ratio improves efficiency and reduces the size of drive train components. This is important in electricity generation, whereas torque is important to lifting water or grinding grain. *See* AIRFOIL.

Control of rotor velocity is important to protect the turbine against overspeed in high winds. All wind

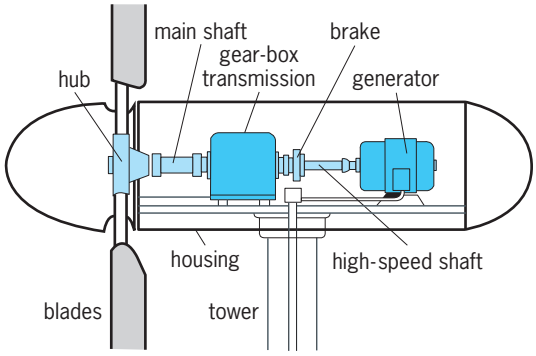


Fig. 15. Main components of a wind turbine. (*After B. H. Bailey, New York State Wind Energy Handbook, New York State Wind Energy Office, 1982*)

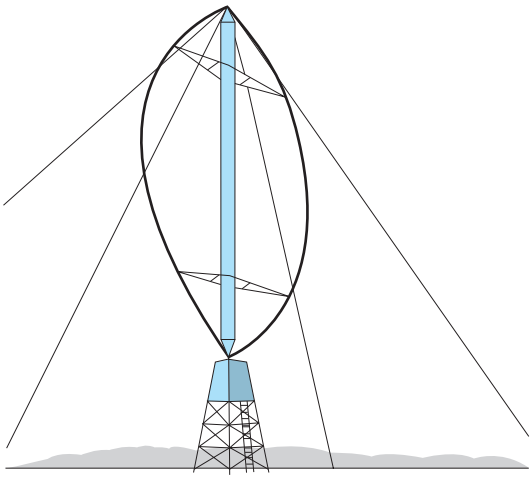


Fig. 16. Darrieus vertical-axis wind turbine. (*After B. H. Bailey, New York State Wind Energy Handbook, New York State Wind Energy Office, 1982*)

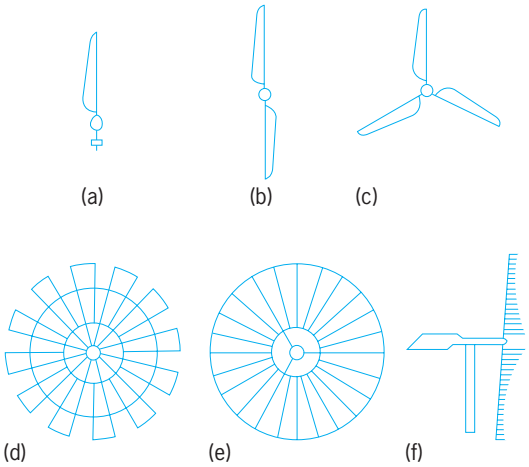


Fig. 17. Some horizontal-axis wind turbine configurations. (*a*) Single-bladed. (*b*) Double-bladed. (*c*) Three-bladed. (*d*) United States farm windmill multibladed. (*e*) Bicycle multibladed. (*f*) Sail wing. (*After Renewable Energy Technology Handbook for Military Engineers, SERI/SP-200-1413, Solar Energy Research Institute, 1982*)

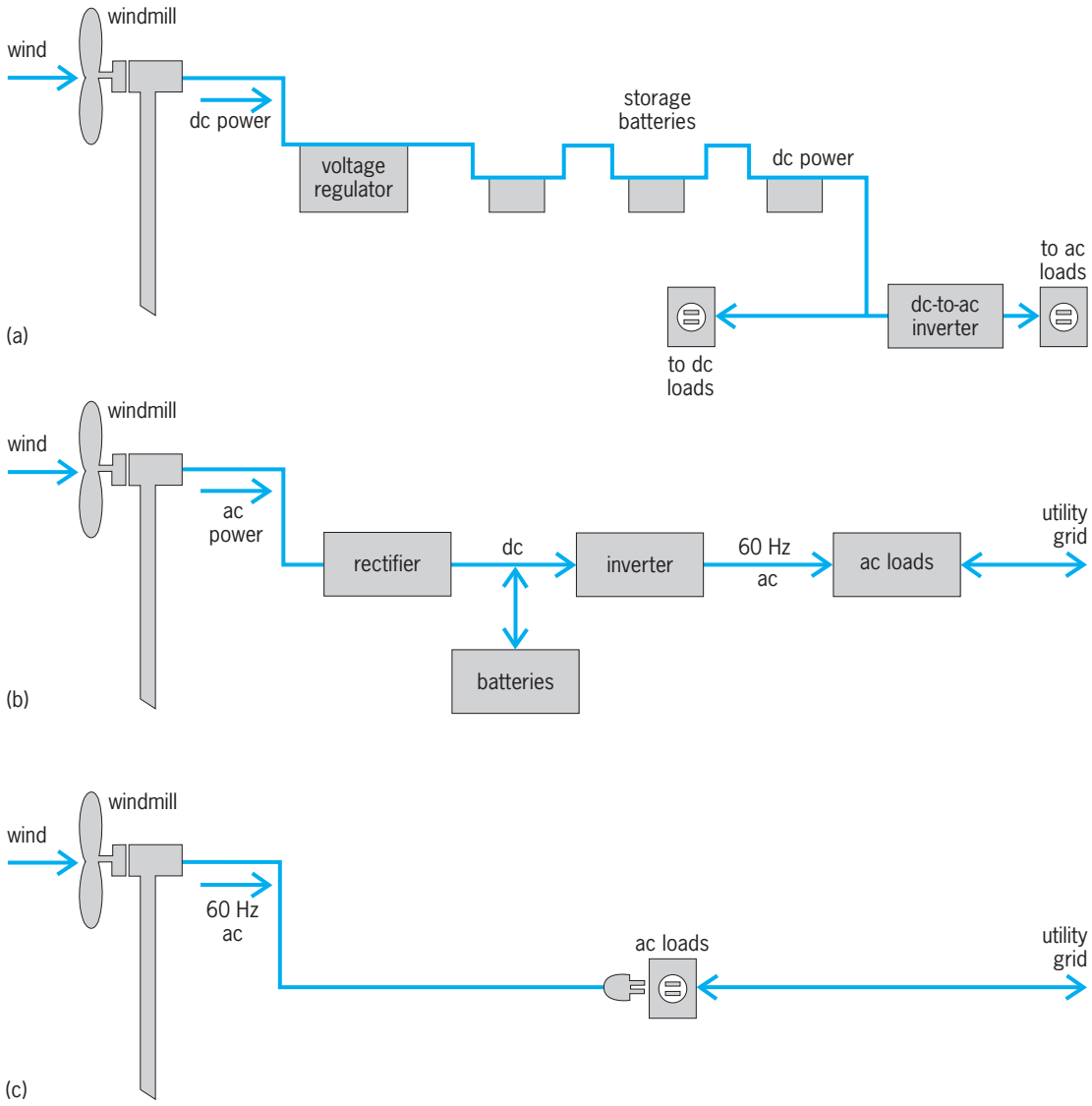


Fig. 18. Possible wind energy system configurations. (*a*) A dc generator windmill. (*b*) Variable-frequency ac generator windmill. (*c*) Induction generator windmill. (*After New York State Wind Energy Handbook, New York State Wind Energy Office, 1982*)

turbines are designed to start at a certain minimum wind speed and stop when the maximum design speed is reached. Various mechanical, hydraulic, and aerodynamic devices are used to brake the rotor or to turn it out of the wind.

The wind energy captured by the rotor is transmitted by shafts and gears to the generator. Generators can produce either direct current or alternating current; the configurations of wind energy systems using different types of generators are shown in **Fig. 18**.

Wind turbines for residential applications are placed on towers usually at least 80 ft (25 m) tall. Towers are either free-standing or guyed, and either type can be a pole or lattice (truss) structure. Installation can cost as much as 30% of the total price of a small wind system. The cost of a tower can be partially offset by the increased output from higher wind speeds at the higher elevation.

Storage of wind energy can be by battery, pumped-hydro equipment, and other devices. Normally, wind turbines are connected to a utility, which provides backup electricity when needed. Excess electricity can, in turn, be fed into the grid and sold to the utility at a negotiated price. *See* TURBINE.

Ocean Thermal Energy Conversion

Ocean thermal energy conversion uses the temperature difference between surface water heated by the Sun and deep cold water pumped from depths of 2000 to 3000 ft (600 to 900 m). This temperature difference makes it possible to produce electricity from the heat engine concept. Since the ocean acts as an enormous solar energy storage facility with little fluctuation of temperature over time, ocean thermal energy conversion, unlike most other renewable energy technologies, can provide electricity 24 h a day. In addition, ocean thermal energy conversion plants can be located onshore, mounted on the continental shelf, or placed on moored or floating platforms. Various shapes have been suggested for floating platforms (**Fig. 19**).

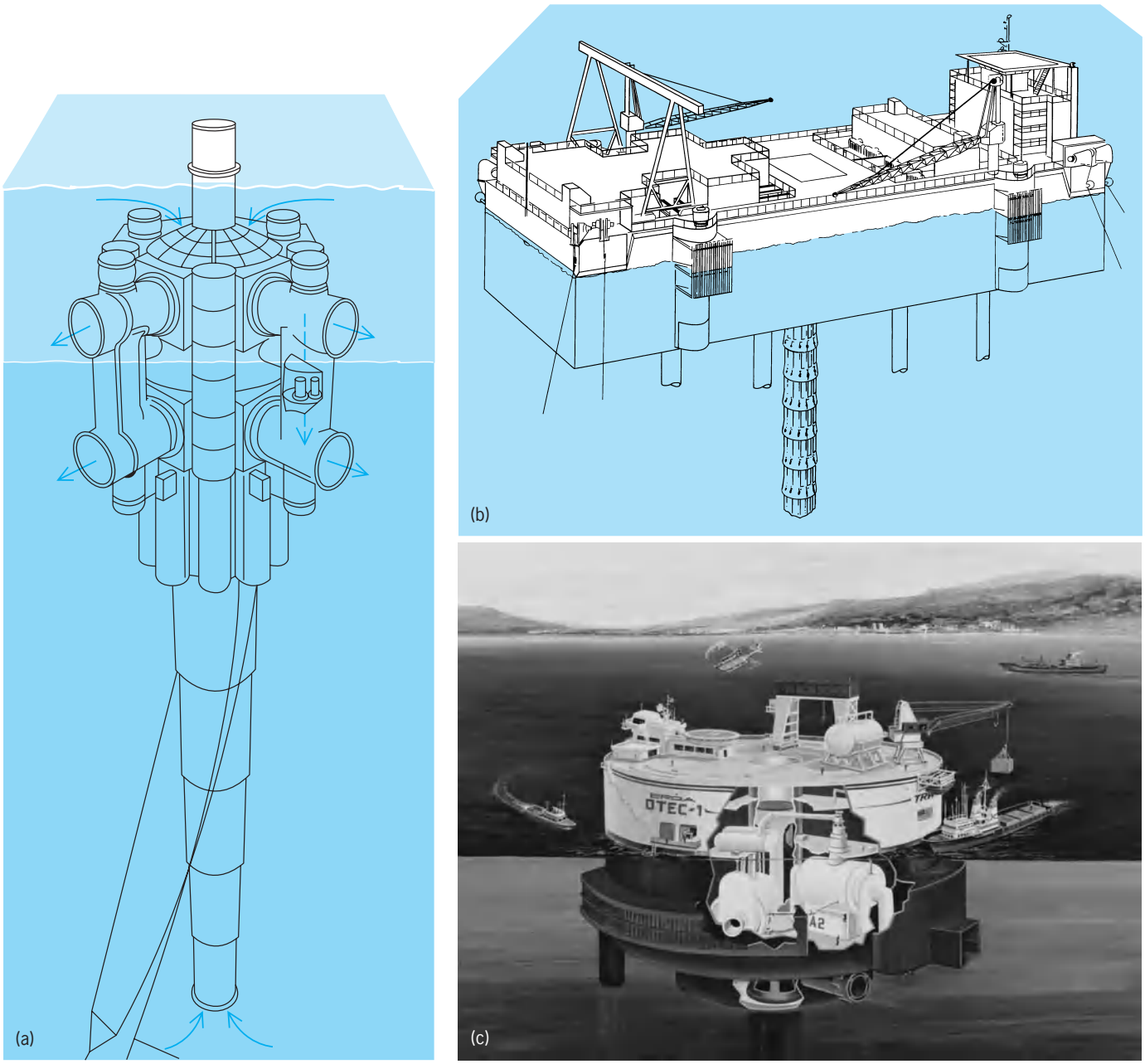


Fig. 19. Platform concepts for ocean thermal energy conversion (OTEC) systems. (*a*) Spar buoy. (*b*) Rectangular barge. (*c*) Cylindrical platform. (*After R. A. Meyers, ed., Handbook of Energy Technology and Economics, John Wiley and Sons, 1983*)

Energy content and electricity. A rough estimate of the thermal energy content of the ice-free, mixed layer of the ocean with a temperature difference of 18°F (10°C) is 70 × 10^{20} Btu (75 × 10^{23} joules). Thus, the ocean thermal energy resource is vast, but that suited for ocean thermal energy conversion facilities is only a fraction of this total, although still an enormous amount. Suitable sites are generally limited to an area within 30° north and 25° south of the Equator. With available technology, an annual average temperature difference of 36°F (20°C) or greater between water at the surface and that at a depth not exceeding 4700 ft (1500 m) is required for continuous and efficient operation of an ocean thermal energy conversion plant. Furthermore, the mean temperature difference during the coldest month should exceed 31°F (17°C). If system efficiency is improved, the required temperature difference will be less and a greater ocean area will be suitable for ocean thermal energy conversion use.

The possibility of generating electricity from solar thermal energy collected and stored by the ocean near the Equator was first suggested by A. d'Arsonval in 1881. In 1930, G. Claude built a power plant in Mantazas Bay, Cuba, and converted ocean thermal energy to electricity, using a temperature difference of 25°F (14°C) between the surface and a depth of 2200 ft (700 m) to generate 22 kWe. In 1979, a small closed-cycle test plant produced 50 kWe and a net output of 12 kWe by using ammonia as the working

fluid. In 1980, a 1000-kW test bed plant was mounted in a converted U.S. Navy tanker to test heat exchangers. A 100-kWe, shore-based plant was tested in 1981 by the Japanese for the island republic of Nauru. The Kyushu Electric Company in early 1983 began operating a 50-kWe test plant in Tokunoshima that uses waste heat from a diesel engine to boost the temperature of the warm seawater.

Conversion systems. There are three principal conversion systems for ocean thermal energy: closed Rankine cycle, open Rankine cycle (Claude cycle), and mist or foam-lift cycle. These systems are listed in order of technological development and in inverse order of theoretical energy conversion efficiency. In other words, the least technologically developed system, the lift cycle, theoretically has the highest potential efficiency. *See* RANKINE CYCLE.

Closed Rankine cycle. In the closed Rankine cycle, warm seawater is pumped through a heat exchanger to transfer stored thermal energy to a working fluid such as ammonia, which has relatively high vapor pressures at low temperatures (**Fig. 20**). The working fluid is vaporized, and the expanding vapor is used to drive a turbine generator. The vapor is condensed by cold seawater in another heat exchanger to complete the cycle. The major advantage of the closed cycle is that the process of heat exchange and energy conversion is established and familiar. The components can be manufactured with existing technology. The major disadvantage is that in adapting an energy conversion process designed for high temperatures and pressures to a low-temperature heat source, inefficiencies in the conversion process become magnified. The heat exchangers must be large, and a large volume of water must be pumped through them. About half of the total available temperature difference is required to transfer heat in the evaporator and condenser. Since the efficiency is a direct function of the temperature difference available after the transfer, the small but inherent drop in temperature through the heat exchange process reduces an already low temperature potential. *See* HEAT EXCHANGER.

Large heat exchanger surfaces and low working temperatures also make the control of slime formation (biofouling) and corrosion very critical. The design, construction, and operation of large closed-Rankine-cycle ocean thermal energy conversion facilities present complex engineering problems even though no technological breakthroughs are needed for any major structural or operating components.

Open Rankine cycle (Claude cycle). Claude showed that the heat exchange process for both evaporation and condensation could be performed by using warm seawater itself as the working fluid. Direct-contact heat exchange eliminates the need for large, costly heat exchangers. The evaporator and direct-contact condenser possibly can be made with inexpensive materials such as poly(vinyl chloride) and concrete.

In the open-cycle process, steam is generated from warm seawater by flash (very rapid) evaporation in an evacuated chamber. The steam turns a turbine and

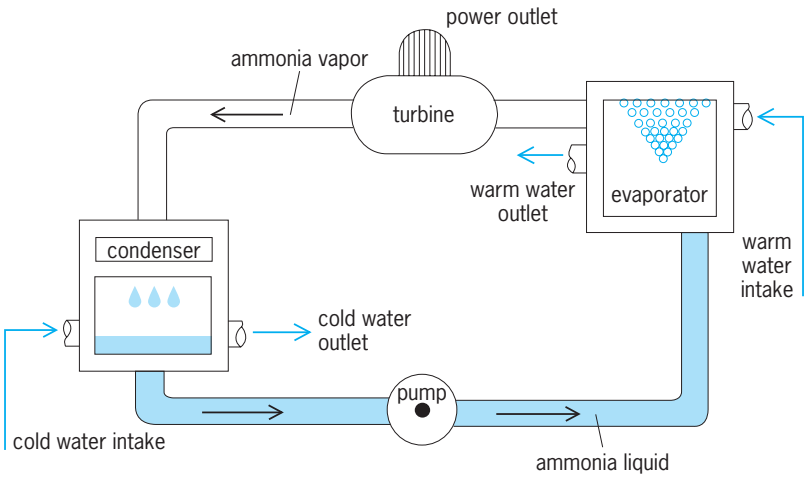


Fig. 20. Closed Rankine (ammonia) cycle. (*After The Ocean Option, SERI/SP-732-334, Solar Energy Research Institute, 1980*)

is condensed by direct contact with cold seawater to complete the cycle. The absence of any barrier between the working fluid and the heating and cooling source makes possible a higher overall system efficiency and lower seawater flow requirement than in the closed-cycle process. Using seawater rather than a substance like ammonia as the working fluid also lessens safety problems. The need for mechanical or chemical measures to control biofouling is also reduced.

The open-cycle process is similar to certain desalination processes, and if a typical closed-cycle surface condenser is used, fresh water can be produced as a by-product at some sacrifice in efficiency. The use of ocean thermal energy conversion facilities to produce fresh water as well as electricity may prove attractive in some locations. *See* WATER DESALINATION.

The elimination of volatile working fluids and heat exchangers means that the open-cycle process will function at very low steam pressures. The steam turbine, in turn, must be very large and housed in a correspondingly large vacuum vessel. Present technology is limited to relatively small plants.

Lift cycles. There are several other approaches to the open cycle that involve the conversion of the thermal gradient into a hydraulic head through the lift of seawater in vertical two-phase flow. The water column created can drive a turbine, as in a hydroelectric dam. An advantage of this approach is that the hydraulic turbine is more compact and less costly than the steam turbine required in the Claude system.

Warm seawater is introduced as a mist that is then lifted against gravity by the flow of steam from a low-pressure region to a lower-pressure region. This action is similar to the naturally occurring cycle that converts water evaporated by solar energy to hydropower. In ocean thermal energy conversion lift cycles, an artificial hydrological cycle is created.

Robert L. San Martin

Bibliography. B. Anderson, *Solar Building Architecture*, 1990; J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, 2d ed.,

1991; M. A. Green, *Third Generation Photovoltaics: Advanced Solar Energy Conversion*, 2003; *Kirk-Othmer Encyclopedia of Chemical Technology*, vol. 17, 4th ed., 1995; Institute of Gas Technology, *Energy from Biomass and Wastes Symposium,* 1989; G. N. Tiwarik, *Solar Energy: Fundamentals, Design, Modelling and Applications*, 2002; S. Wieder, *An Introduction to Solar Energy for Scientists and Engineers,* 1982, reprint 1990.

Solar heating and cooling

The use of solar energy to produce heating or cooling for technological purposes. When the Sun's short-wave radiation impinges upon a blackened surface, much of the incoming radiant energy can be absorbed and converted into heat. The temperature that results is determined by: the intensity of the solar irradiance; the ability of the surface to absorb the incident radiation; and the rate at which the resulting heat is removed. By covering the absorbing surface with a material such as glass, which is highly transparent to the Sun's short-wave radiation but is opaque to the long-wave radiation emitted by the Sun-warmed surface, the effectiveness of the collection process can be greatly enhanced. The energy which is collected can be put to beneficial use at many different temperature levels to accomplish: distillation of seawater to produce salt or potable water; heating of swimming pools; space heating; heating of water for domestic, commercial, and industrial purposes; cooling by absorption or compression refrigeration; cooking; and power generation by thermal or photovoltaic means. *See* SOLAR ENERGY.

Distillation

The oldest of these applications dates back to a 50,000-ft^2 (4600-m^2) installation built in Chile in 1872 to distill saline water and make it potable. **Figure 1** shows in cross section the type of glass-roofed solar still used more than a century ago in Chile and used again in modern times on the Greek islands in the eastern Mediterranean and in central Australia. When the Sun shines through the glass cover into the salt or brackish water contained within the concrete channel, the water is warmed and some evaporates to be condensed on the underside of the glazing. The condensate runs down into the scuppers and then into a suitable container. More than 6000 gal (23,000 liters) of pure drinking water were produced on each sunny day by the Chilean still, and the larger version now operating at Coober Pedy, Australia, produces nearly twice as much.

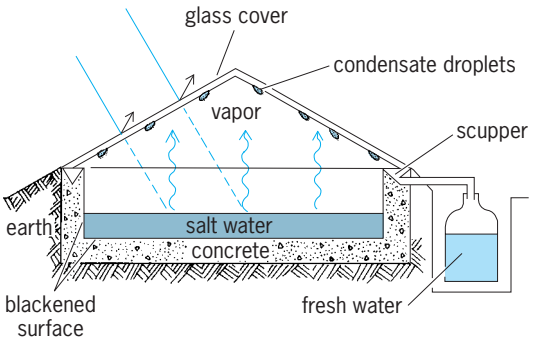


Fig. 1. Roof-type solar still.

Production of salt from the sea has been accomplished for hundreds of years by trapping ocean water in shallow ponds at high tide and simply allowing the water to evaporate under the influence of the Sun. The residue contains all of the compounds that were present in the seawater, and it is sufficiently pure for use in many industrial applications. *See* DISTILLATION; WATER DESALINATION.

Swimming Pool Heating

Swimming pool heating is a moderate-temperature application which, under suitable weather conditions, can be accomplished with a simple unglazed and uninsulated collector similar to the black polyethylene extrusion shown in **Fig. 2**. When the water to be heated is at almost the same temperature as the surrounding air, little or no heat will be lost from the absorber and so there is no need for either glazing or insulation. Such collectors are usually designed to empty themselves when the circulating pump is shut off so they can avoid the danger of freezing at night. For applications where a significant temperature difference exists between the fluid within the collector passages and the ambient air, both glazing and insulation are essential.

Space Heating

Space heating can be carried out by active systems which use separate collection, distribution, and storage subsystems (**Figs. 3** and **4**), or by passive designs which use components of a building to admit, store, and distribute the heat resulting from absorbing the incoming solar radiation within the building itself.

Passive systems. Passive systems can be classified as direct-gain when they admit solar radiant energy directly into the structure through large south-facing windows (**Fig. 5**), or as indirect-gain when a wall (**Fig. 6**) or a roof (**Fig. 7**) absorbs the solar radiation, stores the resulting heat, and then transfers it into the building. If the absorber and the storage components are not a part of the building fabric but are separate subsystems which operate by the natural circulation of warmed or cooled air, the term isolated-gain is generally used (**Fig. 8**).

Passive systems are generally effective where the number of hours of sunshine during the winter months is relatively high, where moderate indoor temperature fluctuations can be tolerated, and where the need for summer cooling and dehumidification is moderate or nonexistent.

Direct-gain systems. Most passive systems make use of the fact that, in winter, whenever the Sun is above the horizon it is in the southern part of the sky, and its altitude above the horizontal plane at noon is relatively low compared to the much higher position which it will attain in summer. This means that in winter

when heat is needed, vertical south-facing windows admit solar radiation freely as long as the Sun is shining. The use of double glazing reduces the transmission of the incoming solar radiation by about 16%, but it can halve the thermal loss due to the indoor-outdoor temperature difference. The use of movable insulation which can be placed over the windows at night and removed during the sunlit hours of the day can also greatly improve the performance of direct-gain systems. Window area on surfaces other than those facing south should be kept to the minimum permissible under local building codes. The building components (walls and floors) on which the solar radiation falls should possess as much mass as possible so that the excess solar heat gained during the day can be stored for use at night.

Indirect-gain systems. The indirect-gain concepts shown by Figs. 6 and 7 interpose a thermal mass between the incoming solar radiation and the space to be heated. The system shown in Fig. 6 uses a south-facing glazed wall of concrete or masonry with an air space between the wall's outer surface and the single or double glazing. This is known as a Trombe wall. Vents with dampers are provided near the floor and at the ceiling level so that cool air can be drawn by chimney action from the room into the space between the glass and the wall. The air is heated by contact with the wall, and rises to reenter the room at the ceiling level. At night, the chimney action stops and the dampers are closed to prevent the downflow of cooled air back into the room.

As the Sun's rays warm the outer surface of the concrete, a wave of heat begins to move slowly (at about 2 in. or 5 cm per hour) through the wall. The thickness of the wall is chosen to delay the arrival of the wave of warmth at the indoor surface until after sunset, when the long-wave radiation emitted by the wall will be welcome. Heat also flows outward from the wall to the glazing and thence to the outdoor environment, but this can be minimized by the use of movable insulation within the air space. Windows may be introduced into the Trombe wall, and the building need not be restricted to one story.

Another indirect-gain system, shown in Fig. 7, uses enclosed bags of water, called thermoponds, which are supported by a heat-conducting roof-ceiling and covered by horizontally movable insulating panels. This system collects solar energy during winter days by rolling the insulation away to a storage area, thus admitting Sun's rays into the thermoponds. At night, the insulating panels are rolled back to provide the insulation needed to retain the collected heat.

In summer, the operation is reversed, and the insulating panels are rolled back at night to expose the thermoponds to the sky and thus to enable them to dissipate heat that has been absorbed from the building during the day. Convection to cool night air, when it is available, and radiation to the sky on clear nights are two of the natural processes by which heat can be rejected. The third process, and the most potent, is evaporation, which can occur when the ponds are provided with exposed water surfaces by flooding or spraying them. *See* HEAT TRANSFER.

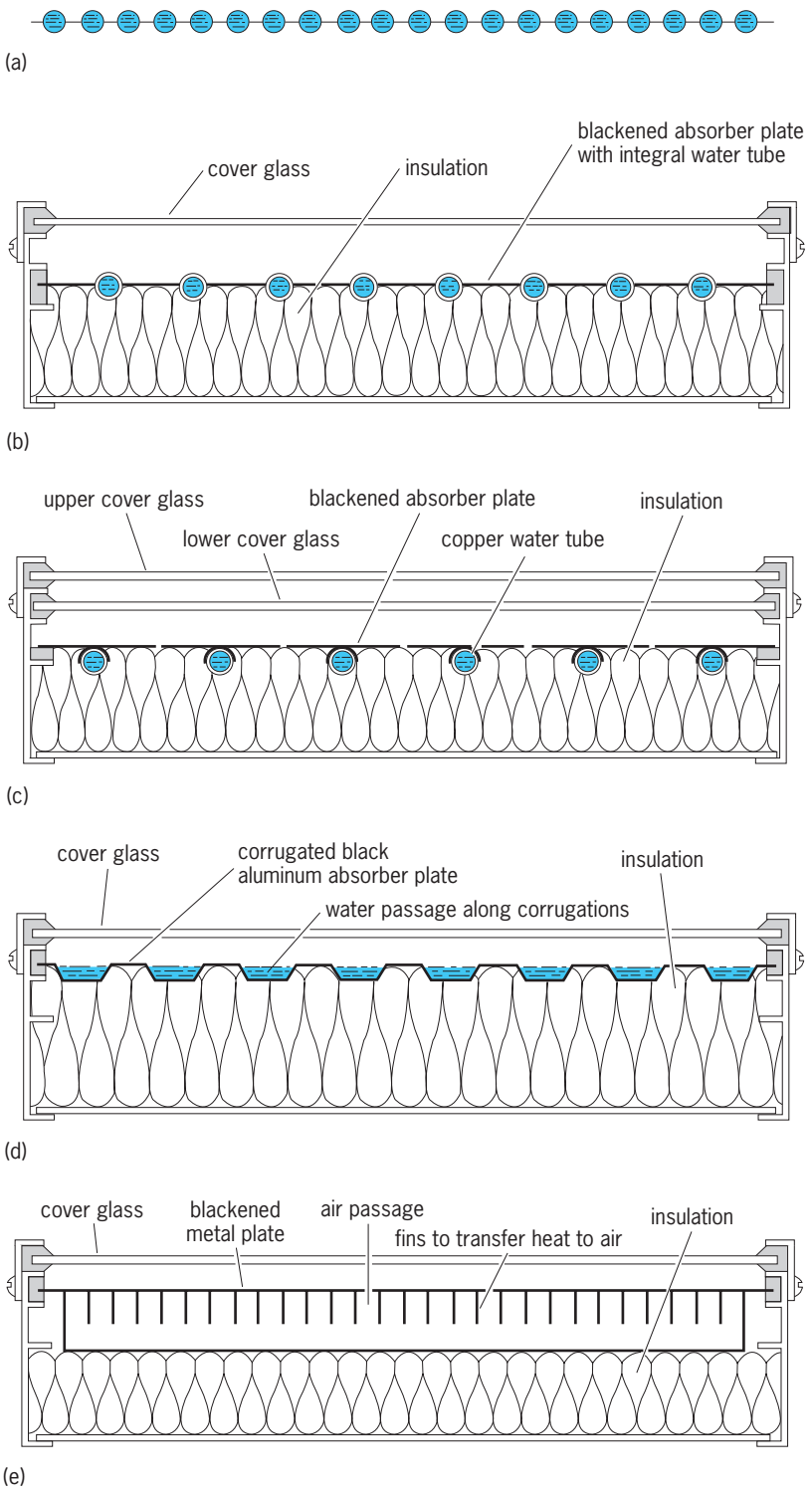


Fig. 2. Flat-plate collector types. (*a*) Extruded black plastic swimming pool heater. (*b*) Single-glazed collector with tube-in-sheet. (*c*) Double-glazed collector with extruded aluminum absorber element and copper water tube. (*d*) Open-flow collector used in the Thomason Solaris system. (*e*) Air heater using finned aluminum heat absorber.

Another indirect-gain system makes use of a greenhouse attached to the south side of a building to gather heat during the day. The wall between the glazed space and the building is warmed by the Sun's rays and by contact with the warm air in the

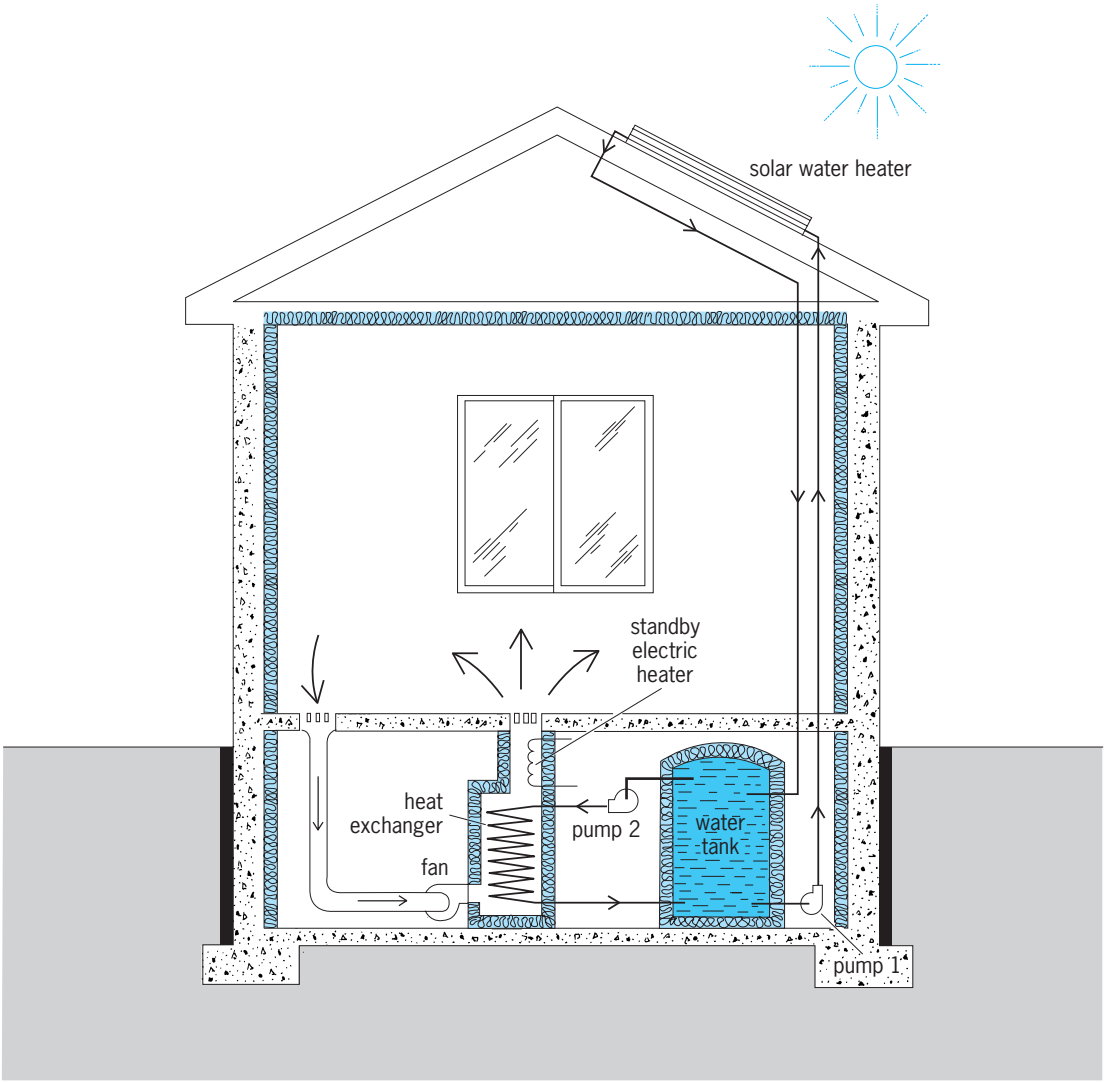


Fig. 3. Active system using roof-mounted solar water heater with storage tank in basement.

greenhouse. Additional stored heat can be transferred into the house by opening windows or vents. The warmth which is gathered even during cold winter days can thus be used to aid in heating the residence as well as to provide an environment in which plants can grow. *See* CONVECTION (HEAT); HEAT RADIATION.

Isolated-gain systems. The isolated-gain system shown in Fig. 8 uses a thermosyphon heater to warm air, which rises as it becomes hot, until it encounters the entrance to the rock-bed thermal storage component. The rocks abstract heat from the air which moves downward as it becomes cooler. The air eventually descends to the bottom of the rock bed and thence to the inlet of the heater. The cycle continues as long as the Sun shines on the collector, and heat is stored throughout most of the day. At night, when the occupants of the house sense the need for heat, the damper settings are changed and warm air from the top of the rock bed is admitted to the living space. At the same time, cool air from the north wall of the house is allowed to flow downward through ducts to enter the base of the rock bed. There the air is warmed, and the heat which has been stored during the day is used to warm the house at night.

Active systems. Active systems may use either water (Fig. 3) or air (Fig. 4) to transport heat from roof-mounted south-facing collectors to storage in rock beds or water tanks. The stored heat may be withdrawn and used directly when air is the transfer fluid. When the heat is collected and stored as hot water, fan-coil units are generally used to transfer the heat to air which is then circulated through the warmed space. Each system has both advantages and disadvantages.

Air cannot freeze or cause corrosion, and leakage is not a serious problem. Water requires relatively small pipes compared with the ducts needed to transport the same amount of heat in the form of warm air. Water tanks can store more than three times as much heat as rocks in a given volume per degree of temperature change, but rock beds are considerably lower in cost than water tanks, and rocks can

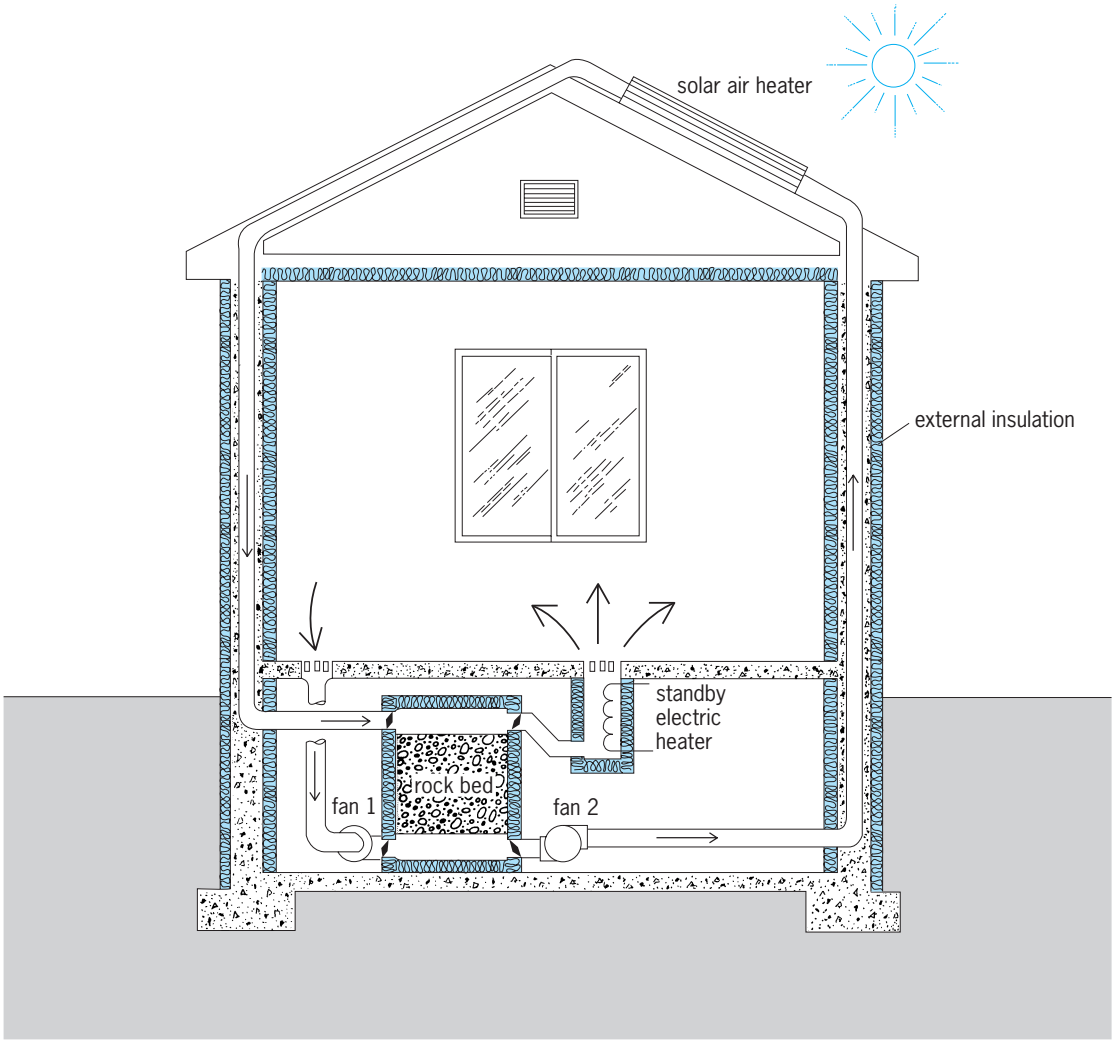


Fig. 4. Active system using air heater with rock-bed heat storage in basement.

tolerate virtually any temperature while water will boil at 212°F (100°C) unless its pressure is raised above atmospheric. The choice between water and air systems must be made carefully, taking into account the many features of each.

Heat storage. Heat storage can be accomplished with specific-heat materials such as water or rocks, which can store and discharge heat by simply undergoing a change in temperature. A different process is involved in the heat-of-fusion materials, which, like water at 32°F (0°C), can freeze when heat is removed and melt again when the same amount of heat is returned to them. Ice has the unique property of increasing in volume by 7% compared with the water from which it is formed, but virtually all other substances become denser when they solidify and so their solid components sink instead of floating. A number of materials are available which, like the well-known Glauber's salt, change their state from liquid to solid and back again at temperatures which are useful for either heating or cooling. None has yet been found, however, which possesses all of the attributes required for successful application to solar heating and cooling systems.

Standby electrical sources. Standby electrical sources are shown in Figs. 3 and 4, since some method of providing warmth must be included for use when

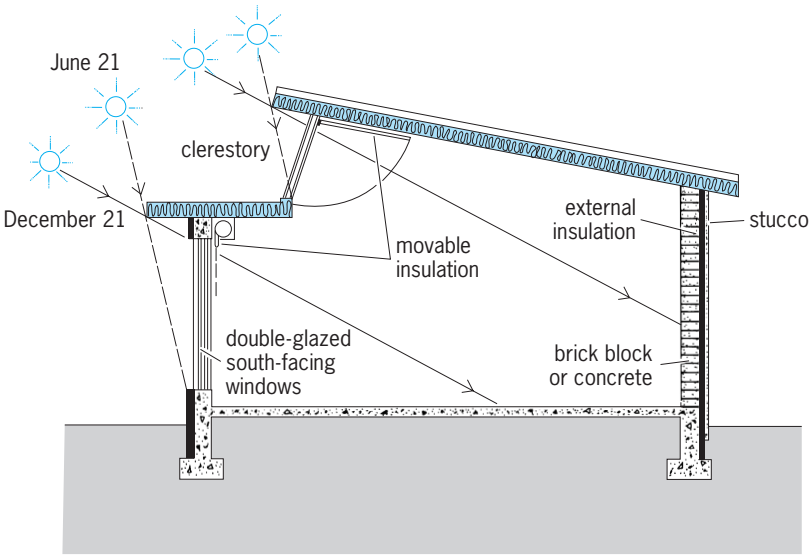


Fig. 5. Direct-gain passive system using large south-facing windows.

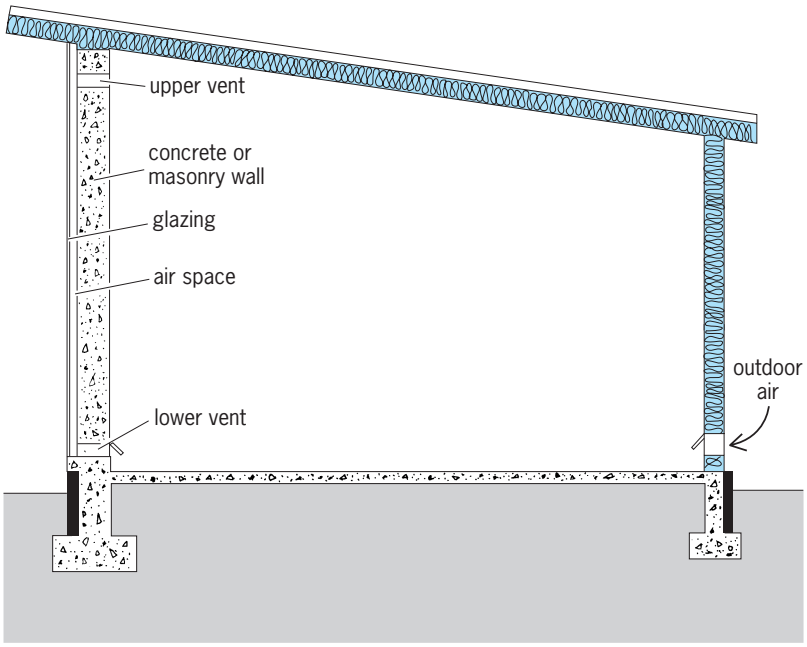


Fig. 6. Indirect-gain passive system using Trombe wall.

the Sun's radiant energy is inadequate for long periods of time. The standby heater may be something as simple as a wood-burning stove or fireplace, or as complex as an electrically powered heat pump. Simple electrical resistance heaters are frequently chosen for this service because of their low first cost, but their operating cost can become excessive in applications where they must be used for long periods of time. The rapidly escalating cost of electricity must be borne in mind when the standby energy source is selected. *See* COMFORT HEATING; HEAT PUMP.

Service Water Heating

Solar water heating for domestic, commercial, or industrial purposes is one of the oldest and most successful applications of solar-thermal technology. The most widely used water heater, and one that is suitable for use in relatively warm climates where freezing is a minor problem, is the thermosyphon type (**Fig. 9**). A flat-plate collector of one of the types shown in Fig. 2 is generally used with a storage tank which is mounted above the collector. A source of water is connected near the bottom of the tank, and the hot water outlet is connected to its top. A downcomer pipe leads from the bottom of the tank to the inlet of the collector, and an insulated return line runs from the top of the collector to the upper part of the storage tank which is also insulated.

The system is filled with water, and when the Sun shines on the collector, the water in the tubes is heated. It then becomes less dense than the water in the downcomer, and the heated water rises by thermosyphon action into the storage tank. It is replaced by cool water from the bottom of the tank, and this action continues as long as the Sun shines on the collector with adequate intensity. When the Sun moves away, the thermosyphon action stops, but the glazing on the collector minimizes heat loss from this component and the insulation around the storage tank enables it to retain the collected heat until needed. The system must be prevented from freezing, and the components must be strong enough to withstand operating water pressure.

For applications where the elevated storage tank is undesirable or where very large quantities of hot water are needed, the tank is placed at ground level. A small pump circulates the water in response to a signal from a controller which senses the temperatures of the collector and the water near the bottom of the tank. Heat exchangers may also be used with water at operating pressure within the tubes of the exchanger and the collector water outside to eliminate the necessity of using high-pressure collectors. Antifreeze substances may be added to the collector water to eliminate the freezing problem, but care must be taken to prevent any possibility of cross-flow between the collector fluid and the potable water. Many plumbing codes require special heat exchangers which interpose two metallic walls between the two fluid circuits. *See* HOT-WATER HEATING SYSTEM.

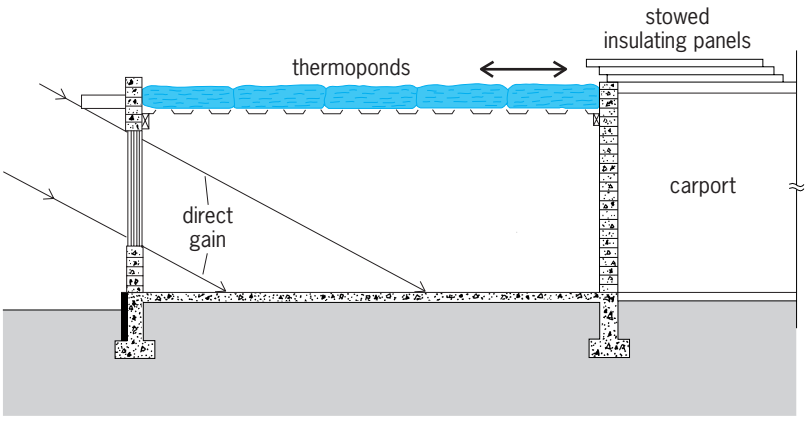


Fig. 7. Indirect-gain Skytherm system with movable insulation.

Cooling

Cooling can be provided by both active and passive systems.

Active cooling systems. The two feasible types of active cooling systems are Rankine cycle and absorption. The Rankine cycle system uses solar collectors to produce a vapor (steam or one of the fluorocarbons generally known as Freon) to drive an engine or turbine. A condenser must be used to condense the spent vapor so it can be pumped back through the vaporizer. The engine or turbine drives a conventional refrigeration compressor which produces cooling in the usual manner. *See* RANKINE CYCLE; REFRIGERATION.

The absorption system uses heat at relatively high temperature (180 to 200°F or 82 to 93°C) and employs a hygroscopic solution of lithium bromide to absorb water vapor at a very low pressure. The evaporating water is cooled to a temperature low enough (about 45°F or 7°C) to provide the chilled water needed to produce air conditioning. The heat introduced into this cycle by the high-temperature

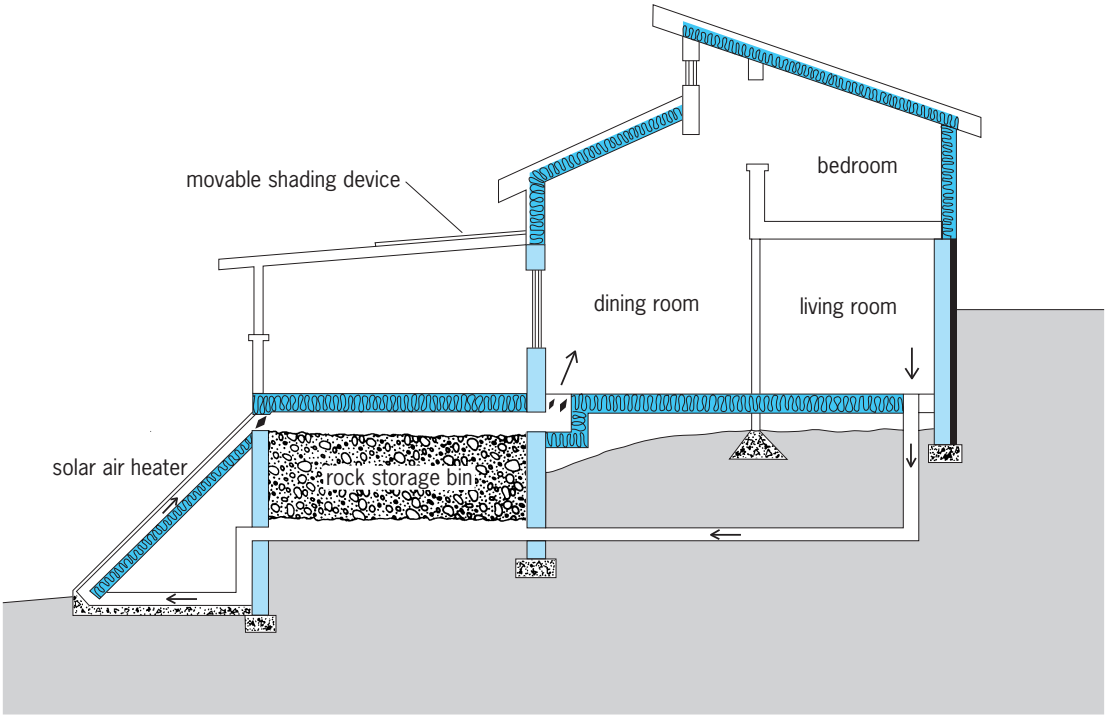


Fig. 8. Isolated-gain system with thermosyphon air heater and rock bed.

activating water and the low-temperature chilling water is removed by a third stream of water which has been cooled to below the outdoor air temperature by passing through an evaporative cooling tower. The cycle, simple in principle but much more complex in practice, is in wide use in large buildings where the necessary hot activating water is available either from solar collectors or as waste heat from other processes. Use of either Rankine cycle or absorption solar cooling systems for residences is restricted to a few experimental installations, but both hold promise for future development.

Passive cooling systems. Passive cooling systems make use of three natural processes: convection cooling with night air; radiative cooling by heat rejection to the sky on clear nights; and evaporative cooling from water surfaces exposed to the atmosphere. The effectiveness of each of these processes depends

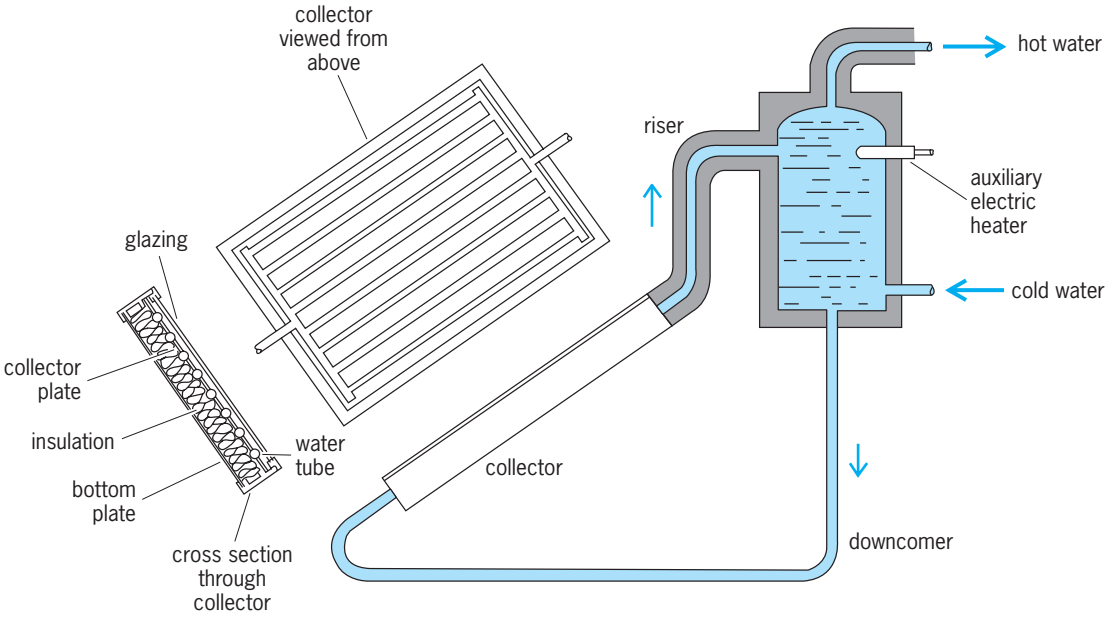


Fig. 9. Passive domestic water heater using thermosyphon system.

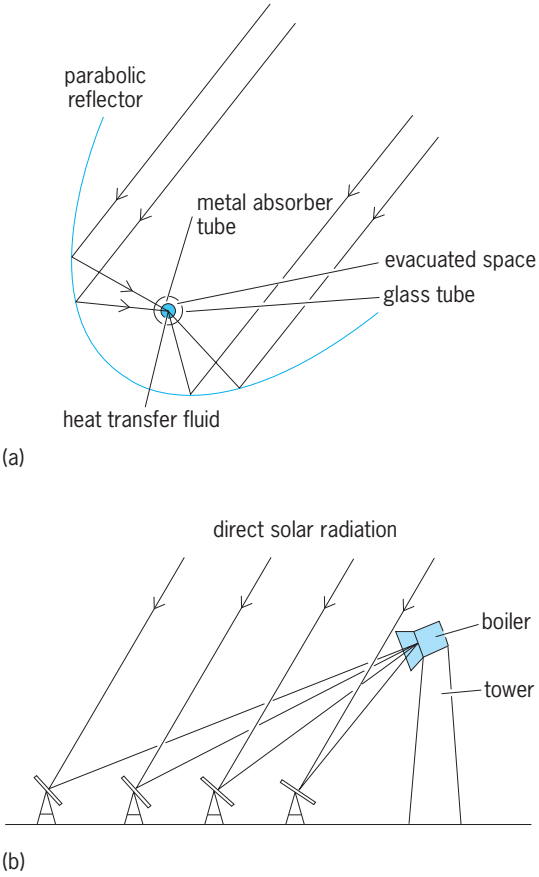


Fig. 10. Concentrating collectors. (*a*) Parabolic trough, for use in installations of moderate size. (*b*) Power tower, for large utility-type plants.

upon local climatic conditions. In the hot and dry deserts of the southwestern United States, radiation and evaporation are the two most useful processes. In hot and humid regions, other processes must be employed which involve more complex equipment that can first dehumidify and then cool the air.

Power Generation

Electric power can be generated from solar energy by two processes, the first of which uses the Rankine cycle described previously. In order to attain vapor temperatures high enough to make the cycle operate efficiently, concentrating collectors must be used. For installations of moderate size, parabolic troughs (**Fig. 10***a*) are adequate, but for very large utility-type plants the "power tower" concept (Fig. 10*b*) is preferred. Concentrators can use only the radiation which comes directly from the solar disk, so some means of tracking must be provided and energy must be stored during the daylight hours to ensure continuity of operation during intermittent cloudy periods. Storage systems for night-long operation are not yet available. Concentrating systems cannot make effective use of the diffuse radiation that comes from the sky, and so their use at the present stage of development of solar technology is limited to areas where bright sunshine is generally available.

Electric power can also be produced from solar radiation, without need for a thermal cycle, by photovoltaic cells which convert radiant energy directly into electricity. Originally used primarily in the United States and Soviet space programs, silicon solar batteries can now be produced at a cost low enough to justify their use in special circumstances for terrestrial applications. One development uses concentrators similar to those in Fig. 10*a* to enable a single solar cell to do the work of ten; the heat which must be removed to make the cells function efficiently can be used for many purposes requiring hot water or hot air. Solar cells produce direct current which can be stored in ordinary automobile-type batteries; the direct current can be converted to alternating current at the standard frequency, 60 Hz in the United States, and at any necessary voltage. One of the goals of the Department of Energy is the reduction of the cost of photovoltaic cells to the point where residential rooftop units become economically feasible. *See* SOLAR CELL.

Conclusion

Heating of space and service hot water is currently practical and economical. Passive heating and cooling are coming into wide use as the cost of alternative fuels continues to rise. Active cooling and power generation are too expensive for general use, but it is expected that the cost of these systems will diminish in the future. The major unsolved problem is how to store large quantities of energy for long periods of time. *See* ENERGY STORAGE.

John I. Yellott

Bibliography. American Society of Heating, Refrigerating and Air Conditioning Engineers, *Active Solar Heating Systems Design Manual,* 1988; B. Anderson and M. Wells, *Passive Solar Energy: The Homeowner's Guide to Natural Heating and Cooling*, 2d ed., 1993; D. Halacy, *Understanding Passive Cooling Systems,* 1987; B. Norton, *Solar Energy Technology*, 1991.

Solar magnetic field

The magnetic field rooted in the Sun and extending out past the planets into the solar system. The field at the Sun's surface is detected remotely by its effect (the Zeeman effect) on atoms whose radiation is observed from Earth. This technique was first applied at the Mount Wilson Observatory in 1908 by G. E. Hale to detect the fields in sunspots. In 1952, H. D. Babcock and his son H. W. Babcock used a scanning technique to make the first magnetic maps of the entire visible disk of the Sun. Their daily "magnetograms" soon revealed a variety of magnetic features, including bipolar regions associated with sunspot groups, and unipolar regions whose fields extend far from the Sun and are responsible for recurrent geomagnetic activity at Earth. The field strengths range from a few gauss (a few hundred microtesla) in quiet areas to 3500 gauss (0.35 tesla) in sunspots. In 1962, the *Mariner 2* spacecraft, en route to Venus, made the first on-site sampling of the extended solar field in space. The average field strength was only 50 microgauss (5 nanotesla), reflecting the rapid (inverse

square) fall-off of field strength with distance from the Sun. *See* ZEEMAN EFFECT.

Bipolar regions and sunspots. Solar magnetic fields are related to the 11-year variation in the occurrence of sunspots. As the new sunspot cycle begins, concentrations of bipolar flux break through the Sun's surface in each hemisphere, beginning at about 40° latitude and gradually progressing toward the solar equator over the next few years. The emergence typically takes a few days, depending on the size of the region, but the subsequent spreading and decay may take several months so that a bipolar region may be visible from Earth again during the next 27-day rotation of the Sun on its axis. Thus, as the sunspot cycle advances and the eruption rate exceeds several per day, the sunspot belts become filled with the remnants of old bipolar magnetic regions.

Unipolar regions and polar fields. If the eruptions of bipolar flux were randomly oriented, the resulting patterns might be a complicated mixture of positive and negative polarities. However, the orientations are not random. The bipolar regions are oriented approximately east-west with their leading polarities (in the sense of solar rotation) all positive in the northern hemisphere and negative in the southern hemisphere during a given sunspot cycle (Hale's law). During the next 11-year cycle, the polarity scheme is reversed, with the leading polarities negative in the north and positive in the south. This systematic property of the sources of flux leads to systematic properties of the evolving magnetic patterns, and occasionally produces a large unipolar region whose opposite-polarity companion is far away.

In fact, the emerging bipolar magnetic regions are not oriented exactly east-west, but are tilted slightly so that their trailing polarities are closer to the Sun's poles and the leading polarities are closer to the equator. This small but systematic effect ultimately leads to the formation of unipolar regions at the Sun's poles, positive in one hemisphere and negative in the other, with an approximate 5-year lag between the peak in the eruption rate and the subsequent peaks in the strengths of the polar fields. When bipolar regions erupt with the opposite polarities during the next sunspot cycle, they will cause the polar fields to become reversed a few years later. This is why the solar magnetic cycle is referred to as a 22-year cycle rather than as an 11-year cycle like that of sunspots.

Mixed-polarity regions. Improvements in spatial resolution and sensitivity have revealed a mixed-polarity component of the solar magnetic field. This component acts like a weak background noise in which small-scale fields are continuously emerging, bumping into their counterparts of opposite polarity, and disappearing again. At sunspot minimum, this process involves more flux eruption than occurs in the larger, but less frequent, bipolar magnetic regions. However, because of its random occurrence, the mixed-polarity component seems to have little net effect on the large-scale solar magnetic field and its interplanetary extension.

Outward extension of the field. The occasional formation of large unipolar magnetic regions, isolated from their opposite-polarity counterparts by great distances, causes their magnetic field lines to extend far from the Sun without looping back to the surface. In these locations, material is able to stream outward from the Sun at relatively high speeds of 750 km/s (470 mi/h). When such open-field regions are located at low latitudes, their high-speed streams repeatedly sweep past Earth as the Sun rotates, inducing auroras and recurrent geomagnetic disturbances. *See* AURORA; GEOMAGNETIC VARIATIONS; MAGNETOSPHERE; SOLAR WIND.

The continual readjustment of the evolving patterns of magnetic fields leads to a variety of transient plasma processes, including solar flares, eruptive prominences, and coronal mass ejections. Without the solar magnetic field, none of these explosive phenomena would occur, and the Sun would be a much more placid late-type star.

Solar variability and other stars. As magnetic fields from bipolar regions spread out on the surface of the Sun, they heat the overlying chromosphere, causing an increase in ultraviolet radiation. The integrated (Sun-as-a-star) component of chromospheric radiation has been measured since 1976 and shows a clear 11-year variation, suggesting that the sunspot cycle would be visible to someone observing the Sun from elsewhere in the Milky Way Galaxy, and conversely that this technique could be used to see solar magnetic cycles in other stars. This idea motivated O. C. Wilson to begin such a survey at the Mount Wilson Observatory in 1966, and the continuation of this work has led to the identification of sunspot cycles in several late-type stars. *See* STELLAR MAGNETIC FIELD.

Origin of bipolar fields. It is presently believed that convection and differential rotation are responsible for the eruption of bipolar magnetic regions. The Sun's equator rotates faster than its polar caps and ought to wind up meridional field lines (like those that may join the Sun's poles), producing strong azimuthal fields below the sunspot belts. These strong fields would become evacuated to balance magnetic pressure against the external plasma pressure, and the lighter flux tube would rise to the surface (magnetic buoyancy) to form bipolar regions. The observed tilts of the bipolar regions are believed to result from cyclonal convection acting on the rising flux tubes. As increasing amounts of magnetic flux are packed more closely together, the very strong fields are believed to inhibit convection and produce the dark sunspots. *See* MAGNETISM; MAGNETOHYDRODYNAMICS; PLASMA (PHYSICS); SUN.

Neil R. Sheeley, Jr.

Bibliography. P. V. Foukal, *Solar Astrophysics*, 2d ed., Wiley, 2004; L. Golub and J. M. Pasachoff, *Nearest Star: The Surprising Science of our Sun*, Harvard University Press, 2001; E. J. Smith, Solar wind magnetic fields, in J. R. Jokipii (ed.), *Cosmic Winds and the Heliosphere*, University of Arizona Press, 1997; J. O. Stenflo, *Solar Magnetic Fields*, Kluwer Academic Publishers, 1994; M. Stix,

The Sun, 2d ed., Springer, 2004; H. Zirin, *Astrophysics of the Sun*, Cambridge University Press, 1988.

Solar neutrinos

Neutrinos produced in nuclear reactions inside the Sun. Neutrinos are produced as well in laboratory nuclear reactions. The first direct tests of how the Sun produces its luminosity (observed most conspicuously on Earth as sunlight) have been carried out by observing solar neutrinos. The results of these experiments confirm the theory of how the Sun shines and stars evolve. Moreover, the results show that neutrinos behave differently than is predicted by the standard model of particle physics.

Nuclear fusion in the Sun. The Sun shines because of fusion reactions similar to those envisioned for terrestrial fusion reactors. The basic solar process, called the proton-proton chain, is the fusion of four protons to form an alpha particle, two positrons (e^+), and two neutrinos (ν); that is, $4p \rightarrow \alpha + 2e^+ + 2\nu_e$ (see **table**). The rate for the initiating proton-proton (PP) reaction is predicted accurately ($\pm 1\%$) by the standard solar model. Unfortunately, the neutrinos from this reaction are below the energy thresholds for the first two experiments to detect solar neutrinos, with chlorine-37 (^{37}Cl) and with ultrapure water (H_2O). However, two experiments, one in Russia and one in Italy, which use gallium-71 (^{71}Ga), have successfully detected neutrinos from the PP reaction.

The proton-electron-proton (PEP) reaction, which is the same as the PP reaction except for having the electron in the initial state, is detectable in the ^{37}Cl experiment. The ratio of PEP to PP neutrinos is approximately independent of which solar model is used for the solar predictions. Two other reactions in the proton-proton chain are of special interest. The capture of electrons by ^{7}Be produces detectable neutrinos in the ^{37}Cl experiment. The ^{8}B decay is expected to be the main source of neutrinos for the ^{37}Cl experiment because of their relatively high maximum energy (14 MeV), although it is a rare reaction in the Sun. The ^{8}B neutrinos are the only substantial contributions to the experiments carried out with ultrapure water. There are also some less important neutrino-producing reactions from the carbon-nitrogen-oxygen (CNO) cycle, which will not be discussed here in detail since this cycle is believed to play a rather small role in the energy-production budget of the Sun. *See* CARBON-NITROGEN-OXYGEN CYCLES; NUCLEAR FUSION; NUCLEAR REACTION; PROTON-PROTON CHAIN.

Experimental tests. The first solar neutrino detector was based on the reaction $\nu_{\text{solar}} + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^-$, which is the inverse of the electron-capture decay of argon-37 (^{37}Ar). This reaction was chosen for the first experiment because of its unique combination of physical and chemical characteristics, which are favorable for building a large-scale solar neutrino detector. Neutrino capture to form ^{37}Ar in the ground state also has a relatively low energy threshold (0.81 MeV) and a high sensitivity, nuclear properties that are important for observing neutrinos from ^{7}Be, ^{13}N, and ^{15}O decay and the PEP reaction.

Like all solar neutrino experiments, the ^{37}Cl detector was built deep underground to avoid cosmic-ray interactions in the detector that simulate the effects of neutrinos. The final detector system consists of an approximately 400,000-liter (100,000-gallon) tank of perchloroethylene, a pair of pumps to circulate helium through the liquid, and a small building to house the extraction equipment, all in the deep Homestake gold mine in South Dakota, 1500 m (5000 ft) below the surface. Sets of experimental runs were carried out in the ^{37}Cl experiment during the 1970s and 1980s, and continued into the 1990s. They have shown that the ^{37}Ar production rate in the tank is 0.48 ± 0.04 ^{37}Ar atoms per day. Even though the tank is nearly a mile (1.6 km) underground, a small amount of ^{37}Ar is produced by cosmic rays. The background rate has been estimated to be of order 0.1 ^{37}Ar atom per day.

Subtracting the cosmic-ray background, a positive signal of (2.6 ± 0.2) solar neutrino units (SNU) is inferred; 1 SNU $= 10^{-36}$ capture per target particle per second.

Principal reactions of the proton-proton chain in the Sun

Number	Reaction	Percentage of solar terminations*	Maximum neutrino energy, MeV
1†	$p + p \rightarrow {}^2\text{H} + e^+ + \nu$ or	99.75	0.420
2‡	$p + e^- + p \rightarrow {}^2\text{H} + \nu$	0.3	1.44 (monoenergetic)
3	${}^2\text{H} + p \rightarrow {}^3\text{He} + \nu$	100	
4	${}^3\text{He} + {}^3\text{He} \rightarrow {}^4\text{He} + 2p$ or	85	
5	${}^3\text{He} + {}^4\text{He} \rightarrow {}^7\text{Be} + \nu$	15	
6	${}^7\text{Be} + e^- \rightarrow {}^7\text{Li} + \nu$	15	0.861 (90%), 0.383 (10%) (both monoenergetic)
7	${}^7\text{Li} + p \rightarrow 2\,{}^4\text{He}$ or	15	
8	${}^7\text{Be} + p \rightarrow {}^8\text{B} + \nu$	0.02	
9	${}^8\text{Be} \rightarrow {}^8\text{Be}^* + e^+ + \nu$	0.02	14
10	${}^8\text{Be}^* \rightarrow 2\,{}^4\text{He}$	0.02	

*Percentage of solar terminations of the proton-proton chain in which this reaction occurs.
†PP reaction.
‡PEP reaction.

The predicted capture rates for the ^{37}Cl experiment (in SNU) for the standard solar model is as follows: PP reaction: 0; ^{8}B beta decay: 5.8; PEP reaction: 0.2; ^{7}Be electron capture: 1.2; ^{13}N decay: 0.1; and ^{15}O decay: 0.3. The total theoretical prediction is about 7.5 SNU. Many investigations have been undertaken of the best values to use for various input parameters for the calculation of the standard solar model. The calculated event rate is relatively insensitive to the input parameters, although the total predicted rate may well differ from 7.5 SNU by 1 SNU.

Solar neutrinos were also observed with the aid of electrons that the neutrinos scatter in a 700-metric-ton (770-short-ton) detector of pure water located in the Kamioka metal mine about 300 km (200 mi) west of Tokyo in the Japanese Alps. These results provided direct evidence that the observed neutrinos originate in the Sun; the electrons are scattered by neutrinos in the forward direction between the Earth and the Sun. Moreover, the neutrino events are registered at the exact time they are detected, making possible a sensitive search for possible time dependencies. The measurement of the Kamiokande collaboration yielded the result given by Eq. (1). The

$$\frac{\text{Observed}}{\text{Predicted}} = 0.55 \pm 0.08 \qquad (1)$$

indicated error is a 1-standard-deviation combined statistical and systematic uncertainty. The predicted event rate in the equation stands for the best estimate of the number of events calculated to occur according to the standard model of the Sun and the standard model of how neutrinos behave. The Kamioka result is sensitive only to high-energy neutrinos from ^{8}B decay.

The Super-Kamiokande detector is a much larger version of the original Kamiokande detector and is located in the same mine. The active detector in the Super-Kamiokande experiment contains 32,000 metric tons (35,000 short tons) of pure water and 11,200 specially designed tubes for observing electrons scattered by neutrinos. This more powerful superdetector finds a rate given by Eq. (2). The new rate is in

$$\frac{\text{Observed}_{\text{superK}}}{\text{Predicted}} = 0.46 \pm 0.01 \qquad (2)$$

good agreement with the previous measurement but is more precise.

Since the Japanese water experiments (Kamiokande and Super-Kamiokande) and the chlorine experiment are both primarily sensitive to the same ^{8}B neutrinos, it was difficult to understand why the discrepancy between observation and calculation is greater for the chlorine than for the water experiments. The favored interpretation is that neutrinos do not behave as predicted by the standard model of particle physics; neutrinos change their type as they travel from the center of the Sun to detectors on Earth. The chlorine detector is sensitive only to neutrinos associated with electrons, electron-type neutrinos, while the water detectors have some sensitivity to all types of neutrinos. This interpretation has been strengthened by the fact that the Super-Kamiokande experiment finds strong evidence that neutrinos produced by cosmic rays in the atmosphere change their type when they travel distances comparable to the Earth's radius.

Experiments have also been performed with gallium detectors located underneath a mountain in the Caucus region of Russia [60 metric tons (66 short tons) of gallium, SAGE experiment] and in an underground laboratory about an hour's drive north of Rome [30 metric tons (33 short tons) of gallium, GALLEX/GNO experiment]. These experiments are similar to the chlorine experiment. The two gallium experiments and the chlorine experiment all use radiochemical techniques to extract a small number of radioactive nuclei produced by neutrinos from huge tanks containing enormous numbers of atoms of chlorine or gallium. Unlike the Kamiokande and Super-Kamiokande experiments, the gallium and chlorine detectors do not determine specific energies for the neutrinos they detect. Instead, the radiochemical gallium and chlorine detectors specify only a minimum energy: 0.23 MeV for gallium and 0.81 MeV for chlorine.

The results from the gallium experiments reinforce the conclusion that new physics—adjustments or changes to the current understanding of the standard model of particle physics—is required. The observed rate is 69 ± 4 SNU, which is much less than the standard-model predicted rate of about 130 SNU. Almost all of the observed rate in the gallium experiments could be attributed to the fundamental low-energy neutrinos from the PP and PEP reactions. The rate expected from these neutrinos is largely independent of astronomical uncertainties. If the PP and PEP neutrinos are being detected at the expected rate, there is practically no room in the experimental results to accommodate the neutrinos from ^{7}Be, whose number is also believed to be reliably calculated with the solar models, and the ^{8}B neutrinos, which are actually observed in the Kamiokande and Super-Kamiokande experiments.

Possible explanations. Many explanations were advanced for the discrepancy between the observed and the predicted event rates in solar neutrino experiments. These explanations can be divided into three classes: (1) the standard solar model must be significantly modified; (2) something is seriously wrong with the experiments; (3) the standard model of how neutrinos behave must be significantly modified.

No one has succeeded in modifying a solar model so that it is consistent with all of the experimental data on solar neutrinos, as well as the astronomical data on the Sun and the laboratory measurements of the parameters used in constructing the solar models. Moreover, precise measurements of the thousands of frequencies with which the Sun pulsates on its surface (characteristic period of order 5 min) have confirmed to an accuracy of 0.1% the predictions of the standard solar model for these pulsation frequencies. This agreement convinced researchers that the standard solar model is an accurate description of the Sun.

All of the solar neutrino experiments have been examined carefully by many different researchers. A variety of checks have been made to test whether there was a significant error or a large uncertainty in one of the experiments that might explain the difference between prediction and observation. No significant previously unknown errors or uncertainties have been found. Moreover, the gallium experiments have been tested in the most direct fashion possible. Intense laboratory sources of neutrinos have been placed near the gallium detectors, and the expected number of events has been observed from these artificial sources. The consensus among scientists in the field is that the solar neutrino experiments are yielding a valid although surprising result.

The only remaining possibility is that the theory of how the neutrino behaves must be changed. Indeed, in March 2000 the results of a decisive experiment were published that showed unequivocally that solar neutrinos change their type on their way to Earth from the center of the Sun.

SNO experiment and combined results. An experiment called SNO (Sudbury Neutrino Observatory) uses 1000 metric tons (1100 short tons) of heavy water in a nickel mine in Sudbury, Ontario, Canada. This experiment began operating in the spring of 1999. It measures both the number of neutrinos of the electron type (associated with electrons) and the number of neutrinos of all types (associated with electrons, muons, and tau particles). In March 2000, the SNO collaboration combined their measurements of just electron-type neutrinos with the measurements from the Super-Kamiokande ordinary-water experiment (some sensitivity to neutrinos of all types) to obtain the total number of solar neutrinos from ^{8}B that reach the Earth. The SNO and Super-Kamiokande result announced in 2001 is given by Eq. (3), where the error shown reflects only

$$\frac{\phi}{\phi_{\text{predicted}}} = 1.08 \pm 0.20 \qquad (3)$$

the measurement uncertainty. Unlike the previous measurement made by Super-Kamiokande alone (for which the ratio of measured to predicted neutrino rate was 0.55), the new SNO plus Super-Kamiokande determination includes all of the ^{8}B solar neutrinos with equal sensitivity. The Super-Kamiokande experiment is mostly sensitive to electron-type neutrinos. By a series of improvements in the experimental measurements and the use of a more powerful analysis technique, the comparison between theory and measurement was refined, with the result, announced in September 2003, given by Eq. (4).

$$\frac{\phi}{\phi_{\text{predicted}}} = 1.01 \pm 0.04 \qquad (4)$$

An empirical value for the number of PP neutrinos has also been extracted from an analysis that includes all the solar neutrino experiments and relevant experiments with reactor neutrinos. The measured number of PP neutrinos is given by Eq. (5).

$$\frac{\phi\,(PP)}{\phi(PP)_{\text{predicted}}} = 1.02 \pm 0.02 \qquad (5)$$

All of the results from solar neutrino experiments are consistent with the conclusion that the standard solar model predicts accurately the number of neutrinos of different energies that are emitted by the Sun but that some of the neutrinos change their type on the way from the center of the Sun to the detectors on Earth. *See* NEUTRINO; STANDARD MODEL; STELLAR EVOLUTION; SUN.

John N. Bahcall

Bibliography. N. Abdurashitov et al., Solar neutrino flux measurements by the Soviet-American Gallium Experiment (SAGE) for half the 22-year solar cycle, *J. Exp. Theor. Phys.*, 95:181–193, 2002; Q. R. Ahmad et al., Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.*, 87:071301-1 to -6, 2001; M. Altmann et al., GNO solar neutrino observations: Results for GNO I, *Phys. Lett. B*, 490:16–26, 2000; J. N. Bahcall and C. Peña-Garay, A road map to solar neutrino fluxes, neutrino oscillation parameters, and tests for new physics, *J. High Energy Phys.*, 11004-0 to 11004-47, 2003; B. T. Cleveland et al., Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *Astrophys. J.*, 496:505–526, 1998; S. Fukuda et al., Solar ^{8}B and hep neutrino measurements from 1258 days of Super-Kamiokande data, *Phys. Rev. Lett.*, 86:5651–5655, 2001; Y. Fukuda, Final solar neutrino results at Kamiokande, *Phys. Rev. Lett.*, 77:1683–1686, 1996.

Solar radiation

The electromagnetic radiation and particles (electrons, protons, alpha particles, and rarer heavy atomic nuclei) emitted by the Sun. The electromagnetic radiation covers a wavelength range from x-rays to radio waves, that is, from about 0.01 nanometer to 30 km. The annual mean irradiance at Earth, integrated over the whole spectrum, amounts to 1365 $W \cdot m^{-2}$, and 99% of its energy is carried by radiation with wavelengths between 278 and 4600 nm, with the maximum at 472 nm. The division into two energetically equal parts is at a wavelength of 731 nm. Although the total solar irradiance is often called the solar constant, it varies with time, for example, with the solar activity cycle with a period of about 11 years. These variations have been monitored since 1978 by radiometers on space platforms and amount to about 0.1%, with the radiation higher during the maximum of solar activity. There are also shorter-term variations such as decreases due to the passage of dark sunspots or increases due to faculae. *See* ELECTROMAGNETIC RADIATION; SOLAR CONSTANT.

The Sun also emits a continuous stream of particles, the solar wind, which originates in coronal holes and the upper corona. The density and speed of solar wind particles, and thus the solar wind's

influence on the Earth environment, such as the ionosphere, vary with the solar cycle. This wind also carries magnetic fields from the surface of the Sun into interplanetary space, and these fields in turn shield the Earth from the highly energetic cosmic rays coming from outside the solar system. Since these fields are stronger during periods of high solar activity, the shielding becomes more efficient and less cosmic rays reach the Earth at these times. On the other hand, high activity means also increased occurrence of explosive events on the Sun: the solar flares and coronal mass ejections. Solar flares are produced by the most powerful explosions, releasing energies of up to 10^{25} joules in 100–1000 s and high-speed electrons that emit intense radiation at radio and x-ray wavelengths. They also produce nuclear reactions in the solar atmosphere with the emission of gamma rays and of neutrons that move nearly at the speed of light. Coronal mass ejections expand away from the Sun at speeds of hundreds of kilometers per second, becoming larger than the Sun and removing up to 5×10^{13} kg of coronal material. Both events are believed to be ignited by the reconnection of magnetic fields, and the emitted particles are much more energetic and numerous than those of the solar wind. If these particles reach the Earth, they give rise to the aurora at high latitudes, and they can damage satellites, endanger humans in space, and on the Earth disturb telecommunications and even disrupt power systems. Such influences on the Earth environment are due to "space weather." Its prediction, and thus the continuous observation of the Sun, is of major importance for the functioning of civilization. *See* AURORA; COSMIC RAYS; IONOSPHERE; MAGNETOSPHERE; SOLAR WIND; SUN.

Claus Fröhlich

Bibliography. J. Burch, The fury of space storms, *Sci. Amer.*, 284(4):72–81, April 2001; L. Golub and J. M. Pasachoff, *Nearest Star: The Surprising Story of Our Sun*, Harvard University Press, 2001; K. R. Lang, *SOHO* reveals the secrets of the Sun, *Sci. Amer.*, 276(3):32–47, March 1997; C. J. Schrijver and A. M. Title, Today's science of the Sun, *Sky Telesc.*, 101(2):35–40, February 2001, and 101(3):35–39, March 2001.

Solar system

The Sun and the bodies moving in orbit around it. The most massive body in the solar system is the Sun, a typical single star that is itself in orbit about the center of the Milky Way Galaxy. Nearly all of the other bodies in the solar system—the terrestrial planets, outer planets, asteroids, and comets—revolve on orbits about the Sun. Various types of satellites revolve around the planets; in addition, the giant planets have orbiting rings. The orbits for the planets appear to be fairly stable over long time periods and hence have undergone little change since the formation of the solar system. It is thought that some 4.56×10^9 years ago a rotating cloud of gas and dust collapsed to form a flattened disk (the solar nebula) in which the Sun and other bodies formed. The bulk of the gas in the solar nebula moved inward to form the Sun, while the remaining gas and dust are thought to have formed all the other solar system bodies by accumulation proceeding through collisions of intermediate-sized bodies called planetesimals. Planetary systems are believed to exist around many other stars in the Milky Way Galaxy. Solid evidence for the existence of Jupiter-mass planets around nearby solarlike stars now exists. *See* PLANET.

Composition. The Sun is a gaseous sphere with a radius of about 7×10^5 km (4×10^5 mi), composed primarily of hydrogen and helium and small amounts of the other elements. The Sun is just one of about 2×10^{11} stars in the Milky Way Galaxy. The solar system lies in the disk of the Galaxy and moves around the galactic center about once every 2×10^8 years on a circular orbit with a radius close to 3×10^4 light-years (about 3×10^{17} km or 2×10^{17} mi). The Sun is somewhat peculiar in that it is a single star, unaccompanied by another star; most stars in the disk of the Galaxy are in double or multiple systems where two or more stars orbit about their common center of mass. The Sun's mass of 2×10^{30} kg (4.4×10^{30} lb), however, is quite typical; stellar masses range from about 0.08 to about 60 times the Sun's mass, with the great majority of stars being similar in mass (and hence in other characteristics) to the Sun. *See* BINARY STAR; MILKY WAY GALAXY; STAR; SUN.

The terrestrial planets (Mercury, Venus, Earth, and Mars) are the closest to the Sun, with orbital radii ranging from 0.39 AU for Mercury to 1.5 AU for Mars. One astronomical unit (AU), the distance from Earth to Sun, equals 1.5×10^8 km or 9.3×10^7 mi. The terrestrial planets are composed primarily of silicate rock (mantles) and iron (cores). The Earth is the largest terrestrial planet, with an equatorial radius of 6378 km (3963 mi) and a mass of 5.972×10^{24} kg (1.317×10^{25} lb); Mercury is the smallest, with a mass of 0.055 times that of Earth. *See* EARTH; MARS; MERCURY (PLANET); PLANETARY PHYSICS; VENUS.

The outer planet region begins with Jupiter at a distance of 5.2 AU from the Sun. The outer planets are subdivided into the gas giant or Jovian planets (Jupiter and Saturn), the ice giant planets (Uranus and Neptune), and Pluto. By far the largest planet is Jupiter, with a mass 318 times that of the Earth, while the other giant planets are more massive by a factor of 15 or more than Earth. Jupiter and Saturn are composed primarily of hydrogen and helium gas, like the Sun, but with rock and ices, such as frozen water, methane, and ammonia, concentrated in their cores. Uranus and Neptune also have rock and ice cores surrounded by envelopes with smaller amounts of hydrogen and helium. Pluto, slightly smaller than the Earth's Moon, is probably composed primarily of rock and ice. *See* JUPITER; NEPTUNE; PLUTO; SATURN; URANUS.

The region between Mars and Jupiter is populated by a large number of rocky bodies called asteroids. The asteroids are smaller than the terrestrial planets, with most known asteroids being about 1 km (0.6 mi) in radius, though a few have radii of

hundreds of kilometers. Some asteroids have orbits that take them within the orbits of Earth and the other terrestrial planets, leading to the possibility of impacts with these planets; asteroids in the main belt beyond Mars are occasionally perturbed into these Earth-crossing orbits. Small fragments of asteroids (or comets) that impact the Earth first appear as meteors in the sky; any meteoric material that survives the passage through the Earth's atmosphere and reaches the surface is called a meteorite. The Allende meteorite, which fell in Mexico in 1969, contains inclusions that have been radiometrically dated at 4.56×10^9 years; this age, the oldest found so far, is presumably the age of the solar system. Along with the lunar rocks returned by the Apollo missions, Allende and other meteorites are samples of other bodies in the solar system. In addition, primitive components of meteorites such as Allende give important information about conditions in the very early solar system. A few meteorites are thought to have come from Mars or the Moon. *See* ASTEROID; METEOR; METEORITE.

Comets are icy bodies (so-called dirty snowballs) with diameters on the order of 10 km (6 mi). In contrast to the orbits of most planets, cometary orbits often are highly elliptical and have large inclinations that take them far above and below the plane where the planets orbit. The region well beyond Pluto's orbit is populated with a very large number (perhaps 10^{12}) of comets, out to a limiting distance of about 10^5 AU, at which point external forces (due to nearby stars, passing molecular clouds, and the Milky Way Galaxy itself) disturb any body trying to stay in a stable orbit around the Sun. The distribution of comets within this huge volume, the Oort Cloud, is uncertain; there may be an inner cloud extending outward to about 2×10^4 AU, and an outer cloud beyond that. In addition, comets have been detected orbiting in the plane of the solar system at distances of 30 to 55 AU; this flattened distribution is called the Edgeworth-Kuiper Belt, named for the scientists who predicted its existence in 1949 and 1951. The largest objects in the Edgeworth-Kuiper Belt can now be observed from Earth, with diameters in the range from 100 to 1000 km (60 to 600 mi); Pluto is one the largest bodies in this family of icy objects. Comets that enter the inner solar system are eventually evaporated by solar radiation, leaving behind interplanetary dust particles and perhaps rocky cores similar to certain asteroids. *See* COMET; KUIPER BELT.

Satellites orbit around all of the planets except Mercury and Venus. The Earth and its rocky Moon are unusual in that the Moon is about one-fourth the size of the Earth; nearly all other satellites are much smaller than their planet, with the exception of Pluto and its satellite Charon, which is half the size of Pluto. While Mars has only two tiny, rocky satellites, the Voyager missions, followed by ground-based observations, have shown that the giant planets all have large systems of satellites (Jupiter has at least 60), ranging in size from 0.5 to 2700 km (0.3 to 1700 mi) in radius and composed of varying amounts of rock and ice. Each giant planet also has a number of rings of particles whose orbits are interspersed with and controlled by the planet's innermost satellites. *See* MOON; SATELLITE (ASTRONOMY).

Origin. A number of imaginative theories of the origin of the solar system have been presented in the last several hundred years, but nearly everyone presently working on solar system cosmogony agrees on one fundamental concept, advanced in 1796 by P. S. de Laplace. This concept, the nebular hypothesis, holds that the Sun and the rest of the bodies in the solar system formed from the same rotating, flattened cloud of gas and dust, now called the solar nebula. The nebular hypothesis at once explains the gross orbital properties of the solar system: all planets orbit (and most rotate) in the same sense as the Sun rotates, with their nearly circular orbits being confined largely to a single plane almost perpendicular to the Sun's rotation axis.

While the grand experiment that led to the formation of the solar system can never be repeated, present-day regions of star formation in the Milky Way Galaxy can be observed to gain insight into the processes involved in the formation of stars similar in mass to the Sun. Assuming that present-day star formation proceeds much as it did 4.56×10^9 years ago, this approach should yield many constraints on the general process. These observations confirm the stellar implications of Laplace's nebular hypothesis: very young stars (protostars) are indeed found embedded in dense clouds of gas and dust that often show evidence for flattening and rotation. While understanding of solar system formation is almost certain to undergo further evolution, the following summarizes the basic framework in which cosmogonists are working.

Collapse phase. The solar nebula was produced by the collapse of a dense interstellar cloud (**illus.** *a*) composed primarily of molecular hydrogen (about 77% by mass) and helium gas (about 21%), with about 2% of the cloud mass being in the form of the other elements, mostly frozen into dust grains. Elements heavier than hydrogen and helium were formed by nucleosynthesis in previous generations of stars and then ejected into interstellar clouds through events such as supernovae. *See* NUCLEOSYNTHESIS; SUPERNOVA.

Radio telescopes have shown that dense interstellar clouds exist with masses comparable to that of the Sun, containing about 10^4–10^6 molecules/cm^3, and with quite low temperatures, about 10 K (-263°C or -441°F). Because of their low temperatures, gas pressure generally is not sufficient to support these clouds against their own self-gravity, which tries to pull the cloud into a smaller configuration. Many clouds appear to be supported primarily by magnetic fields; however, magnetic field support requires a continual contraction of the largely neutral cloud past the field lines, and eventually magnetic forces can no longer support the increasingly dense cloud. At this point, dense clouds enter the collapse phase, where supersonic inward motions develop that lead to the formation of a stellar-sized core at the center of

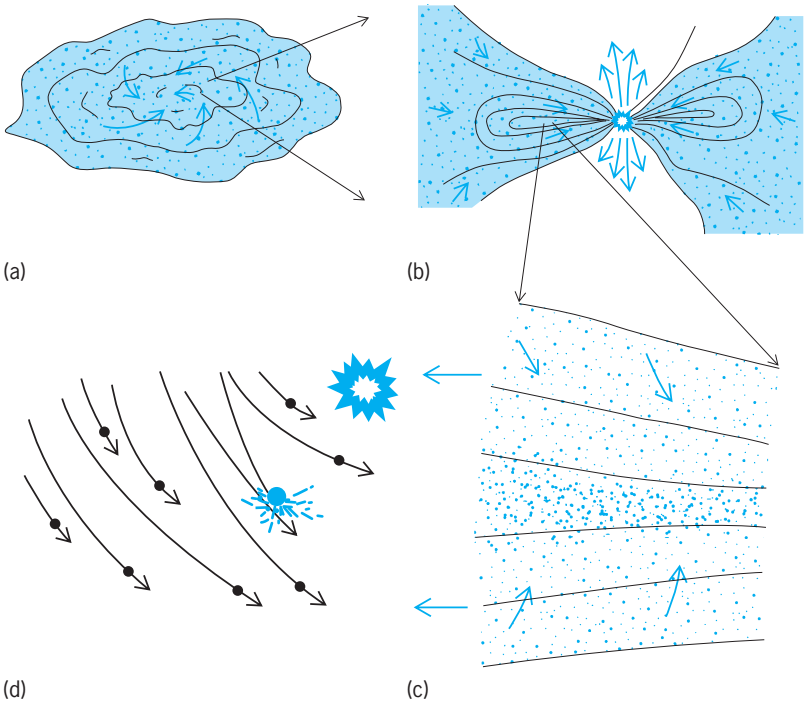
the cloud in about 10^5–10^6 years. Some of the gas in the cloud continues to fall onto this protostellar core, where it is suddenly stopped and its kinetic energy is converted into thermal energy. This heat leaves the cloud as infrared radiation; infrared telescopes have shown that many dense cloud cores contain deeply embedded young stars. *See* INFRARED ASTRONOMY; INTERSTELLAR MATTER; RADIO ASTRONOMY; STELLAR EVOLUTION.

Solar nebula. In a rotating cloud, not all of the infalling gas and dust falls directly onto the central protostar, because of the conservation of angular momentum. Instead, a disklike solar nebula forms, where the inward pull of gravity is expended in maintaining the rotational motion of the nebula. If too much angular momentum is present in the initial cloud, the collapse process is likely to result in the formation of a double or multiple star system instead of a single star and a planetary system; if extremely little angular momentum is present, only a single star forms.

Because the Sun contains 99.9% of the mass of the solar system but only about 2% of the angular momentum, the planetary system has considerably more angular momentum per unit mass than the Sun. This discrepancy is more than can be accounted for by forming the Sun out of the lowest-angular-momentum portions of a dense cloud, or by allowing for the fact that the Sun has lost substantial angular momentum during the age of the solar system through magnetic braking associated with the solar wind. *See* ANGULAR MOMENTUM; SOLAR WIND.

The solution to this angular momentum problem appears to lie in substantial evolution of the solar nebula after its formation. In order to have enough angular momentum to lead to planetary system formation, much of the mass of the newly formed solar nebula must reside in the disk rather than in the early Sun. The disk must then evolve in such a way as to transfer mass inward to feed the growing Sun, while transporting outward the excess angular momentum undesired by the Sun but required for the planets. While this sort of evolution may appear to be contrived if not miraculous, it is actually to be expected on very general grounds for any viscous disk that is undergoing a loss of energy, as the solar nebula will, through radiation to space. Several different physical mechanisms have been identified that potentially are capable of driving solar nebula evolution: turbulence caused by convective instability or velocity shear, gravitational torques between a bar-shaped protosun and spiral arms in the nebula, and magnetic torques associated with remanent magnetic fields or with a nebula magnetic field generated by the dynamo mechanism. These mechanisms are likely to lead to evolution of the bulk of the nebula gas inward onto the protosun (illus. *b*), on a time scale on the order of 10^5–10^7 years. Magnetic fields are the currently favored driver of nebula evolution.

Planetesimal formation. The portion of the nebula that is to form the planets must decouple from the gaseous nebula to avoid being swallowed by the Sun. This occurs by the process of coagulation of dust



Schematic diagram of four phases of the formation of the solar system. (*a*) A dense, rotating interstellar cloud collapsed to form the solar nebula, a flattened cloud of gas and dust. (*b*) Most of the gas in the solar nebula flowed onto the growing protosun, whose energetic wind first flowed outward in a bipolar pattern and later removed the last vestiges of the nebula. (*c*) Dust grains sedimented to the nebula midplane and coagulated into kilometer-sized planetesimals. (*d*) Gravitational forces between the planetesimals resulted in collisions and their accumulation into planetary-sized bodies.

grains through mutual collisions; when solid bodies become large enough (roughly kilometer-sized), they will no longer be tied to the nebula through brownian motion (as is the case with dust grains) or gas drag (as happens with smaller bodies).

The interstellar dust grains that eventually form the solid bodies in the solar system start out with sizes on the order of 0.1 micrometer. These minuscule particles must stick together in order to begin the planet formation process; sticking can be caused by van der Waals forces. Coagulation is enhanced by the increase in the spatial density of dust grains, which occurs as the dust grains sediment down through the gaseous nebula to form a dense subdisk composed largely of solid particles (illus. *c*). Both dust grain sedimentation and coagulation are inhibited by the vigorous stirring expected in a highly turbulent nebula (the latter because grains may hit each other at relative velocities that are high enough to produce fragmentation rather than coagulation), but once the turbulence dies down, sedimentation and coagulation can produce centimeter- to meter-sized bodies in about 10^3–10^4 years. *See* INTERMOLECULAR FORCES.

The next phase of growth has long been thought to involve a collective gravitational instability of the dust subdisk that would rapidly produce bodies of kilometer size or larger, termed planetesimals. However, this instability may be prevented by turbulence induced by the difference in orbital speed between the gaseous nebula and the dust subdisk. (The gas orbits the Sun somewhat more slowly than the solid

bodies because the gaseous disk is partially supported against the Sun's gravity by gas pressure.) If the gravitational instability is prevented, then growth to planetesimal size may occur through further collisional coagulation; gas drag preferentially slows down smaller bodies, producing relative motions between larger and smaller bodies that can lead to collisions. In this case, growth to kilometer-sized planetesimals is expected to occur in about 10^4 years. There is currently much debate as to whether or not a dust disk instability was actually involved in growth to this size.

Planetary accumulation. About 10^{12} kilometer–sized planetesimals are needed to form just the terrestrial planets; significantly greater numbers of similarly sized bodies would be needed to form the giant planets. These plantesimals are already roughly the size of many asteroids and comets, suggesting that many of these bodies are simply leftovers from intermediate phases of the planet formation process.

Planetesimals are massive enough for gravitational forces to determine their collision probabilities; self-gravity is also needed to prevent debris from escaping following collisions. The subsequent growth of the planetesimals through gravitational accumulation (illus. *d*) is in two distinct phases. In the first phase, planetesimals grew by accumulation of other planetesimals at essentially the same distance from the Sun. Once the nearby planetesimals were all swept up, this phase ended. The first phase lasted about 10^4 years in the inner solar system and produced planetesimals about 500 km (300 mi) in size. This phase is likely to have been characterized by the runaway growth of relatively few bodies, because once one body becomes larger than its neighbors (as must happen after any collision occurs), its increased gravitational pull increases the chances that another body will impact it, and so on.

In the second phase, accumulation requires bodies at significantly different distances from the Sun to collide. This can happen only if the planetesimal orbits become highly elliptical and hence intersecting. Gravitational forces between the orbiting planetesimals can produce elliptical orbits, but only over relatively long time periods; about 10^7–10^8 years is required for accumulation to proceed by this means in the inner solar system. The second phase is currently thought to have proceeded differently from the first phase, with collisions occurring primarily between more or less equal-sized bodies, rather than the highly unequal-sized collisions that occur in a runaway accumulation process. The second phase may then have involved violent collisions between planetary-sized bodies, a spectacular finale to the entire planet formation process. A glancing collision between a Mars-sized and an Earth-sized body appears to be the best explanation for the formation of the Earth-Moon system; debris from the giant impact would end up in orbit around the Earth and later form the Moon.

Astronomical observations of young stars similar to the Sun imply that the gaseous portion of the solar nebula was removed some time between 10^5 and 10^7 years after the Sun began to form. While most of the gas presumably was added to the Sun through nebula evolution, the residual nebula gas and dust were probably removed from the solar system by the early solar wind, which had a mass loss rate roughly 10^6 times larger than at present (illus. *b*). The final formation of the terrestrial planets, lasting 10^7–10^8 years, then occurred in the absence of appreciable nebula gas, which explains the absence of significant hydrogen and helium in the inner solar system. However, formation of the giant planets must have occurred prior to the loss of the nebula gas; otherwise their rock and ice cores would not have been able to capture the gas needed to account for their present compositions. Formation of the giant planets within about 10^6 years appears to require that the second phase of planetesimal accumulation proceeded through runaway accretion all the way to bodies about 10 earth masses in size in the outer solar system. The satellite systems of the giant planets were largely formed in minisolar nebulae orbiting around each protoplanet, through processes that are similar to those described for planet formation.

Forming gas giant planets by the two-step process requires about 10^7 years for a 10-earth-mass core to form and then accrete a massive gaseous envelope, which may be longer than the life of typical gaseous disks. The alternative means for forming the gas giant planets is much more rapid, requiring only about 10^3 years for a gravitational instability of the gaseous nebula to produce a massive clump of gas and dust. The dust will settle to form a core at the center of the clump on a similar time scale. If gas giant planets form by the latter mechanism, then even the youngest stars will show evidence for Jupiter-mass companions; while if the two-step mechanism predominates, most young stars will not be old enough to have such companions. The formation of ice giant planets could occur in the disk instability mechanism if the solar nebula was born in a region of high-mass stars. Their radiation would photoevaporate the disk gas beyond Saturn's orbit as well as the gaseous envelopes of the protoplanets orbiting there.

Extrasolar planetary systems. Confirmation of the implications of the nebular hypothesis for planetary formation requires the detection of planetary systems orbiting around other stars; if the basic theory of solar system formation is correct, planetary systems should not be rare in the Milky Way Galaxy.

A new era in the study of planetary origins emerged in 1995, when Michel Mayor and Didier Queloz found evidence for a Jupiter-mass planet in orbit around the Sun-like star 51 Pegasi. Since then, over 130 other Jupiter-mass planets have been discovered, mostly by Geoffrey Marcy and Paul Butler, with orbital distances ranging from a surprising 0.04 AU to about 6 AU; the latter was to be expected based on the solar system, but not the former. These discoveries appear to show that some planets can migrate inward toward their stars following their formation. Other surprises undoubtedly await as additional extrasolar planets are discovered and characterized.

Alan P. Boss

Bibliography. A. P. Boss, *The Race to Find New Solar Systems*, Wiley, 1998; J. J. Lissauer, Planet formation, *Annu. Rev. Astron. Astrophys.*, 31:129–174, 1993; V. Mannings, A. P. Boss, and S. S. Russell (eds.), *Protostars and Planets IV*, 2000; V. S. Safronov, *Evolution of the Protoplanetary Nebula and Formation of the Earth and the Planets*, 1969, transl., NASA TTF-677, 1972; S. R. Taylor, *Solar System Evolution*, 2d ed., 2001; G. W. Wetherill, Formation of the Earth, *Annu. Rev. Earth Planet. Sci.*, 18:205–256, 1990.

Solar wind

A flow of completely ionized material away from the Sun. As this plasma moves through space, it creates a bubble known as the heliosphere. The processes that heat and accelerate the solar wind are not well understood at present. At heights of a few solar radii above the Sun's visible surface, the flow becomes supersonic, reaching speeds of 350–750 km/s (220–470 mi/s) in interplanetary space. The solar wind finally terminates at a distance of perhaps 150 astronomical units (1 AU is the distance from the Sun to Earth), where it interacts with the local interstellar medium.

This flow can be conveniently categorized as either ambient or time-dependent. The combination of solar rotation together with the ejection of material of different speeds from the Sun leads to a dynamical interaction between neighboring parcels of plasma and the formation of interaction regions and shocks that bound them. This forms the basis of the ambient solar wind. These structures appear stationary in a frame of reference rotating with the Sun and are thus known as corotating interaction regions (CIRs). Coronal mass ejections (CMEs) represent the most spectacular form of time-dependent phenomena.

During the course of the 11-year solar cycle, the structure of the solar wind changes at a fundamental level. At the minimum of the solar activity cycle a simple flow pattern exists consisting of (1) a slow wind about the magnetic equator (tilted or warped about the rotational equator), and (2) a fast wind emanating from large polar coronal holes. In contrast, at solar maximum the flow is considerably more complex and time-dependent. The rate, launch site, and properties of coronal mass ejections also vary with the phase of the solar cycle. The interaction of the solar wind with the Earth's geomagnetic environment represents an important and practical application of space physics known as space weather, which seeks to describe the conditions in space that affect both the Earth and its technological systems.

The study of the solar wind is a relatively new topic. As recently as the 1950s, its mere existence was speculation. Now, in part through the successful launch of a number of innovative spacecraft, considerably more is known. The first clue that the Sun was emitting a continuous wind can be traced back to observations of comet tails, which were found to always point away from the Sun no matter whether the comet was moving toward or away from the Sun. It turns out that comets often have two tails: a dust tail, produced by radiation pressure (that is, sunlight), and an ion tail. Differences in the properties of these two tails led L. Biermann in 1957 to propose the presence of a continuous stream of particles propagating away from the Sun that pushed the ions making up the ion tail. The following year, E. N. Parker derived a theory for the solar corona that required there to be a flow of plasma away from it. (Previously it had been thought that the extended atmosphere of the Sun was static and gravitationally bound to the Sun.) He named this flow the solar wind. In 1959, the Soviet spacecraft *Lunik 2* made the first direct measurements of the particles making up the solar wind. In the intervening years a number of interplanetary spacecraft have uncovered a broad spectrum of phenomena. *See* COMET; SPACE PROBE.

Coronal heating and solar wind acceleration. The visible surface of the Sun is known as the photosphere. Beyond this lies the chromosphere, and above this lies the corona. The corona extends many solar radii away from the Sun, eventually becoming the solar wind.

The temperature of the photosphere is a modest 10,000 K, while the solar corona is more than 10^6 K. Since the only source of energy that can heat the outer layers of the Sun comes from below, a major unsolved problem in space physics has been: How can the corona be hotter than the photosphere? Thermal conduction cannot transfer heat from something that is cold to something that is hot. Radiation can be ruled out too because the solar corona is transparent. What is known is that the continuous, agitating motion of the convective zone just below the photosphere provides the energy to heat the corona. However, how it is transported into the corona is still not well understood. It was once thought that magnetic or acoustic waves generated by the convective zone motions propagate up into the corona and dissipate their energy there. Unfortunately, it appears that acoustic waves are reflected back to the Sun at low altitudes and magnetic waves do not appear to dissipate sufficient energy into the corona.

A more promising explanation for coronal heating comes from a process known as magnetic reconnection. Coronal heating is strongest in regions of intense magnetic fields. In these so-called active regions, magnetic fields are constantly merging with one another and liberating large amounts of energy. This process manifests itself by the presence of jets, x-ray bright points, and many low-intensity flares. **Figure 1** illustrates how fountains of plasma lead to the formation of giant loops hundreds of thousands of kilometers above the surface of the Sun. The plasma on these loops can reach temperatures of several million kelvin and may be the energy source that heats the solar corona. *See* SOLAR CORONA.

Ambient solar wind. To a first approximation, two types of ambient solar wind can be distinguished: slow and fast (see **table**). The slow flow, which is associated with coronal streamers (features of the white light corona that have an archlike base

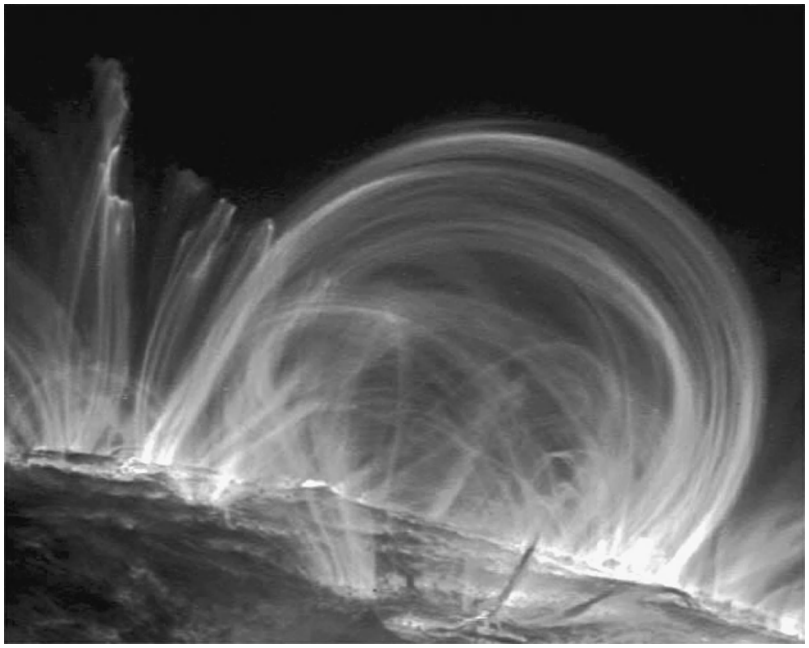


Fig. 1. Hot, dense plasma, showing the location of magnetic loops in the solar corona. This image was taken by the *Transition Region and Coronal Explorer (TRACE)* spacecraft, launched in 1998. (*Courtesy of the TRACE Consortium, Stanford-Lockheed Institute for Space Research, and NASA*)

and rays extending away from the Sun out to about one solar radius), travels away from the Sun at approximately 350 km/s (220 mi/s). It is denser and colder than the fast wind. The fast wind travels at more than twice the speed of the slow wind. It is also considerably steadier and emanates from coronal holes. The helium-to-proton abundance (the ratio of the number density of doubly ionized helium to protons) acts as a unique tracer of the origin of the solar wind in which the spacecraft is immersed. More sophisticated composition instruments on board *Ulysses* and *ACE* spacecraft now make it possible to measure the ratios of other minor species. Scientists are beginning to use these measurements to uncover the conditions and processes in the corona where the wind is accelerated as well as identifying the source regions of the solar wind in the corona.

This basic two-state flow pattern of the solar wind is illustrated in **Fig. 2**, which shows the speed measured at *Ulysses* (between 1.3 and 5.4 AU) as it completed its first orbit around the Sun. These measurements were taken during predominantly solar minimum conditions. At low latitudes, near the ecliptic, *Ulysses* observed slower, more variable solar wind. Comparison with the white-light image of the Sun shows that this flow is associated with coronal streamers. At higher latitudes, on the other hand, the flow is faster and significantly less variable. This flow is associated with the coronal holes (seen as dark regions in the ultraviolet disk image). The speed has been coded according to the magnetic polarity of the interplanetary magnetic field and shows that during this time interval the magnetic field was largely outward in the northern hemisphere and inward in the southern hemisphere, reflecting the dipole character of the solar field at this time.

Interplanetary magnetic field. The electrical conductivity of the solar wind plasma is sufficiently high that the magnetic field is effectively frozen to it. Thus, as the solar wind propagates away from the Sun, it drags the solar field with it. Because the field lines remain attached to the Sun at one end, solar rotation causes them to become wrapped into spirals. An initially radial field line at the Sun will have a pitch of approximately 45° at 1 AU and will lie almost perpendicular to the radial direction by 10 AU. Parker first provided a mathematical description of this pattern, known as the Parker spiral. *See* MAGNETOHYDRODYNAMICS; PLASMA (PHYSICS).

Spacecraft observations over several decades have confirmed Parker's basic model. However, there are some notable departures. For example, turbulence and waves produce large deviations of the magnetic field. It has also been suggested that motion of the foot points of the field lines on the surface of the Sun could lead to deviations from the Parker spiral at large distances from the Sun.

The heliospheric current sheet (HCS) is a surface that separates regions of opposite magnetic polarity. It is a fundamental feature of the heliosphere that is intimately related to the global dynamical flow of the solar wind. While it is the largest coherent structure within the heliosphere, it is also completely passive, responding to the dynamics of the solar wind flow pattern. The shape of the heliospheric current sheet also plays an important role in the modulation of galactic cosmic rays, which drift along it. *See* COSMIC RAYS.

Dynamic processes. As material flows away from the Sun, the combined effects of solar rotation together with different intrinsic speeds of the flow lead to slower material being caught by faster material, and faster material outrunning slower material. In the former case, the fast material plows into the slower material compressing and heating it. This compressed region is known as an interaction region (IR), and when the flow pattern back at the Sun does not change much with time, it is further known as a corotating interaction region. The interaction region or corotating interaction region is

Average properties of the slow and fast solar wind as observed by interplanetary spacecraft

Parameter	Slow wind	Fast wind
Proton speed	350 km/s (220 mi/s)	750 km/s (470 mi/s)
Proton density	11 cm^{-3}	3 cm^{-3}
Proton temperature	55,000 K	300,000 K
Helium to proton abundance	2.5%	4.3%

initially bound by waves that propagate away from it. As the entire structure propagates away from the Sun, these waves steepen nonlinearly, and by about 2–3 AU they have become shocks. In the case of faster material outrunning slower material, a region of rarefied plasma, known as a rarefaction region (or expansion fan), is created. *See* SHOCK WAVE.

Figure 3 summarizes many of these concepts. These are results from a global numerical simulation of the inner heliosphere during solar minimum conditions. The heliospheric current sheet is essentially flat with a single warp in it. The solar wind speed at high latitudes (over the poles of the Sun) is fast and steady. About the equator the flow is slower and more variable. The heliospheric current sheet folds back on itself by 4 AU. This is due to the acceleration of some material and the deceleration of other material. Magnetic field lines, which were initially radial near the Sun, are wound up into Parker spirals.

Time-dependent effects. Coronal mass ejections are spectacular events that involve the expulsion of massive amounts of solar material into the heliosphere. To initiate such an eruption requires a large amount of energy. It is thought that this energy is supplied by the magnetic field in the corona; however, how magnetic energy is converted into the kinetic energy needed to drive the eruption is not well understood. It is believed that a magnetic reconnection plays a central role.

Knowledge of coronal mass ejections is important for a number of reasons. First, coronal mass ejections appear to play a fundamental role in the way the corona responds to changes in the solar magnetic field associated with the 11-year solar activity cycle. The coronal mass ejections effectively provide a mechanism for the Sun to shed magnetic flux, which is thought to be essential to the cyclic renewal of the solar dynamo. Second, coronal mass ejections contribute a small, but significant amount to the overall solar wind flow at low latitudes. Third, fast coronal mass ejections have been identified as the leading cause of strong geomagnetic storms.

Figure 4 shows the evolution of a coronal mass ejection as it erupts from the Sun. Observations such as these have led to the suggestion that coronal mass ejections are often composed of helical magnetic fields, or flux ropes. Measurements of the magnetic field by interplanetary spacecraft subsequently added support to this picture.

At high heliographic latitudes, observations by the solar polar-orbiting spacecraft *Ulysses* have led to the identification of a new class of coronal mass ejections. These so-called overexpanding coronal mass ejections are probably ejected with internal pressures considerably higher than that of the surrounding solar wind, but with comparable speeds. This "overpressure" causes the coronal mass ejections to expand relative to the ambient solar wind, eventually producing an expansion shock that surrounds the coronal mass ejections and a region of low pressure within their central portions. While it is possible that these events are a distinct class from the low-latitude coronal mass ejections that have been studied for

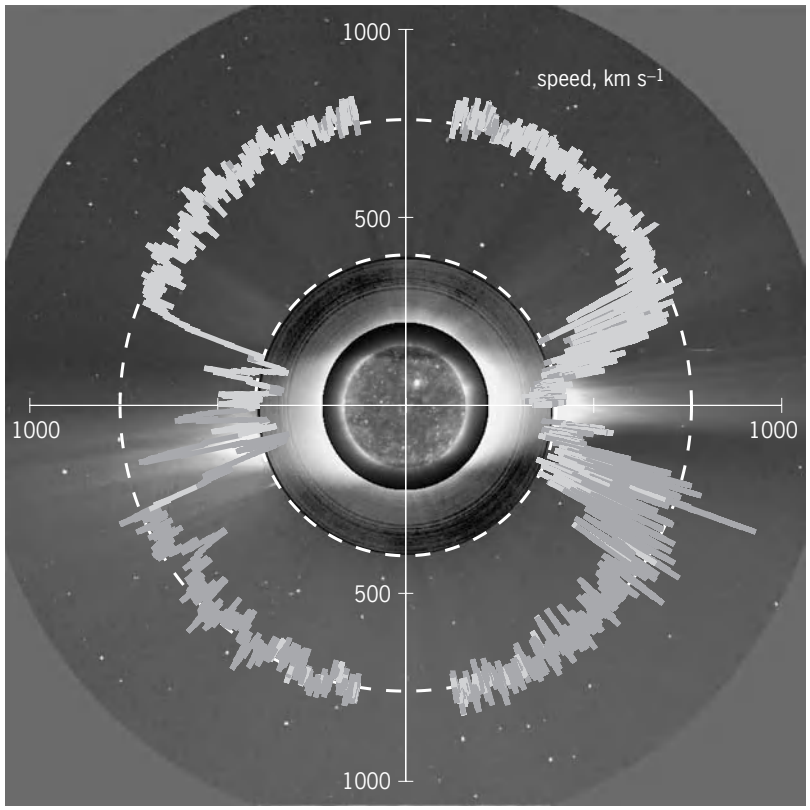


Fig. 2. Circular graph of the speed of the solar wind as measured by the *Ulysses* spacecraft, superposed on an image of the Sun and solar corona. The image is a composite made of an ultraviolet image of the solar disk from the EIT instrument on the *Solar and Heliospheric Observatory (SOHO)* and two white-light coronagraph images, one from the ground-based Mauna Loa MK3 and the other from *SOHO's* LASCO C2 instrument. The speed has been shaded to reflect the magnetic polarity of the field measured at *Ulysses*, light shading indicating outward (away from the Sun) polarity and dark indicating inward polarity.

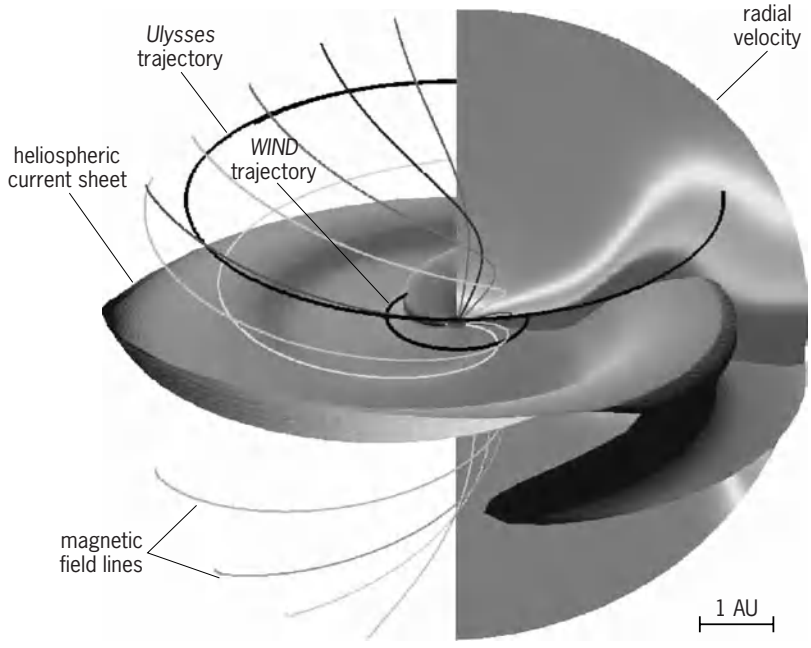


Fig. 3. Model of the heliosphere during solar activity minimum. The heliospheric current sheet is shown out to 5 AU. A slice of the radial velocity is shown at an arbitrary longitude. The darkest shading corresponds to slowest speeds (350 km/s or 220 mi/s) while the lightest shading corresponds to fastest speeds (750 km/s or 470 mi/s). Superimposed is a selection of interplanetary magnetic field lines originating from different latitudes. The trajectories of the *WIND* and *Ulysses* spacecraft are also marked.

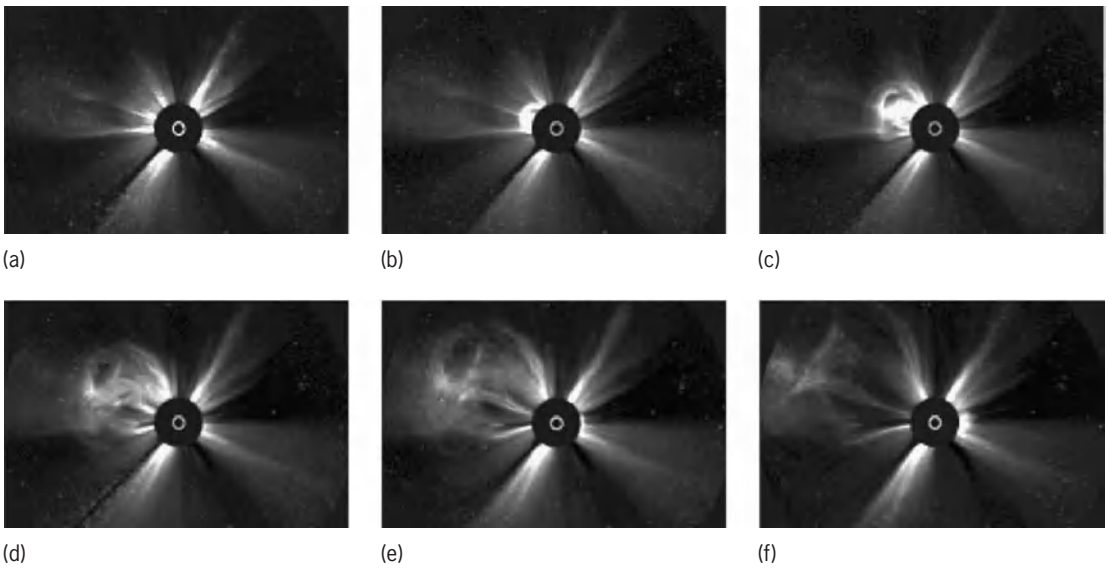


Fig. 4. Evolution of a coronal mass ejection (CME) observed over approximately 8 h on August 5–6, 1999. The white circle marks the location of the Sun, while the dark occulting disk in the center blocks the Sun. Images were taken by the LASCO C3 Coronagraph on board the *SOHO* spacecraft. (*a*) 18:18 Universal Time, August 5. (*b*) 18:42. (*c*) 21:18. (*d*) 23:18. (*e*) 00:42, August 6. (*f*) 2:42. (*Courtesy of the SOHO LASCO consortium. SOHO is a project of international collaboration between ESA and NASA.*)

decades, it is more likely that they represent a high-latitude extension of them.

Solar cycle variations. Over the course of approximately 11 years, the number of sunspots observed on the surface of the Sun rises from almost zero (solar minimum) to several hundred (solar maximum) and then falls back to zero. Other solar and solar wind parameters exhibit similar variations. For example, the pattern of coronal holes, which are the source of the fast solar wind, varies systematically with the solar cycle. At solar minimum, there are large, well-developed coronal holes over the northern and southern polar regions. By solar maximum, they have disappeared, being replaced by smaller, more transient coronal holes at essentially all heliographic latitudes. In turn, the fundamental character of the solar wind changes with the solar cycle. Whereas at solar minimum the solar wind consists of a simple two-state wind—slow, variable flow at low latitudes and fast, quiescent flow at higher latitudes—at solar maximum the flow pattern is much more complex. Fast flow can be found in equatorial regions as often as slow flow exists over the poles of the Sun.

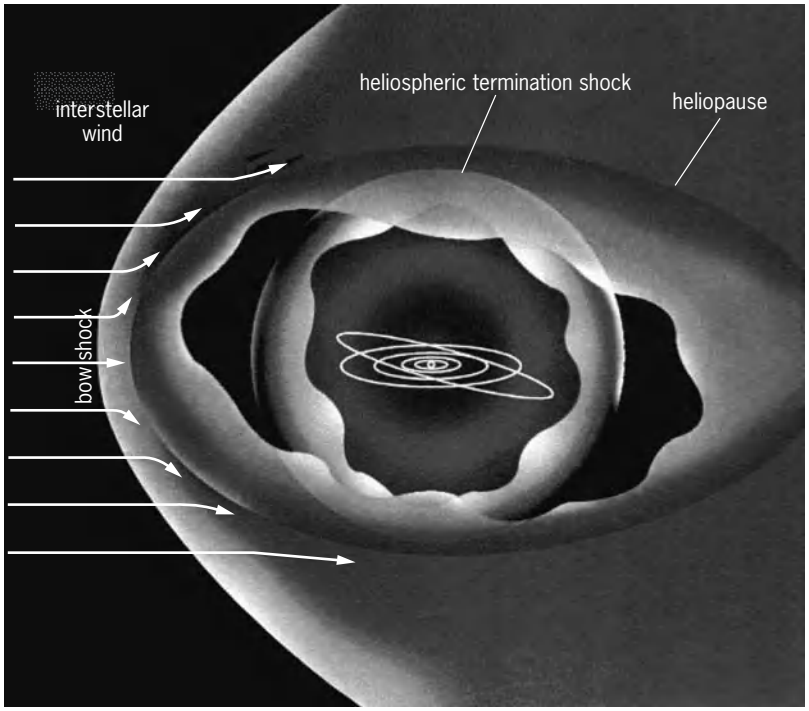


Fig. 5. Interaction of the heliosphere with the local interstellar medium. The heliopause marks the boundary between the heliosphere and the interstellar gas outside. Solar wind approaching this boundary slows suddenly, forming a shock wave. The orbits of the planets are shown by the closed curves. (*Jet Propulsion Laboratory*)

Outer heliosphere. Much of our knowledge of the heliosphere comes from interplanetary spacecraft, which measure the local properties of the surrounding plasma. As such, early knowledge of the solar wind was restricted to the environment surrounding the Earth, where spacecraft were most easily placed. As more ambitious missions were developed, explorations began in the region between the Sun and Earth (for example, *Helios* and *Pioneer Venus Orbiter*), high latitudes (*Ulysses*), and the outer heliosphere (*Pioneer 10* and *11* and *Voyager 1* and *2*).

Farther away from the Sun, the general characteristics of the solar wind change. High-frequency variations are damped out, leaving only large-scale structures. In much the same way that corotating interaction regions are formed from fast streams overtaking slower streams in the inner heliosphere, at larger distances from the Sun, more complex interactions can arise when one corotating interaction

region (or coronal mass ejection) overtakes another. The structures produced by such interactions are termed merged interaction regions (MIRs). They may be localized in latitude and longitude (such as when two coronal mass ejections interact) or global—essentially smearing into a thin spherical shell spanning 360° in longitude and up to high latitudes. Such structures play an important role in the modulation of galactic cosmic rays.

The solar system is traveling through the local interstellar medium at a speed of approximately 26 km/s (16 m/s). As the solar wind blows away from the Sun it carves out a cavity around it, called the heliosphere (**Fig. 5**). Upstream a bow shock is believed to exist, since the heliosphere is moving through the local interstellar medium supersonically. The surface where the solar wind and local interstellar medium meet is known as the heliopause. However, before reaching this point, the solar wind must be slowed down. This is accomplished by the presence of a shock wave, known as the heliospheric termination shock. Currently both of the *Voyager* spacecraft are moving away from the Sun and toward the termination shock. It is likely that at least one of them will encounter it before 2014.

Kinetic effects. In the preceding discussion, it has been implicitly assumed that the solar wind plasma behaves like a fluid. For environments like the Earth's atmosphere and oceans, this is a reasonable approximation: The collision rate between particles is very high, and the distribution of speeds within the gas closely approximates a Maxwellian distribution. In the solar wind, however, collisions are very rare. Several effects, however, conspire to give the solar wind a fluidlike behavior. The magnetic field, for example, constrains particles to orbit tightly about field lines so that they do not travel large distances between collisions. In addition, plasma instabilities tend to occur whenever there are large deviations from a Maxwellian distribution.

Kinetic effects, however, can be important. The proton temperature parallel to the magnetic field is typically 40% larger than the temperature perpendicular to it. And, while the bulk of the particles may define a Maxwellian-like distribution, secondary proton and alpha peaks in the velocity distribution imply the presence of a beam of particles streaming along the magnetic field. It has been suggested that these secondary beams are produced by reconnection in the lower corona and thus may contribute to the heating and acceleration of the solar wind. Solar wind electrons also display a secondary population, known as the halo. These halo electrons stream along field lines away from the Sun. Sometimes streams are seen in both directions, suggesting that the field lines are connected to the Sun at both ends. This usually coincides with the passage of a coronal mass ejection.

Space weather. The field of space weather is a relatively new discipline that aims to understand the conditions on the Sun, solar wind, magnetosphere, ionosphere, and thermosphere that can (1) affect the performance and reliability of space-borne and ground-based technological systems, and (2) endanger human life or health. An important component of space weather is the interaction of the solar wind with the Earth's magnetosphere. The boundary between these two regions is known as the magnetopause. Under quiet conditions the magnetopause provides a barrier to the flow of the solar wind, allowing only a fraction of the energy carried by the solar wind to enter the magnetosphere and be stored there. However, when solar activity increases, such as in a fast coronal mass ejection, the system is sufficiently perturbed that this stored energy is released suddenly, leading to a geomagnetic storm. Ultimately, it is desirable to be able to predict these events before they occur. For example, although no astronaut has yet been caught outside the shuttle during one of these storms, if such a situation were to occur, severe radiation sickness might result. *See* AEROSPACE MEDICINE; GEOMAGNETIC VARIATIONS; MAGNETOSPHERE; SPACE BIOLOGY; SUN.

Pete Riley

Bibliography. M. J. Carlowicz and R. E. Lopez, *Storms from the Sun: The Emerging Science of Space Weather*, Joseph Henry Press, 2002; K. R. Lang, *The Sun from Space*, Springer-Verlag, 2000; J. B. Zirker, *Journey from the Center of the Sun*, Princeton University Press, 2002.

Soldering

A low-temperature metallurgical joining method in which the solder (joining material) has a much lower melting point than the surfaces to be joined (substrates). Because of its lower melting point, solder can be melted and brought into contact with the substrates without melting them. During the soldering process, molten solder wets the substrate surfaces (spreads over them) and solidifies on cooling to form a solid joint.

The most important technological applications of solders are in the assembly of electronic devices, where they are used to make metallic joints between conducting wires, films, or contacts. They are also used for the routine low-temperature joining of copper plumbing fixtures and other devices. In addition, solder is used in the fusible joints of fire safety devices and other high-temperature detectors; the solder joint liquefies if the ambient temperature exceeds the solder's melting point, releasing a sprinkler head or triggering some other protective operation.

Solder alloys. To be effective, a solder must have a low melting point and form a good wetting bond to the metals it joins. A number of solder compositions are listed in the **table**, along with their melting points. Note that most of these are eutectic (lowest melting point alloy) compositions, and contain significant amounts of tin (Sn) or indium (In). Alloys of metals that are not completely miscible usually have an intermediate, eutectic composition at which the melting point has its minimum value and the molten

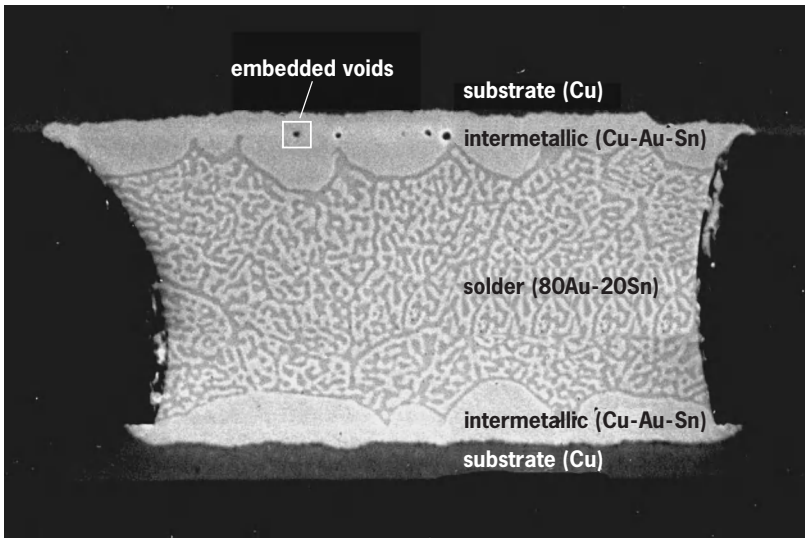
Compositions and melting points of simple solders

Alloy composition, wt %	Melting temperature, °C
51In-49Sn (eutectic)	120
57Bi-43 Sn (eutectic)	139
97In-3 Ag (eutectic)	141
62Sn-38 Pb (eutectic)	183
96.5Sn-3.5 Ag (eutectic)	221
95Sn-5 Sb	240
80Au-20 Sn (eutectic)	280
97Pb-3 Sn	320
87.5Au-12.5 Ge (eutectic)	361

alloy solidifies directly into a mixture of two solid solutions. Tin or indium content is included in solder to facilitate bonding to the metals that are most commonly soldered, such as copper (Cu), nickel (Ni), and gold (Au). Tin and indium form stable intermetallic compounds with copper and nickel, and indium also forms intermetallics with gold. The intermetallic reaction at the solder-substrate interface creates a strong, stable bond. *See* ALLOY; EUTECTICS; INTERMETALLIC COMPOUNDS.

Eutectic tin-lead (Sn-Pb) has been, by far, the most widely used solder, with its melting point conveniently located at 183°C (361°F). Its high tin content is suitable for rapid, efficient bonding to copper, nickel, and gold. In addition, tin-lead solders are easy to manufacture, widely available, and low in cost. This situation is, however, changing rapidly because of legislation intended to minimize the release of lead into the environment. Lead-rich solders are already prohibited in most plumbing fixtures and may be restricted in electronics. *See* LEAD ALLOYS; TIN ALLOYS.

The solders in the table are only a few of the many compositions that are commercially available. Different solders are used because of the need for specific melting-point ranges, for hierarchies of melting points, and for environmentally compatible compositions.



Polished and etched cross section of a eutectic Au-Sn solder joint between copper pads. The distinct phases are shown, as well as the fine voids present at the interface. (*Mark McCormack, Fujitsu Computer Packaging Systems*)

The melting point of the solder must be chosen to satisfy both manufacturing and service issues. For example, low-melting solders are needed for many microelectronic devices, particularly optoelectronic devices containing semiconductors that can be damaged by exposure to even moderate temperature. High-melting solders are required for "underhood" applications in automotive electronics, since the electronic components installed near the engine are exposed to relatively high temperatures.

A hierarchy of solder melting points is required in manufacturing operations where soldered devices are assembled in stages. In manufacturing a microelectronic device, for example, chips may be soldered to chip carriers, which are then soldered to circuit boards. It is desirable that subsequent stages of the operation use a solder with a sufficiently low melting point that the joints fabricated in previous operations are not melted or damaged. To accomplish this, chips may be mounted with 97Pb-3Sn and chip carriers mounted with 62Sn-38Pb. *See* ELECTRONIC PACKAGING; ELECTRONICS; INTEGRATED CIRCUITS; PRINTED CIRCUIT.

Environmental compatibility includes both chemical integrity in the device and environmental tolerance when the device is ultimately discarded. Solders may be metallurgically incompatible with one another [for example, a small amount of lead can cause a serious deterioration in the mechanical properties of eutectic bismuth-tin (Bi-Sn)]; they can exacerbate corrosion problems within the device; or they can be environmentally harmful on disposal, as are tin-lead alloys. *See* CORROSION; METAL, MECHANICAL PROPERTIES OF.

Circuit assembly. The manufacture of a sound solder joint requires a clean substrate to facilitate wetting and intermetallic formation. In particular, the oxide coatings that form spontaneously on copper and nickel must be removed or prevented. The oxide coatings are ordinarily removed by washing with appropriate cleaning agents called fluxes. The flux may be applied to the surface prior to soldering. Alternatively, the flux may be incorporated into the solder. This is usually done by making a solder paste, a mixture of fine solder particles with an appropriate flux. Since the most efficient fluxes are acidic, and pose environmental and safety problems, substrates are often cleaned during their original manufacture and coated for protection. The most common coatings are the solder iteself (in which case the substrate is said to be pre-tinned) and gold, which does not oxidize. Concepts such as flux-free soldering are under development, and include methods such as sputter cleaning of the substrate prior to soldering.

In typical electronic manufacturing, a solder joint is fabricated in two steps. The substrates to be joined are first coated with solder. They are then fitted together and heated to remelt or "reflow" the solder and make the joint. Solder paste is selectively applied to the substrate by stencil printing. A stencil is a mask with open areas that match the substrates that are to be coated, and the solder paste is pressed through the stencil openings using a squeegee. When molten

solder is deposited, a common method is wave soldering, in which the part or board containing the substrates is passed through a continuous stream of molten solder (solder wave). Because of its wetting characteristics, the molten solder adheres only to the substrates.

Solder joint metallurgy. In the **illustration**, a finished solder joint comprises copper substrates joined with eutectic Au-Sn solder. The solder is bonded to the substrates through an intermediate layer of intermetallic which, in this case, is a Cu-Au-Sn ternary compound. The thickness of the intermetallic layer is largely determined by the time of contact with the molten solder, and, hence, increases with the number of reflow operations that must be performed. In extreme cases, the intermetallic may bridge the joint. Since the intermetallic phase is ordinarily brittle, a thick intermetallic layer is undesirable.

The microstructure of the bulk solder in the as-solidified joint depends on its composition and on the rate at which it is solidified. The solidification rate is largely determined by the joint size; the smaller the joint, the higher its surface-to-volume ratio and the more rapidly it cools. Increasing the cooling rate refines the effective grain size of the solder and, generally, improves its mechanical properties. The bulk solder in the illustration has a classic eutectic microstructure, with alternating lamellae of gold- and tin-rich phases. At very high cooling rates the eutectic microstructure tends to be replaced by a fine mixture of small, equiaxed (having equal dimensional in all directions) grains of the two phases. Solder joints in microelectronic devices experience temperatures near their melting points during operation. These high homologous temperatures cause the microstructure to coarsen gradually as the device is used. *See* CRYSTAL STRUCTURE; GRAIN BOUNDARIES; METALLOGRAPHY; METALLURGY.

Another significant feature of the joint cross section shown in the illustration is the array of small voids along the interface. Such voids are common and form during solidification due to entrapped flux or outgassing from the substrate.

Reliability. The mechanical integrity of solder joints in microelectronic devices is a particular concern, since a joint fracture can break electrical contact, causing failure of the device. There are two principal sources of mechanical failure: overload fracture during handling, and thermal fatigue fracture in service. Overload fractures tend to happen at the interface and involve the brittle intermetallic layer. The strength deteriorates as the intermetallic layer thickens, and also decreases with the areal fraction of interfacial voids.

Thermal fatigue failures are caused by cyclic stresses that result from differential thermal expansion and differential heating, mainly when the device is turned on and off. These failures tend to occur in shear, by gradual crack propagation through the body of the solder. Thermal fatigue in solder is complicated by the high homologous temperature at which the joint operates. The deformation mode is dominated by high-temperature creep, whose rate depends strongly on both the temperature and the local microstructure. Moreover, the microstructure is not fixed, but evolves in a pattern that can be strongly affected by the deformation as well as the temperature. Given the difficulty of predicting thermal fatigue life in service, solder joints in critical locations tend to be designed with large margins of safety. *See* CREEP (MATERIALS); SHEAR. J. W. Morris, Jr.

Bibliography. D. R. Frear et al. (eds.), *The Mechanics of Solder Alloy Interconnects*, Van Nostrand Reinhold, New York, 1994; J. S. Hwang, Modern solder technology for competitive electronics manufacturing, McGraw-Hill, New York, 1996; J. H. Lau (ed.), *Solder Joint Reliability: Theory and Applications*, Van Nostrand Reinhold, New York, 1991; R. R. Tummala, E. J. Rymaszewski, and A. G. Klopfenstein (eds.), *Microelectronics Packaging Handbook*, 2d ed., Chapman & Hall, New York, 1997.

Solenodon

A rare insectivorous mammal classified in the family Solenodontidae. Solenodons are among the largest living insectivores. There are only two extant species. The Hispaniolan, or Haitian, solenodon (*Solenodon paradoxus*) is restricted to the remote, wet, densely vegetated central highland regions on the island of Hispaniola (Dominican Republic and Haiti). The coarse pelage (coat) varies from blackish to reddish brown. Most individuals possess a small, square, whitish area on the nape of the neck. Adults are 28–33 cm (11–13 in.) in body length, have a 17.5–26-cm-long (7–10.4-in.) tail, and weigh 600–1000 g (1.2–2.2 lb). The smaller Cuban solenodon, or almique (*S. cubanus*), has longer and finer fur which is blackish brown with white or buff. It is now restricted to Oriente Province in Cuba. Adults are 20–30 cm (10–12 in.) in body length, have a 26–30 cm-long (10.5–12 in.) tail, and weigh 600–700 g (1.3–1.5 lb). Two additional species are known only from skeletal remains in Haiti and Cuba.

Morphology. Solenodons resemble shrews in general appearance except that they are much larger and more stoutly built (see **illustration**). They have relatively large heads, tiny eyes, and large, partially naked ears. The head tapers to a long, flexible, and mobile proboscis that extends well beyond the lower jaw. A unique ball-and-socket joint attaches the proboscis to the skull in the Hispaniolan solenodon and provides even greater mobility. The nostrils are located at the sides of the tip of the proboscis. The senses of touch,

Solenodon, of the family Solenodontidae.

smell, and hearing are highly developed. The front legs are longer than the hind legs. All limbs possess five toes, with the toes on the forelimbs possessing long, stout, sharp claws. The tail is thick, scaly, and almost hairless. Scent glands in the armpit and the groin produce a fatty, musky, greenish secretion that has a goatlike odor. The mammary glands are located in the inguinal region. The dental formula is I 3/3, C 1/1, PM 3/3, M 3/3 × 2, for a total of 40 teeth. The second lower incisor has a goove on the lingual side (solenodon means "animal with grooved teeth"). The submaxillary salivary glands produce a toxic saliva. The grooves on the second lower incisor act as channels for carrying the venom.

Ecology. Solenodons are nocturnal and inhabit forests and brushy edges. Although able to climb, solenodons spend most of their time on the ground. They forage and nest in extensive burrow systems that they excavate in deep layers of humus. Although adult Cuban solenodons are considered to be solitary except when a mother is accompanied by her young, adult Hispaniolan solenodons may live in family groups in caves and natural hollows, with up to eight having been found in the same shelter. Vocalizations consist of chirps, squeaks, clicks, snorts, and grunts. A high-frequency click may function in echolocation. Food consists of a variety of invertebrates, including millipedes, ground beetles and other insects, insect larvae, earthworms, and snails, as well as amphibians and small reptiles and birds. The front claws are used to expose prey beneath rocks, bark, and in the soil.

The frequency and timing of reproduction in the wild in unknown. Mating is believed to be irregular and independent of the seasons. Normally a single young is born. At birth, young solenodons are naked except for long vibrissae on the snout and a small amount of fine, transparent body hair. They weigh about 40–55 g (1.4–1.9 oz). The first solid food is consumed at about 13 weeks of age. A young solenodon will remain with its mother for several months. Longevity in the wild is unknown. In captivity, a Hispaniolan solenodon has lived for more than 11 years, while a Cuban solenodon survived for 6.5 years.

Status. Both species are rare and are listed as highly endangered in the Red Data Book of the International Union for Conservation of Nature and Natural Resources (IUCN) and by the U. S. Department of the Interior. Their decline is due to their low reproductive rate, habitat destruction (deforestation), increasing human activity, and predation by introduced carnivores (dogs, feral cats, and mongooses). Mongooses were introduced to Hispaniola in 1872 for snake control, but they have also killed many other forms of wildlife, including solenodons. Prior to human settlement and the introduction of other carnivores, solenodons were the only natural terrestrial predators on the two islands on which they live.

See INSECTIVORA.

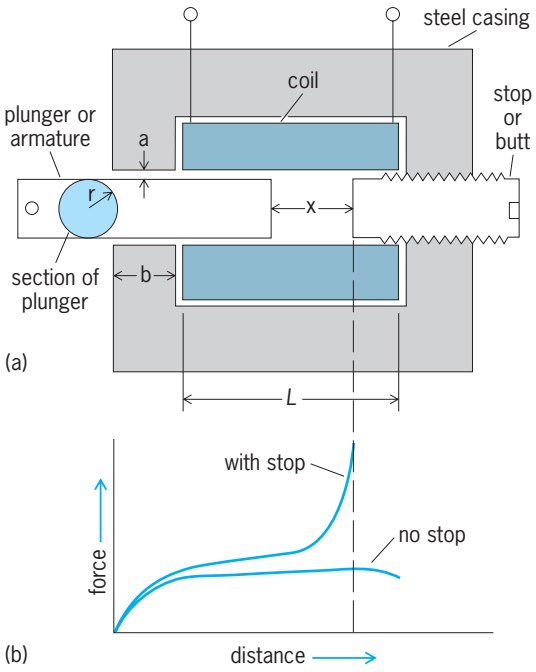
Donald W. Linzey

Bibliography. D. Macdonald (ed.), *The Encyclopedia of Mammals*, Andromeda Oxford, 2001; R. M. Nowak, *Walker's Mammals of the World*, 6th ed., Johns Hopkins University Press, 1999.

Solenoid (electricity)

An electrically energized coil of insulated wire which produces a magnetic field within the coil. If the magnetic field produced by the coil is used to magnetize and thus attract a plunger or armature to a position within the coil, the device may be considered to be a special form of electromagnet and in this sense the words solenoid and electromagnet are synonymous. In a wider scientific sense the solenoid may be used to produce a uniform magnetic field for various investigations. So long as the length of the coil is much greater than its diameter (20 or more times), the magnetic field at the center of the coil is sensibly uniform, and the field intensity is almost exactly that given by the equation for a solenoid of infinite length.

When used as an electromagnet of the plunger type, the solenoid usually has an iron or steel casing. The casing increases the mechanical force on the plunger and also serves to constrain the magnetic field. The addition of a butt or stop at one end of the solenoid greatly increases the force on the plunger when the distance between the plunger and the stop is small. The **illustration** shows a steel-clad solenoid with plunger and plunger stop. The relation of force versus distance with and without the stop is also shown.



Steel-clad solenoid. (*a*) Cross-sectional view. (*b*) Relation of the force acting on the armature to the displacement of the armature.

The force for the solenoid rapidly increases as the plunger enters the coil because of the rapid rate of change of the reluctance for the magnetic path. For the solenoid shown in the illustration, the force on the plunger is given approximately by the equation below, where I is the coil current (in amperes), N is

$$F = ANI\left(\frac{\mu_0 NI}{2(c+x)^2} + \frac{1}{CL}\right) \text{ (newtons)}$$

the number of turns, $A = \pi r^2$ is the cross-sectional area of the plunger (in square meters; 1 ft^2 = 0.0929 m^2), μ_0 is the permeability of free space ($4\pi \times 10^{-7}$ henry/m), r, a, b, x, and L are as shown in the illustration (in meters; 1 ft = 0. 3048 m), and $c = ra/2b$. The value of the constant C depends upon the dimensions of the solenoid and the degree of plunger magnetization. For normal designs, C has a value in the range 0.6 to 2.9. For a conservative design, one may use the largest value for C or neglect the second term entirely. This term is due to leakage fluxes and becomes negligible as the plunger closes the air gap x.

In principle a solenoid works with either ac or dc excitation. In the dc solenoid the flux is always at its maximum value for a fixed plunger position. In an ac solenoid the force varies at twice the frequency of the supply voltage. The variation, which is caused by the variation of the magnetic flux, gives rise to excessive chattering or vibration unless a shading coil is embedded in the face of the plunger stop. The shading coil acts to smooth the variation of flux and thereby smoothes the attractive force.

When direct current is used, only the resistance of the coil wire limits the final value of current, while with ac excitation the inductance of the coil must also be considered. It is difficult to calculate the proper value of inductance, because it is a function of the position of the plunger. As the plunger moves into the solenoid, the inductance, or flux per ampere, becomes much greater because the reluctance of the flux path is much less. The current drawn by the coil for constant applied ac voltage becomes smaller as the plunger moves into the solenoid. It is common practice to laminate the plunger, stop, and casing in ac service to reduce eddy current losses. *See* ELECTROMAGNET. Jerome Meisel

Bibliography. D. G. Fink and H. W. Beaty (eds.), *Standard Handbook for Electrical Engineers*, 14th ed., 2000; G. R. Jones, M. A. Laughton, and M. G. Say (eds.), *Electrical Engineer's Reference Book*, 15th ed., 1993; D. B. Montgomery, *Solenoid Magnet Design*, 1969, reprint 1980; Y. M. Pulyer, *Electromagnetic Devices for Motion Control and Signal Processing*, 1992.

Solenoid (meteorology)

In meteorological usage, solenoids are hypothetical tubes formed in space by the intersection of a set of surfaces of constant pressure (isobaric surfaces) and a set of surfaces of constant specific volume of air (isosteric surfaces) or density (isopycnic surfaces). The isobaric and isosteric surfaces are such that the values of pressure and specific volume, respectively, change by one unit from one surface to the next. The state of the atmosphere is said to be barotropic when there are no solenoids, that is, when isobaric and isosteric surfaces coincide. The number of solenoids cut by any plane surface element of unit area is a measure of the torque exerted by the pressure gradient force, tending to accelerate the circulation of air around the boundary of the area. *See* BAROCLINIC FIELD; BAROTROPIC FIELD; ISOPYCNIC.

Frederick Sanders; Howard B. Bluestein

Bibliography. S. L. Hess, *Introduction to Theoretical Meteorology*, 1959, reprint 1979; H. G. Houghton, *Physical Meteorology*, 1985.

Solenopora

A genus of extinct calcareous red algae (Rhodophyta). *Solenopora* is characterized by encrusting, rounded, nodular growth forms ranging in size from several millimeters to a few centimeters. Internally, this alga consists of calcified rows or filaments of cells, commonly polygonal in shape, ranging from 20 to 100 micrometers in diameter. Although the cellular tissue is similar to modern coralline algae, the cells in *Solenopora* are generally larger and are not differentiated into more than one kind of cellular arrangement.

Solenopora and related fossil genera have been referred to invertebrate groups, but now there is general agreement that these fossils constitute a coherent group of extinct red algae separate from the Corallinaceae. Solenoporids probably provided an ancestral stock to the true coralline algae.

Solenopora appeared first in Cambrian time and became extinct in the early part of the Cenozoic. The genus flourished during the Ordovician, Silurian, and Jurassic periods. The environmental distribution of *Solenopora* was comparable to some modern coralline algae. *See* ALGAE. John L. Wray

Bibliography. J. L. Wray, *Calcareous Algae*, 1977.

Solid (geometry)

A three-dimensional geometric figure consisting of points continuously connected, and separated from the rest of space by a surface called the boundary of the solid. Points on the boundary are usually considered part of the solid. A solid is bounded if there exists a sphere having finite radius that could enclose the solid. A solid is convex if all points of any line segment having end points on the boundary also are points of the solid.

Volume. A measure of a solid is its volume: the amount of space enclosed by the boundary surface. A unit cube is a cube having an edge that is one unit long, where the unit normally is a standard unit of length. The volume of a unit cube is defined to be 1 cubic unit, or 1 $unit^3$. The volume of a solid is a number mulitiplied by the volume of a unit cube, with the number telling how many of these unit cubes would be required to occupy the same amount of space as does the solid. The volume of a bounded solid is finite; an unbounded solid might have finite or infinite volume. *See* UNITS OF MEASUREMENT.

Many solids have volumes that can be determined from formulas that involve one or more dimensions associated with the solid. A dimension is any measure associated with a geometric figure that has units of

length, such as height, width, or perimeter. Angles are not dimensions (they do not have units of length) but appear in some formulas for volumes.

Cavalieri's principle. The volumes of a large number of solids can be found with Cavalieri's principle: If a solid can be oriented so that all horizontal plane sections have the same cross-sectional area A, and if b is the distance between the lowest and highest plane sections, then the volume V is given by Eq. (1).

$$V = Ab \tag{1}$$

Prismoidal formula. Some solids have volumes given by the prismoidal formula: If the solid can be oriented so that the area A of any horizontal plane section satisfies $A = f(d)$, where d is the distance of that section from the lowest point of the solid, and f is a polynomial function of degree three or less, then the volume V is given by Eq. (2), where L, U,

$$V = \tfrac{1}{6}b(L + 4M + U) \tag{2}$$

and M are, respectively, the areas of the lowest plane section, the highest plane section, and a plane section midway between the lowest and highest (one or more of these areas might be zero); and b is the height of the solid (the distance between the highest and lowest plane sections). Prismatoids, spheres, ellipsoids, pyramids, frustrums, cones, cylinders, and many other solids satisfy this property. However, since the area of the midsection is usually not easily determined, the prismoidal formula is seldom a practical method for determining a volume.

Integral calculus techniques. Integral calculus supplies general techniques for determining the volumes of solids whose boundaries can be described with equations. These techniques can be classified as one-, two-, or three-variable. *See* CALCULUS; INTEGRATION.

1. One-variable: If the solid can be oriented so that cross sections perpendicular to an axis (say, the x axis) have areas $A(x)$ [the area is a function of x], then the volume V is given by Eq. (3), where a and b de-

$$V = \int_a^b A(x)\,dx \tag{3}$$

fine the limits of the x axis for which this functional relationship holds. Some solids might have volumes that can be found by partitioning the solid, applying the technique to each part, and summing the volumes thus obtained.

A solid of revolution is the part of space that is said to be swept out when a region of a plane is revolved about a line in that plane that does not contain an interior point of the region. A solid of revolution also can be defined as a solid whose boundary consists only of surfaces of revolution having a common axis. The volume of a solid of revolution might be found by the so-called cylindrical shell method, in which the solid is positioned so that the axis of revolution is the y axis. If the region that is being revolved can be defined by $y = U(x)$ and $y = L(x)$ [the upper and lower curves that are boundaries of the region], and $x = a$ and $x = b$ (the left and right straight-line boundaries of the region), then the volume V is given by Eq. (4).

$$V = \int_a^b 2\pi x[U(x) - L(x)]\,dx \tag{4}$$

See SURFACE (GEOMETRY).

2. Two-variable: If the solid can be oriented so that the up direction is the positive z axis, the upper surface of the solid is defined by $z = U(x,y)$, the lower surface is defined by $z = L(x,y)$, and the remaining boundary is a cylindrical surface whose generator is parallel to the z axis, then the volume V is given by Eq. (5). Here A represents the region of the xy plane

$$V = \iint_A U(x,y) - L(x,y)\,dA \tag{5}$$

enclosed by the cylindrical surface—the region over which the double-integral is to be evaluated.

3. Three-variable: If $\mathscr{VV}$ represents the part of space occupied by a solid, then the volume V is given by Eq. (6).

$$V = \iiint_{\mathscr{VV}} dV \tag{6}$$

The evaluations of the double and triple integrals above are normally attempted by replacing them with iterated integrals. The proper choice of coordinate system (rectangular, cylindrical, or spherical) and the order of integration can have an effect on the ease (or even the possibility) of obtaining the volume. If an expression for the volume cannot be obtained analytically, then numerical techniques can be used to obtain an approximate volume.

Volume theorem of Pappus. Some solids of revolution have volumes that are best found by using the volume theorem of Pappus: If a plane region is revolved about a line in that plane that does not contain an interior point of the region, then the volume of the solid of revolution generated is equal to the area of the region multiplied by the distance that the centroid of the region moves during one revolution. (If a thin material, such as sheet metal, were cut in the shape of the region, then the centroid of the region is the point where the region would balance.) This theorem is illustrated by a solid torus, the solid of revolution generated when a circular region is revolved about a line that does not intersect the circle. Clearly, the centroid of this region is the center of the circle. If R is the distance between this center and the axis of revolution, and r is the radius of the circle, then the area of the circle is πr^2 and the distance traveled by the centroid during one revolution is $2\pi R$. The volume V of a solid torus, therefore, is given by Eq. (7).

$$V = 2\pi^2 Rr^2 \tag{7}$$

See TORUS.

Important solids. A number of solids are frequently encountered in pure and applied mathematics, in addition to those already discussed.

Solids often receive the name of their boundary surface. In formal mathematical usage, "sphere,"

"cylinder," and "cone" are surfaces, but often these words refer to the solids having such surfaces as part or all of their boundary. Usually the distinction can be made from the context. The use of the word "solid" can remove any ambiguity.

Polyhedron. A polyhedron is any of a large class of bounded solids whose boundary surface consists of plane polygonal regions. Examples include cubes, pyramids, prisms, and many other familiar solids. *See* CUBE; POLYHEDRON.

Sphere. A solid sphere has a boundary that is a sphere. The size of a sphere is completely determined by either its radius or its diameter, which refer to the length of a radius (a line segment that joins the center with a point on the sphere) or the length of a diameter (a line segment containing the center whose end points lie on the sphere). The volume V of a solid sphere having radius R is given by Eq. (8).

$$V = \tfrac{4}{3}\pi R^3 \tag{8}$$

Two parallel planes that intersect a sphere, together with the part of the spherical surface between those planes, define a solid: a spherical segment (shaped like a tomato slice). The intersection of each plane with the sphere is a circle (a tangent plane intersects at only one point, which can be considered a degenerate circle with radius 0), and if the radii of these two circles are R_1 and R_2, and h is the distance between the planes, then the volume V of the spherical segment is given by Eq. (9).

$$V = \tfrac{1}{6}\pi h(3R_1^2 + 3R_2^2 + h^2) \tag{9}$$

A plane that contains the center of a sphere divides the sphere (both the surface sphere and the solid sphere) into two hemispheres (yet another word that can refer to either a surface or a solid). Another plane that contains the center will cut a solid hemisphere into two lunar segments (the smaller will be shaped like an orange segment). If θ is the radian measure of the dihedral angle formed by the planes that define the lunar segment, and R is the radius of the sphere, then the volume V of the lunar segment is given by Eq. (10).

$$V = \tfrac{2}{3}\theta R^3 \tag{10}$$

See SPHERE.

Cylinder. A cylindrical solid is formed when a cylindrical surface is intersected by two parallel planes (**Fig. 1**). The cylindrical surface between the planes is the lateral surface of the solid; the remaining two congruent regions are the bases of the solid. If each base has finite area A, and the distance between the planes (the height of the cylindrical solid) is h, then the solid's volume V will be given by Cavalieri's principle, Eq. (1). A cylindrical solid often is named by giving information about the shape of a base, and how the plane intersects a generator of the cylinder. The word "right" indicates that the plane is perpendicular to the generator. A prism is a special case of a cylindrical solid, in which the lateral surface

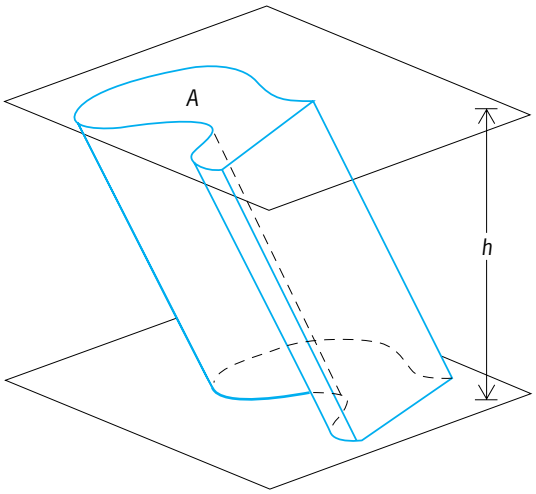


Fig. 1. Cylindrical solid. The height of the solid is *h*, and the area of the bases is *A*.

consists of plane regions. *See* PRISM.

When referring to a solid, the word "cylinder," unaccompanied by any modifying words, invariably means a right circular cylinder. The axis of revolution is the axis of the cylinder. The bases of the cylinder are the circular regions defined by the intersection of the planes and the cylindrical surface. The radius of the cylinder is the radius of a base; the height (or altitude) of the cylinder is the distance between the bases. If r and h denote the radius and height, then the volume V of the cylinder is given by Eq. (11).

$$V = \pi r^2 h \tag{11}$$

See CYLINDER.

Cone. A conical solid is bounded by one nappe of a cone and a plane that intersects that nappe (**Fig. 2**).

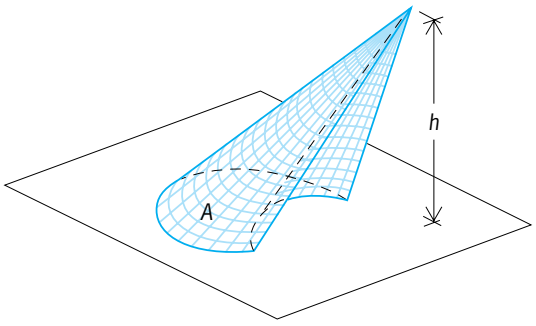


Fig. 2. Conical solid. The height of the solid is *h*, and the area of the base is *A*.

The plane region defined by this intersection is the base of the conical solid; the part of the nappe between the plane and the vertex is the lateral surface of the conical solid. The vertex of the solid is the vertex of the cone; the distance between the vertex and the plane is the height of the conical solid. If the base has area A and the height is h, then the volume

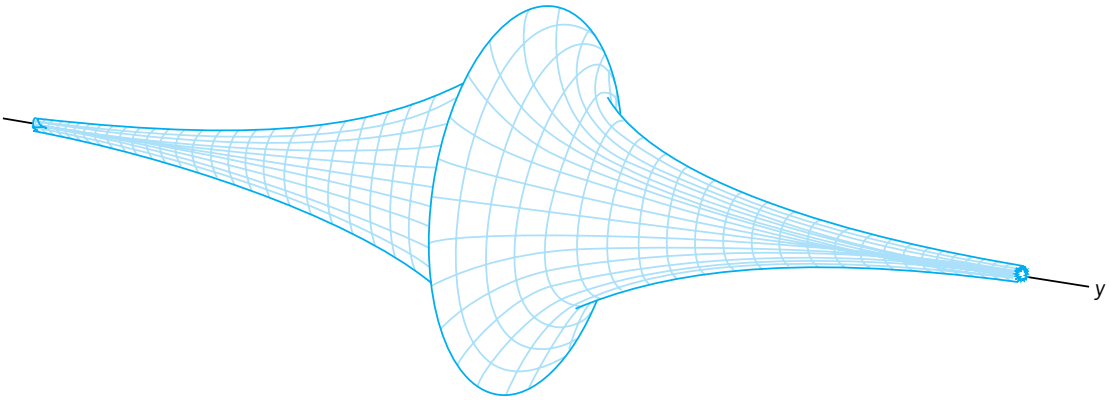


Fig. 3. Pseudosphere formed by revolving a tractrix about the *y* axis.

V of the conical solid is given by Eq. (12). A pyramid

$$V = {}^1\!/_3 Ah \tag{12}$$

is a special case of a conical solid, in which the lateral surface consists of plane regions.

When used as the name of a solid, the single word "cone" invariably means a right circular cone: a solid whose boundary consists of one nappe of a circular cone and a plane perpendicular to the axis of revolution, which also is the axis of the cone. If r and h denote the radius and height, then the volume V of the cone is given by Eq. (13).

$$V = {}^1\!/_3 \pi r^2 h \tag{13}$$

See CONE.

Frustrum. A frustrum (or frustum) of a conical solid is the solid remaining if a plane parallel to the base intersects the solid, and the part between that plane and the vertex is discarded. The intersecting plane defines a second (smaller) base of the frustrum. If B and b are the areas of the bases of the frustrum, and H is the height of the frustrum (the distance between the bases), then the volume V of a frustrum of a conical solid is given by Eq. (14).

$$V = {}^1\!/_3 H(B + \sqrt{Bb} + b) \tag{14}$$

Ellipsoid. A solid ellipsoid is the solid whose boundary is an ellipsoid. A diameter of an ellipsoid is a line segment containing the center that has end points on the surface. The shortest and longest of all the diameters of the ellipsoid will be perpendicular, and a third diameter will be perpendicular to both. If these three diameters have lengths $2a$, $2b$, and $2c$, then the volume V of the solid ellipsoid is given by Eq. (15). If two of the three diameters are of equal

$$V = {}^4\!/_3 \pi abc \tag{15}$$

length, then a solid spheroid (a solid ellipsoid of revolution) is generated. The spheroid will be oblate or prolate, according to whether the third diameter is respectively shorter than or longer than the diameters having equal length.

Pseudosphere. A solid pseudosphere is an unbounded solid of revolution formed when the region between a certain tractrix and a certain line is revolved about that line. If an object is tied to a cord of constant length, and the other end of the cord follows some curve, then a tractrix is the path followed by the object as it is dragged slowly through a resisting medium. A pseudosphere can be formed by revolving this tractrix about the y axis (**Fig. 3**): The end of the cord moves from the origin along the positive y axis, the initial position of the dragged object is the point $(a,0)$, and the length of the cord is a. (The entire tractrix consists of this first-quadrant curve and its reflection about the x axis.) A solid pseudosphere is formed by revolving the region lying between this tractrix and the y axis. The solid thus formed is not bound (since the tractrix is asymptotic to the y axis) but has a finite volume V given by Eq. (16).

$$V = {}^2\!/_3 \pi a^3 \tag{16}$$

See TRACTRIX.

Harry L. Baldwin, Jr.

Bibliography. A. A. Albert, *Solid Analytic Geometry*, 1966; J. S. Frame, *Solid Geometry*, 1948; D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, 1932, reprint 1990; P. Pearce and S. Pearce, *Polyhedra Primer*, 1978.

Solid solution

Compositional variation of a crystalline substance due to substitution or omission of various atomic constituents within a crystal structure. Solid solutions can be classified as substitutional, interstitial, or omissional. They may also be categorized by the nature of their thermodynamic properties, such as enthalpy, entropy, and free energy (for example ideal, nonideal, and regular solid solutions).

Characteristics. The concept of solid solution can be understood by considering a specific mineral group such as olivine. Although olivine-group minerals may exhibit a range of compositions they all have a similar crystal structure consisting of isolated silicate $(SiO_4)^{4-}$ tetrahedra cross-linked by cations

in two specific types of sites, both of which are octahedrally (eightfold) coordinated to oxygen atoms. Thus all of the minerals in the olivine group are isostructural (having a similar crystal structure). Yet within this structural framework, compositions vary considerably. Such chemical variation can be described by defining the nature and extent of the atomic substitutions involved or by describing intermediate compositions in terms of limiting end-member compositions. For example, the composition of the mineral olivine ranges between the end members forsterite (magnesium silicate; Mg_2SiO_4) and fayalite (ferrous silicate; Fe_2SiO_4), representing varying amounts of iron (Fe^{2+}) and magnesium (Mg^{2+}) ion substitution. A complete solid solution series involving iron/magnesium substitution also exists between the olivine group minerals monticellite ($CaMgSiO_4$) and kirschteinite ($CaFeSiO_4$). There is, however, little solid solution between the monticellite series and the olivine series because the larger calcium ion (Ca^{2+}) does not readily substitute for either the magnesium ion or the ferrous iron ion. Thus in some instances solid solution between end members may be complete, whereas in other cases it may be limited. In some minerals, atomic substitution must be coupled in order to maintain electrical charge balance. In plagioclase feldspar, defined by the end-member compositions albite ($NaAlSi_3O_8$) and anorthite ($CaAl_2Si_2O_8$), compositional substitution of sodium (Na^{1+}) ion for calcium ion is coupled with the substitution of the silicon (Si^{4+}) ion for the aluminum (Al^{3+}) ion. *See* FELDSPAR; OLIVINE.

In addition to substitution within the crystal structure, compositional variation may take place by interstitial substitution or omission solid solution. Interstitial solid solution occurs when ions or atoms occupy a position in a crystal structure that is usually vacant. For example, this is relatively common in the zeolite mineral group and in beryl. When charged ions are thus incorporated in the crystal, additional coupled substitution elsewhere in the structure is necessary in order to maintain charge balance. Omission solid solution occurs when an atom or ion is missing from a specific crystallographic position. For example, the composition of the mineral pyrrhotite deviates from pure ferrous sulfide (FeS) by the omission of ferrous ion from the structure. In order to maintain charge balance, as ferrous ion is omitted, some of the iron present in pyrrhotite most likely exists as ferric (Fe^{3+}) ion. *See* BERYL; ZEOLITE.

Exsolution. Exsolution refers to the separation of a mineral solid solution that was initially stable as a single homogeneous substance into two or more substances with different compositions. This process generally occurs in response to decreasing temperature, with the aggregate composition of the new mixture remaining the same as the initial, single, homogeneous solid solution. Because this separation takes place in a solid substance, nucleation of the newly formed material commonly occurs on specific crystallographic planes or directions in the original mineral, producing exsolution lamellae.

Occurrence. Solid solution is widespread among minerals. In fact, very few naturally occurring minerals exist as pure end-member substances but exhibit trace to extensive solid solution. The extent of solid solution depends upon the relative sizes of the atoms or ions involved, the charges of the ions, the coexistence and composition of other minerals or liquid (for example, magma), and the temperature and pressure conditions of formation (with temperature having a more pronounced effect). Generally, solid solution among ions that differ in size by less than 15% is significant, whereas solid solution among ions that differ in size by more than 30% is very limited. If ions differ in charge, a coupled solid solution is required to maintain charge balance. In addition, the extent of solid solution reflects the abundance of various elements at the time of mineral formation. Although a complete solid-solution series between calcium carbonate ($CaCO_3$) and cadmium carbonate ($CdCO_3$) has been experimentally verified, it has not been observed in naturally occurring specimens because of generally low Cd^{2+} concentrations.

Temperature affects solid-solution limits in two ways. In general, increased temperature expands the limits of solid solution, because increased thermal vibration within a crystal structure reduces the geometrical size constraints on a specific structural site. Second, temperature (and pressure) variation affects the chemical equilibria among minerals and between minerals and fluids (such as magma). This affects the distribution of a given element among coexisting substances (phases) which, in turn, controls the composition of individual minerals. Because solid-solution limits are sensitive to temperature and pressure, mineral compositions and exsolution textures have been widely used to determine conditions of formation and subsequent cooling histories of many igneous rocks.

Compositional variation among metamorphic minerals (such as garnets and amphiboles) and exsolution phenomena have been used to determine temperature and pressure relationships during metamorphism. In addition, many of the optical properties of gemstones and minerals are the direct result of compositional substitutions. For example, the red color of ruby is due to chromium ion (Cr^{3+}) substitution in the mineral corundum (Al_2O_3), whereas titanium ion (Ti^{4+}) substitution in the same mineral produces blue sapphire. Similarly, beryl ($Be_3Al_2Si_6O_{18}$) exhibits a range of color (blue-green: aquamarine; green: emerald; pink: morganite) caused by trace atomic substitutions in the crystal structure. Alloys represent metallurgical solid solutions in which composition is controlled to produce specific physical, chemical, and electrical characteristics. *See* ALLOY STRUCTURES; AMPHIBOLE; CRYSTAL STRUCTURE; CRYSTALLOGRAPHY; GARNET; GEM; METAMORPHISM; MINERALOGY.

John C. Drake

Bibliography. I. S. E. Carmichael and H. P. Eugster (eds.), *Thermodynamic Modeling of Geological Materials: Minerals, Fluids, and Melts*, 1987; C. Klein and C. S. Hurlbut, Jr., *Manual of*

Mineralogy, 21st ed., 1993, rev. 1999; A. Putnis, *Introduction to Mineral Science*; 1992.

Solid-state battery

A battery in which both electrodes and the electrolyte are solids. Solid electrolytes are a class of materials also known as superionic conductors and fast ion conductors, and their study belongs to an area of science known as solid-state ionics. As a group, these materials are very good conductors of ions but are essentially insulating toward electrons, properties that are prerequisites for any electrolyte. The high ionic conductivity minimizes the internal resistance of the battery, thus permitting high power densities, while the high electronic resistance minimizes its self-discharge rate, thus enhancing its shelf life. Examples of such materials include Ag_4RbI_5 for Ag^+ conduction, LiI/Al_2O_3 mixtures for Li^+ conduction, and the clay and β-alumina group of compounds ($NaAl_{11}O_{17}$) for Na^+ and other mono- and divalent ions. At room temperature the ionic conductivity of a single crystal of sodium β-alumina is 0.035 S/cm, comparable to the conductivity of a 0.1 *M* HCl solution. This conductivity, however, is reduced in a battery by a factor of 2–5, because of the use of powdered or ceramic material rather than single crystals. Of much interest are glassy and polymeric materials that can be readily made in thin-film form, thus enhancing the rate capability of the overall system. *See* ELECTROLYTE; ELECTROLYTIC CONDUCTANCE; IONIC CRYSTALS.

In some cases, solid-state batteries also make use of fast ion conductors for the electrodes. These materials are, however, good conductors for both ions and electrons. Examples include the layered-structure disulfides of titanium and vanadium, TiS_2 and VS_2, which can be used as the cathode or sink for lithium, and aluminum-lithium alloys, which can be used as the anode or source of lithium in lithium batteries. These materials can sustain high reaction rates because of their unique crystalline structures which allow the incorporation of ions into their crystalline lattices without destruction of those lattices (**Fig. 1**). An example of such a cell is given by notation (1), having anode reaction (2) and cathode reaction (3). At the anode the face-centered cubic lattice of aluminum expands from 0.405 to 0.638 nanometer, whereas at the cathode the titanium disulfide sandwich structure expands only 10% on incorporation of the lithium.

$$\text{Li-Al}//\text{Li glass}//\text{TiS}_2 \quad (1)$$

$$\text{LiAl} \rightleftharpoons \text{Al} + \text{Li}^+ + e^- \quad (2)$$

$$\text{TiS}_2 + \text{Li}^+ + e^- \rightleftharpoons \text{LiTiS}_2 \quad (3)$$

See INTERCALATION COMPOUNDS; SOLID-STATE CHEMISTRY.

Use. Potentially the largest use of solid electrolytes in the battery field is in the high-energy-density sodium/sodium-β-alumina/sulfur battery. This system is expected to find application for powering of automotive vehicles and for energy storage at central power plants, such as solar plants, provided the lifetime of the ceramic electrolyte can be improved. However, this battery operates at about 570–750°F (300–4000°C), so that both electrodes are in the liquid state. *See* ENERGY STORAGE.

Solid-state batteries generally fall into the low-power-density and high-energy-density category. The former limitation arises because of the difficulty of getting high currents across solid–solid interfaces. However, these batteries do have certain advantages that outweigh this disadvantage: They are easy to miniaturize (for example, they can be constructed in thin-film form), and there is no problem with electrolyte leakage. They tend to have very long shelf lives, and usually do not have any abrupt changes in performance with temperature, such as might be associated with electrolyte freezing or boiling. Being low-power devices, they are also inherently safer. The major applications of these batteries are in electronic devices such as cardiac pacemakers, cameras, electrochromic displays, watches, and calculators.

Configuration. In a solid-state cell (**Fig. 2**), the polycrystalline pressed electrolyte is interspaced between a metallic anode and the solid cathode material. The electrodes are applied to the electrolyte by mechanically pressing the materials together, or in some cases the electrolyte is formed in place by reaction between the two electrodes. These cells are then stacked together to form a battery of the required voltage. A carbon current collector is often used on the cathode side, and this is frequently admixed with the cathode material. In some cases, where the

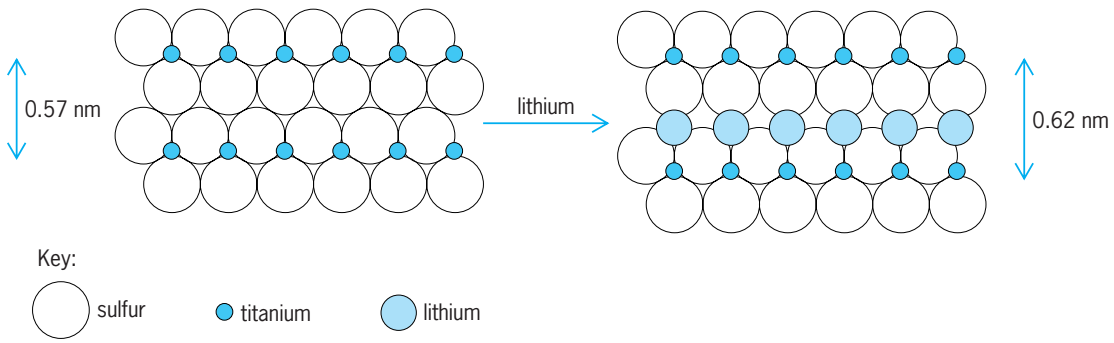


Fig. 1. Schematic illustration of an intercalation cathode reaction.

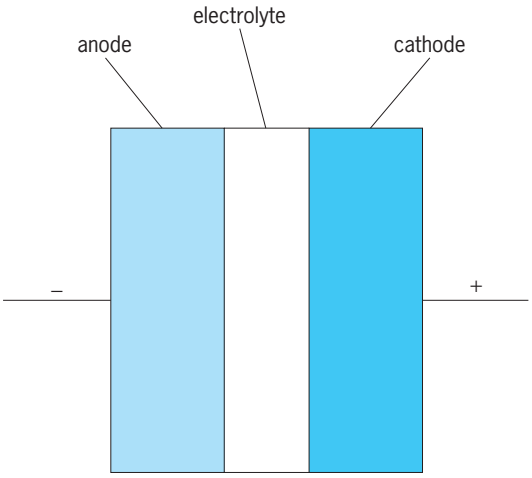


Fig. 2. Schematic diagram of solid-state battery.

cathode material is sufficiently conductive, for example titanium disulfide, no carbon conductor is needed. Although some of the earliest solid-state batteries were based on silver, these have not found any significant applications because of their low energy densities, their high cost, and competition from lithium batteries.

Lithium battery. The lithium-iodine battery employed in many cardiac pacemakers is constructed with a lithium anode and an iodine–polyvinyl pyridine cathode. This cathode has a tarlike consistency in the fresh battery, then solidifies gradually as the battery is discharged. In this cell the solid electrolyte, lithium iodide, is formed in place by the reaction of the lithium and iodine. Because the electrolyte continues to form as the battery discharges, the overall resistance of the cell continually increases with discharge. This results in a drop in cell voltage for a given current drain. The initial cell voltage is around 2.8 V.

The most important characteristics of the lithium battery for this application are its high reliability and very low rate of self-discharge. A typical implantable battery has a volume of 0.4 $in.^3$ (6 cm^3), weighs 0.8 oz (23 g), and has a 2-amperehour capacity. The lifetime of this battery is 5–10 years since pacemakers typically draw only 15–30 microamperes. The battery is considered to be discharged when the output voltage drops to 1.8 V. This may be detected by monitoring the patient's pulse rate. *See* MEDICAL CONTROL SYSTEMS.

A reversible lithium solid-state battery uses a glass as the electrolyte, pure lithium as the anode, and an intercalation compound as the cathode. These compounds, such as titanium disulfide, react reversibly with lithium, as discussed above. The overall resistance, unlike that of lithium-iodine batteries, does not increase with discharge, nor is a conductive element such as carbon necessary in the cathode. The emf of such a cell is around 2 V and decreases slightly but continuously with loss of capacity; this allows the cell voltage to be used as a fuel gage, unlike most conventional batteries where there is an abrupt loss of voltage without warning when the cell is depleted.

An advanced lithium battery uses a lithium carbon compound as anode and lithium cobalt oxide ($LiCoO_2$) as cathode. The anode incorporates lithium between the layers of graphitic coke, thereby containing the lithium safely and making cell rechargeability much more feasible. The lithium cobalt oxide cathode has the same structure as lithium titanium disulfide (Fig. 1) and behaves in the same manner. However, it is much more oxidizing, leading to a cell emf of about 3.5 V. Since this is almost exactly three times that of Ni/Cd or Ni/hydride batteries, this battery can easily be used in their place. Moreover, since a single cell will suffice for such applications as portable computers, battery management is much easier.

Polymeric electrolytes. Solid-state batteries based upon polymeric electrolytes are being actively investigated, particularly for use with lithium and sodium anodes. Energy densities of 5–10 times that of lead-acid cells have been calculated based upon very thin cells combined with high-voltage arrays. These cell designs could eventually lead to rechargeable solid-state batteries with energy densities of 100–200 watthours per kilogram and comparable power levels. They are most likely to be used with lithium anodes and conductive cathodes such as titanium disulfide. The polymer most under study is polyethylene oxide mixed with a lithium salt such as $LiCF_3SO_3$. *See* BATTERY.

M. Stanley Whittingham

Bibliography. B. V. R. Chowdari and S. Radhakrishna (eds.), *Solid State Ionic Devices*, 1988; J. P. Gabano, *Lithium Batteries*, 1983; G. Pistoia (ed.), *Lithium Batteries,* 1994; J-M. Tarascon, Recent advances in rechargeable Li batteries, *Solid State Ionics,* 69:171–368, August 1994; C. A. Vincent, *Modern Batteries*, 2d ed., 1998; M. S. Whittingham and A. J. Jacobson (eds.), *Intercalation Chemistry*, 1982.

Solid-state chemistry

The science of the elementary, atomic compositions of solids and the transformations that occur in and between solids and between solids and other phases to produce solids. Solid-state chemistry deals primarily with those microscopic features which are uniquely characteristic of solids and which are the causes for the macroscopic chemical properties and the chemical reactions of solids. As with other branches of the physical sciences, solid-state chemistry also includes related areas that furnish concepts and explanations of those phenomena which are more characteristic of the subject itself.

The overlap of solid-state chemistry and solid-state physics is extensive. However, the perspectives of the two are sufficiently different that they can easily be identified. In general, solid-state physics treats properties, such as energy and entropy, which are continuously variable in the solid, whereas solid-state chemistry concerns those properties which are discontinuous because of chemical reactions; in a sense, by definition the properties are chemical because they are discontinuous. In another perspective,

solid-state chemistry tends to be based on structure in configuration space, whereas solid-state physics tends to be based on momentum space. *See* INORGANIC CHEMISTRY; SOLID-STATE PHYSICS.

Solid-state chemistry has no single unifying theoretical base and tends to be largely an experimental science supported by several theoretical bases. Consequently, its separation into topics is not well established. Those which are considered to be aspects of solid-state chemistry include, among others, chemical binding, crystal defects, crystal structures, crystal field theory, diffusion in solids, ionic crystals, lattice vibrations, and nonstoichiometry.

As with some other subdivisions of chemistry, the syntheses of new, unusual compounds and phases is one of the primary activities. In solid-state chemistry these compounds and phases have complex and varied structures with large unit cells; they are frequently composed of several different atoms and large integral numbers of atoms, and they tend to be nonstoichiometric phases of variable composition. One of the longer-ranged objectives is to classify these structures and to quantify the multiplicities of chemical bonds involved.

The techniques and procedures used to synthesize the compounds and phases are varied. They include (1) direct combination of the elements to produce an isolated compound or combination of compounds at high temperatures, high pressures, or both; (2) preparation from a solution or crystallization from molten compounds of the same composition or through the use of a molten flux; (3) chemical vapor transport from the reactants at one temperature to a product at another temperature; (4) growth of a single crystal at the interface of two gels, each containing one of the reactants or mixtures of such; (5) arc melting of a mechanical mixture that is sufficiently conductive to maintain an arc; (6) intercalation or deintercalation to produce metastable phases; and (7) implosive, thermomechanical impaction with reaction to produce a dense product. *See* VAPOR DEPOSITION.

Structures

Studies of structures provide a basis for understanding the chemical bonding in solids, their properties, reactions among them, and their variabilities of composition.

Spatial geometries. The spatial arrangement to form a stable compound of atoms or ions in a solid is determined by a minimum in the potential energy of the possible assemblages. For the most part, the factors involved are the atomic or ionic radii, the relative strengths of the bondings, and the directionality of the bonds. One of the objectives of solid-state chemistry is to develop a basis through which the minimum in the potential energy can be predicted for a given collection of atoms or ions. However, the problem is so complex that it has not been solved. Hence, the immediate objective is to obtain sufficient information so that more complete theories can be developed.

The placement of atoms in space is describable in terms of a unit cell. This cell is the smallest part of the crystal that contains all the symmetries of the structure. These are defined at four levels:

1. There are seven possible crystal systems defined by the shape of the unit cell or for a perfect crystal by its geometric shape. They are cubic, tetragonal, orthorhombic, monoclinic, hexagonal, triclinic, and trigonal.

2. To describe the lattices in these shapes, subdivisions of some of these are required. There are three cubic lattices, two tetragonal, four orthorhombic, two monoclinic, and one each of the hexagonal, triclinic, and trigonal.

3. A further subdivision is defined in terms of the rational and reflective invariance about a given lattice position. These consist of the 32 point groups.

4. A final classification is achieved by combining translation with rotation and reflection. These symmetry operations yield 230 space groups.

These idealized structures exist, strictly stated, only at 0 K (absolute zero). At any finite temperatures, these structures are imperfect or distorted in at least two general respects: the atoms or ions vibrate with amplitudes which increase with increasing temperature; and the lattice sites of the space groups become vacant or interstitial positions become occupied, or both occur, with the valencies of the cations or anions becoming altered. *See* ABSOLUTE ZERO; CRYSTAL DEFECTS; CRYSTAL STRUCTURE; LATTICE VIBRATIONS; VALENCE.

The concentration and spatial distribution of the defects in the solid produce various kinds of distortions of the idealized structures. When the defects interact only weakly and do not aggregate, the distortions are mostly localized near the point defects, which may be either randomly distributed or ordered on a long range. For this distribution, the x-ray or neutron diffraction pattern is essentially that of the idealized structures with a charged internuclear distance or lattice parameter and with additional diffraction maxima caused by the long-range order, that is, superstructure. This configuration of point defects occurs especially in several nonstoichiometric phases. In some cases the defects appear to be aggregated, with the result that a structured distortion occurs. Such a situation exists in cases known as shear structures that occur in tungsten trioxide (WO_3) and molybdenum trioxide (MoO_3), where extensive regions appear as an idealized structure connected by a plane in the lattice in which a displacement equivalent to a shear translation exists. *See* NONSTOICHIOMETRIC COMPOUNDS.

Layered structures. Numerous binary compounds or phases have layered structures that in themselves or in combination with other elements can be considered as extensions of the descriptions given above. The layers are displaced relative to each other in such a way that the structures can be classified in the space groups, but when other elements or compounds are incorporated between the layers, the structure is highly distorted. The most widely recognized material to possess the layered structure and the associated solid-state chemistry is graphite. Carbon atoms in this structure are arranged in hexagons connected along the edges, and thus form a plane throughout

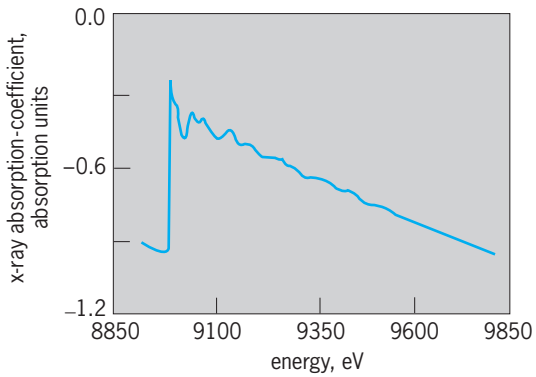


Fig. 1. X-ray absorption spectrum at 20 K (−420°F) for copper in $Ba_4Ca_3Sr_3Cu_4O_x$ obtained with x-radiation near the CuK x-ray edge. The extended x-ray absorption fine structure (EXAFS) occurs after the sharp increase. (*After P. P. Lottici, G. Antonioli, and F. Licci, T-dependence of EXAFS in the Bi-Ca-Sr-Cu-O system, Physics C, 152:468–474, 1988*)

the crystal. These planes are stacked in layers, with each layer displaced with respect to its adjacent layers by a distance such that the structure still belongs to one of the space groups. The bonding within a layer is covalent; between the layers it is metallic. Consequently, graphite is highly anisotropic. When it is exposed to bromine, alkali metals, and other chemical substances, these substances, at high temperatures, enter the spaces between the layers to form intercalation compounds; the idealized structure thus becomes highly distorted. A large number of compounds with layered structures are known to form a variety of intercalation compounds. Among them are titanium disulfide (TiS_2) and molybdenum disulfide (MoS_2). Two others, which are similar in that additional elements or compounds are located between lattice planes, are the tungsten bronzes and the β-aluminas. *See* GRAPHITE; INTERCALATION COMPOUNDS.

Multicomponent phases. For three or more components, the structural situation is generally complex. The number of possible compounds and phases is extremely large, and so is the variety of structures and distortions. Classification of the distorted space groups is difficult and is still being developed. The actual structures are generally described as distortions of some well-recognized structures, which in themselves are complex. Because many of these complex structures were found in minerals, the distorted structures are frequently classified as being derived from these minerals. Examples of such are spinels, perovskites, bixbyites, feldspars, and garnets.

Glasses. All of the structures described above are based on a lattice with defects but with sufficient long-range order to produce well-defined diffraction patterns. There exists another group or class of solids in which the long-range order is insufficient to yield such patterns. These are the glasses or amorphous solids that generally exist only in metastable conditions. The most readily known among these are the silicate glasses, but there are many others, including some even in metallic systems. *See* GLASS; METALLIC GLASSES.

Molecular solids. In all the cases described, there occur no decisive tendencies toward molecular formation in the solids. However, when the intramolecular forces are great enough, as in organic molecules, the solids are composed of molecular crystals. These solids in general have high volatilities and insufficient defects to cause reactions associated with the transfer of mass in the solids, but they have electronic structures which participate in photochemical reactions in organic solids.

Local and extended structures. All the structural information described above is represented by the location of atoms on the lattice network. The structures are determined by diffraction of x-rays, neutrons, or electrons. Two other features of the structure of solids are the local structure about a given atom and the extended structure on a more global scale. The first is determined by the absorption of x-radiation having a wavelength near that of the characteristic x-rays of a given atom (resonance absorption). The spectrum on the high-energy side of the absorption edge (that is, the sudden increase in **Fig. 1**) has fine structure extending beyond the edge and determined by the distance and number of neighbors in the successive shells surrounding the given, central atom [extended x-ray fine structure (EXAFS)]. The global structure is determined by high-resolution electron microscopy (HREM) that produces lattice imaging (**Fig. 2**). Regions of different lattice structures can frequently be identified, such as the contiguous intergrowth of microdomains in a matrix of the primary lattice, a situation that apparently exists in a nonstoichiometric phase that evolves as the composition is varied between two diphasic regions. *See* EXTENDED X-RAY ABSORPTION FINE STRUCTURE (EXAFS); ELECTRON MICROSCOPE.

Electronic structures. Aside from combinations near the center of the periodic system where covalency is dominant, the elements exist in solids as ions. For the regular elements, the outer, valence

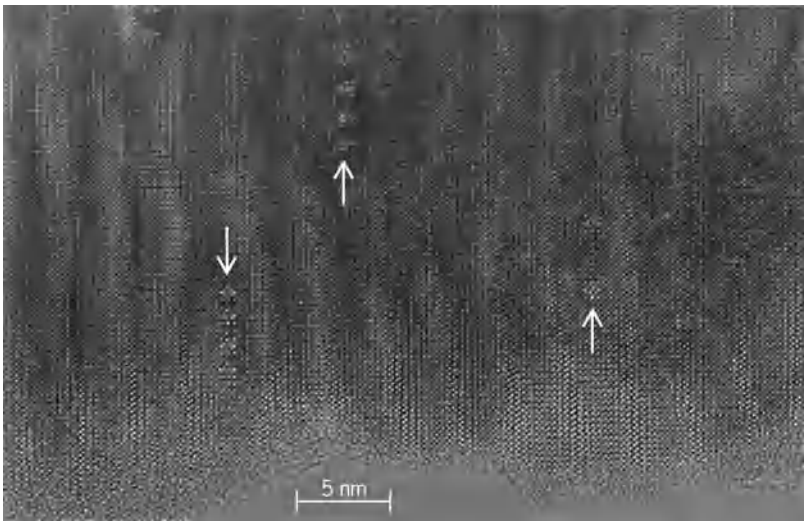


Fig. 2. High-resolution electron microscopic (HREM) lattice image of $Bi_{15}Ti_9Fe_{11}O_{57}$, an Aurivillius phase. Small defect clusters (microdomains) in the center of the perovskite slab are indicated by arrows. (*From M. Lundberg, J. L. Hutchison, and D. J. Smith, The structure of $Bi_{15}Ti_9Fe_{11}O_{57}$ and related compounds derived by high-resolution electron microscopy, J. Sol. State Chem., 80:178–188, 1989*)

orbitals of the ions are closed, with the inert gas shells having *s*-orbital configurations on the cations and hybridized *p*-orbitals on the anion. In both cases, the ions have filled orbitals and tend to be spherical; they are not highly directional. These ions have no low-lying states, and the crystal fields are insufficient to decouple the spins. Consequently, the electronic structures of solids formed with these ions are simple, and they contribute a basis for solid-state chemistry in terms of ionicity and normal polarization. These solids, which are essentially those formed from groups 1, 2, 14, and 17 in the periodic system, are transparent in the visible and near-infrared. In addition, some of the cations of the transition, lanthanide, and actinide groups have closed *p*-orbitals with no low-lying electronic states. These are the scandium and titanium subgroups (thorium is included in the latter). All the other transition, lanthanide, and actinide ions in solids display electronic transitions in their infrared and visible spectra. A large number of the observed transitions are attributable to the fact that the d^n or f^n degenerate states are separated by the electrostatic field of the crystal. In the cases of the lanthanide and actinide ions, some of the observed maxima are also attributable to *f*-to-*d* transitions. The presence of all of these states contributes to anomalous polarization or ionicity and a complexity of the chemistry. *See* ELECTRON CONFIGURATION; PERIODIC TABLE.

Photoacoustic spectroscopy with powdered materials can be used to study the spectra of such solids. Such studies furnish microscopic information for the solid-state chemistry associated with charge transfer, polarization, and valence states. *See* PHOTOACOUSTIC SPECTROSCOPY.

A feature of the electronic structure of solids, which is particularly essential to solid-state chemistry, is the valence state of the ion and the energy required to change the valency in the solid. For the transition, lanthanide, and actinide ions with open *d*- or *f*-orbitals, these energies are sufficiently small that transitions from one valency to another can be promoted by thermal (phonon) activation or the oxidation potentials of anions. The relative concentrations of two valencies in solids can be determined chemically through the stoichiometry or through chemical reactions which presumably do not change the relative concentrations. The known stoichiometries in such solids as the iron oxides Fe_3O_4 and Fe_2O_3 and the uranium oxide U_3O_8 imply mixed valence states. If these phases are dissolved under nonoxidizing conditions, then the initial ratio of the two valence states can be determined. The question of the existence of two distinct valence states in the solid has been debated for decades. Some have argued that a resonance condition exists so that two well-defined oxidation states cannot be identified. Most of these arguments are based on the results of x-ray diffraction observations, which give no evidence of two different cation sites. Two techniques have been developed that enable direct observation of the valence states in the solid, photoelectron spectroscopy and Mössbauer spectroscopy. *See* ELECTRON SPECTROSCOPY; MÖSSBAUER EFFECT; VALENCE; X-RAY DIFFRACTION.

Chemical Bonding

The structure of a solid is the result of the operation of interatomic or interionic forces and the size and shape of the atoms or ions. Hence, logically, bonding should be described first and structure second. However, the detailed role of the electrons in interionic forces is so complex, and the quantitative aspects of the problem of minimizing the potential energy with respect to all the possible configurations is so difficult, that structures cannot be derived. Rather, it is necessary to derive some information about bonding from structures, cohesive energies, refractive indices, electron binding energies, polarizabilities, and other properties through the use of models. Because of the number of ions and electrons involved, the wave-mechanical formulation of bonding in solids is extremely difficult to solve. Fortunately, some rather simple, classically based models modified by quantum-mechanical concepts have been and continue to be quite useful. These can be classified generally as ionic, covalent, and metallic bonding and combinations of the three. *See* CHEMICAL BONDING; STRUCTURAL CHEMISTRY.

Ionic bonding. In the ionic model, the solid is composed of positive (cations) and negative (anions) charges located on the respective lattice sites. These imagined point charges create through their collective coulombic interactions a net attractive force. The self-potential of the interionic forces can be calculated with high precision by summing the electrostatic energies. For crystals with high symmetries, this potential is the product of the ratio of a constant, the Madelung constant, to a common internuclear distance and the sum, characteristic for each structure, of terms whose numerators contain the point charges and whose denominators contain some multiple of the common internuclear distance. For crystals of lower symmetries, more sophisticated summations involving Patterson functions are required. The repulsive forces to balance the attractive forces at equilibrium consist essentially of the overlap of the valence electron orbitals. Additional attractive forces are frequently included in the model. These are occasionally referred to as the dispersive forces because they are the ones related to the dipole and multipole moments in ionic crystals and their indices of refraction. Solids having the most ionic character (ionicity) are those formed from combinations of the most electropositive with the most electronegative elements. In general, the ionicity increases with increasing cationic radius and with decreasing anionic radius. *See* IONIC CRYSTALS; X-RAY CRYSTALLOGRAPHY.

Covalent bonding. In the covalent model, the electrons in the bond are shared equally between the atomic cores, and the electrons involved are in bound orbitals, frequently referred to as molecular orbitals. In essence, the sources of the attractive and repulsive forces are electrostatic, but they are formulated wave-mechanically. In purely covalent

bonding, the atomic cores are identical, as in solids of the inert gases, halogens, oxygen, and sulfur.

Metallic bonding. In the metallic model, the solid is composed of cations on lattice sites surrounded by a uniform negative charge of the conduction or free electrons.

Combination bonding. In all real solids composed of different elements, the bonding consists of admixtures of these models. The bonding in ioniclike crystals composed of the regular metallic elements and the more electronegative elements can be rationalized as admixtures of ionic and covalent bonding. The degree of ionicity or covalency can be evaluated through optical dispersion theories of crystals, wherein the ionicity is evaluated from the index of refraction, the density of valence electrons, and some measure of the separation of the bonding and antibonding orbitals. In these formalisms, the group 14 elements carbon, silicon, and so on, are used as the reference for covalency. Thus covalency increases as the pairs of elements in the solid become more nearly the same as group 14 elements. In the case of the transition, lanthanide, and actinide elements in valence states having open shells, rationalization of the bonding on the basis of this description is inadequate. The crystal field of the solid removes the degeneracy of the orbitals, and the electrons in the bond can occupy states which increase the strength of the chemical bond. Thus there occurs a crystal field stabilization and the attendant anomalous dispersion accompanying the absorption in the infrared and visible spectra. Low-lying *f*-to-*d* transitions may also contribute to the increased bonding. In a sense, the role of these crystal field splittings and *d* states is to increase the polarizabilities of the cations.

In metals formed by higher-valent cations, starting at least with 3+, the anionic-forming elements can react to produce compounds and phases which still contain free electrons. Thus solids such as subhalides of scandium, the monosulfides and carbides of several elements, and some oxides such as titanium oxide (TiO) or vanadium oxide (VO) contain cations, anions, and conduction electrons or a bandgap sufficiently small to be *n*-type semiconductors. The roles of the valence states of the cations and the oxidation of the free electrons are illustrated through the variation of the conductivities and lattice parameters of the lanthanide monosulfides and monoselenides. All of these are conductors except those of europium, ytterbium, and samarium. The rationalization of this behavior is that in the metals the cations are in the 3+ valency in all cases except europium, ytterbium, and samarium, in which the valency is 2+. Thus the anionic elements, sulfur and selenium, are reduced by the free electrons to the 2− state so that the three cases with 2+ cations have no free electrons, whereas each of the others has approximately one free electron. The cationic radii and the oxidation-reduction potentials are consistent with this description.

Measurement. One of the more direct ways of measuring the bonding in a solid, and particularly one that measures it at the cationic and anionic sites separately, is by measuring the electron binding energies by photoelectron spectroscopy. Thus, through the ejection of electrons from valence orbitals with x-rays or ultraviolet radiation $h\nu$ and the measurement of the electrons' kinetic energy, their binding energies are determined by reaction (1). In this reaction *cs* represents a cation site; the asterisk indicates that the cation site is unrelaxed after the ionization; and the electron is in the free gaseous (*g*) state. Solid-state physicists generally refer binding energies to electrons in the Fermi level. In reaction (1) the reference is the free electron; the two differ by the instrumental work function. This technique permits determination of the valence state, $q+$, and the lattice self-potential from which measures of polarization, ionicity, and so on are derived. The lattice self-potential is derived through a comparison of the binding energy with the ionization potential of the gaseous cation.

$$M^{q+}cs \xrightarrow{h\nu} M^{(q+1)+}cs^* + e(g) \qquad (1)$$

Chemical Composition

In a broad sense, the chemical composition of a solid can be classified into one of two groups. The first class comprises compositions in which the ratios of atoms are integers, albeit large in some cases; the structures of these are described in terms of the space groups. This situation is essentially the law of definite proportions, and the compounds are frequently referred to as daltonides. The second class comprises compositions that are continuously variable throughout a phase, because either one atom or ion can be substituted for another of comparable radius (usually known as solid solutions) or atoms or ions can be accommodated through the occupation of interstitial or created vacant sites. A change in valency of either the cation or the anion or both occurs as the composition is varied. This group is frequently referred to as berthollides. *See* DEFINITE COMPOSITION, LAW OF; SOLID SOLUTION.

At finite temperatures, and particularly at high temperatures, the partial vapor pressures of the components in a solid are different. Consequently, in general there occurs a preferential loss of one component so that any solid at equilibrium tends to contain lattice defects and to become nonstoichiometric. However, in a large number of cases the deviations from stoichiometry are not detectable, so that the number of nonstoichiometric phases that can be studied in solid-state chemistry is not extremely large. Among those which have been studied extensively are $Fe_{1-x}O$, CeO_{2-x}, UO_{2+x}, $YBa_2Cu_3O_{7-\delta}$, and various metal hydrides, as well as the other oxides, carbides, and hydrides.

Whenever the structure of a solid is such that an interstitial position of one of the sublattices can be occupied, both interstitials and vacancies occur. However, over the compositional range of nonstoichiometry, one or the other of the defects is usually in the higher concentration. Thus UO_2 has a fluorite structure such that the position at the center of the

unit cell can accommodate an oxygen ion with an attendant shift in the oxidation state of the uranium ions. In UO_{2+x} the defects are predominantly oxygen ions on interstitial sites and U^{5+} ions on some of the cation sites. In $UO_{2-x'}$ which exists at temperatures near 2000°C (3600°F), vacancies on the oxygen sublattice are at higher concentration. In a phase such as $Th_yU_{1-y}O_{2+x'}$ thorium ions are substitutional, and in $Pu_yU_{1-x}O_{2\pm x}$ plutonium is substitutional. In the last case, because plutonium has the 3+ and 4+ oxidation states accessible and uranium has the 4+ and 5+ states, the anion composition has an extensive range on both sides of stoichiometry. In metallic carbides, the carbon is often interstitial, and in UC_{1+x} the carbon tends to be incorporated in the UC lattice as a C_2 unit.

The role of the compositional variable in oxidation at the electronic level in the solid is illustrated in the x-ray photoelectron spectra of the valence band region of uranium at four stages of oxidation from metal to UO_3. For uranium metal and dioxide, the intensities of the 5*f* orbital electrons are the same; no change occurs in the number of 5*f* electrons. Hence, the oxidation of the metal to UO_2 involves only the conversion of the free electrons to bound electrons in the hybridized 2*p* orbital on the oxygen ion. When UO_2 is oxidized to U_4O_9, however, the intensity of the 5*f* part decreases and that of the 2*p* increases. With further oxidation to U_3O_8, the 5*f* intensities decrease further and the 2*p* intensity increases. Finally, in UO_3, the 5*f* electrons are completely oxidized to 2*p* electrons on the oxygen ion.

One of the measurements, which furnishes information needed to understand the compositional variable in solid-state chemistry, is the observation of partial vapor pressures, and their dependence on temperatures is determined. The chemical potentials and the partial molar enthalpies and entropies and their variation with composition are derived from these measurements. For a binary system, the chemical potential increases monotonically in the single-phase region, and is constant in the diphasic region. The construction and mathematical evaluation of statistical models to describe the chemical potential throughout both regions from the interionic forces, valence states, and defect energies contain conceptual problems associated with discontinuities at the diphasic region. Models can be constructed and evaluated to describe the homogeneous regions. One model is described by Eq. (2), in which θ_i is

$$\mu = \mu_0 + RT\left[\ln\left(\frac{\theta_i}{1-\theta_i}\right) - E_i - 2\theta_i E_{ii}\right] \qquad (2)$$

the fraction of interstitial sites occupied, E_i is the energy required to remove an ion from the interstitial site, and E_{ii} is the energy of interaction between two occupied interstitial sites. Above some critical temperature ($T_c = E_{ii}/2R$), this function has a sigmoidal shape with no maximum or minimum. Thus for $T > T_c$, the variation of μ with θ_i (composition) (**Fig. 3**) represents a monophasic region. At the critical temperature, the isotherm has a zero slope. Below this critical temperature, the function has a maximum, a minimum, and an inflection between them (as shown by the isotherm for $T < T_c$ in Fig. 3). Because the slope of μ versus composition in a real system cannot change sign, this behavior is unrealistic. The artifice that is introduced to excuse these van der Waals loops is to construct a horizontal line through the inflection point and use this to represent the diphasic region. A realistic model that describes the evolution to a diphasic region would be one in which the horizontal line is contained in the mathematics. In one technique that has been suggested, the chemical potential is described through the complex (mathematical) variable. In some cases (praesodymium oxides), the compositional isotherm displays hysteresis loops because of the defect complexes that form differently in the two directions of μ versus composition.

When a substantial nonstoichiometric phase is formed between an insulator and a metallic conductor, the electrical conductivity and concentration of free electrons change continuously. However, in some cases and perhaps in all cases, the change is so rapid over a small compositional range that the change can be viewed as an insulator-to-metal transition. One system that displays this behavior is samarium neodymium selenide ($Sm_{1-x}Nd_xSe$). SmSe

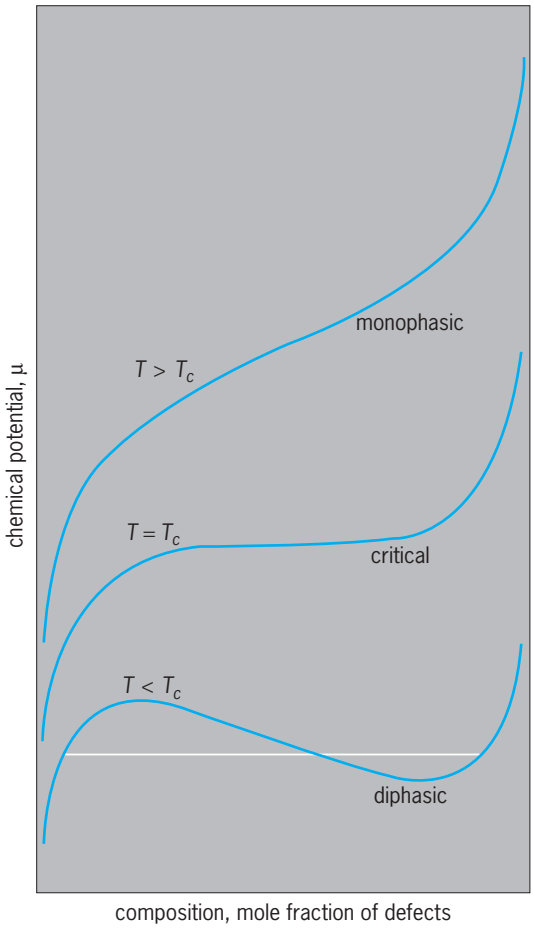


Fig. 3. Three isotherms for the variations of chemical potential with composition of defects.

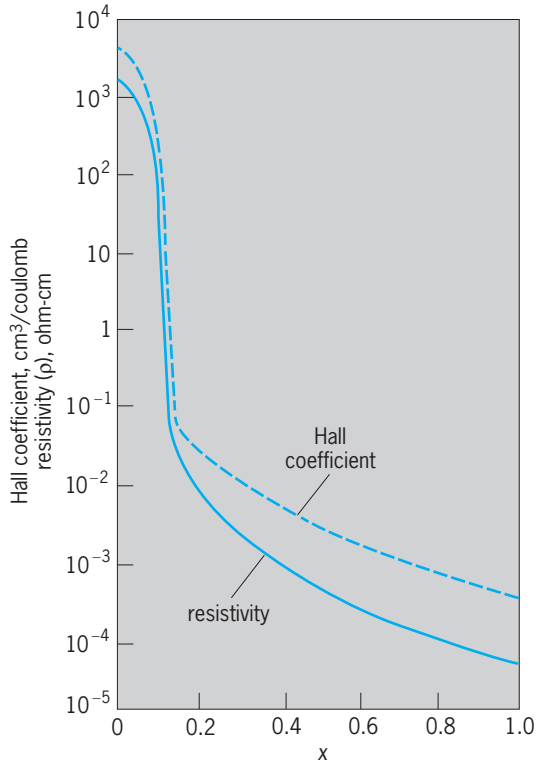


Fig. 4. Electrical properties versus composition in *n*-type $Sm_{1-x}Nd_xSe$, where *x* represents the variation in stoichiometry. The value of the Hall coefficient at $x = 1$ is calculated with the assumption of one electron per neodymium atom. Temperature = 300 K (80°F). (*After F. J. Ried et al., Electrical conduction in rare-earth monoselenides and monotellurides and their alloys, J. Phys. Chem. Sol., 25:969–976, 1964*)

is an insulator with Sm^{2+} and S^{2-}, and NdSe is a conductor with Nd^{3+}, S^{2-}, and one free electron per unit. As the composition is varied, the conductivity and concentration of free electrons change significantly, but at $x \cong 0.1$ the two change by five orders of magnitude (**Fig. 4**).

One of the essential features of the superconducting oxides is nonstoichiometry. The extent varies, but all appear to be phases of variable composition. One of the early ones to be discovered is the barium-lead-bismuth oxide $BaPb_{1-x}Bi_xO_3$, which is a single phase from $BaPbO_3$, a metalliclike conductor to $BaBiO_3$, a semiconductor. Both lead and bismuth have dual valences. The oxide that stimulated great interest in superconducting oxides is $La_{2-x}Ba_xO_{4-\delta}$ (at 35 K or −397°F), $0 \leq x \leq 0.15$ and δ small. The one having the largest range of oxygen composition is $YBa_2Cu_3O_{7-\delta}$ (at 92 K or −294°F), where Y = yttrium, $6 \leq \delta \leq 7$, and the formal oxidation states of copper (Cu) are from 1+ to 3+. Several others were found in a relatively short time. Intense research on these superconducting oxides awakened significant interest in solid-state chemistry.

The oxide $YBa_2Cu_3O_{7-\delta}$ has a perovskite structure (**Fig. 5**). The central part of this structure, which can be written as $(YCu_3O_6Ba_2)$, appears to play a rather minor role in determining the essential properties. The yttrium can be replaced by several lanthanide elements, especially in the second half of the series where the structures resemble those for yttrium. Substitution of holmium or gadolinium, with large magnetic moments, for yttrium has little if any effect. Barium that is adjacent to the basal plane where the chemical entity can be written as $(CuO_{1-\delta})^+$ appears to play an indirect role; its replacement by strontium does affect the properties; it decreases the temperature, but the primary variabilities are attributable to the oxygen content in the basal plane. For $\delta = 0$, the structure is orthorhombic; for $\delta = 1$, it is tetragonal. The transition from orthorhombic to tetragonal is second-order or order-disorder. The diamagnetic susceptibility varies sigmoidally with decreasing temperature during the transition to the superconducting state from a value near zero to a value near $-1/(4\pi)$ for the optimum situation (**Fig. 6**). As the oxygen content decreases from a value near 6.9, the asymptotically attained value of the susceptibility increases. *See* SUPERCONDUCTIVITY.

Chemical Reactions

The mechanisms of chemical reactions within and between solids are through lattice vibrations, lattice defects, and changes in valence states. These are the structural features through which migration of mass, charge, and energy occur. Consequently, diffusion and conductivity are integral, basic parts of solid-state chemistry, even though their quantitative roles in the totality of chemical reactions have not been developed. So long as the solid phase produced during reaction has a density nearly the same as those for the reactants, the microscopic description of the

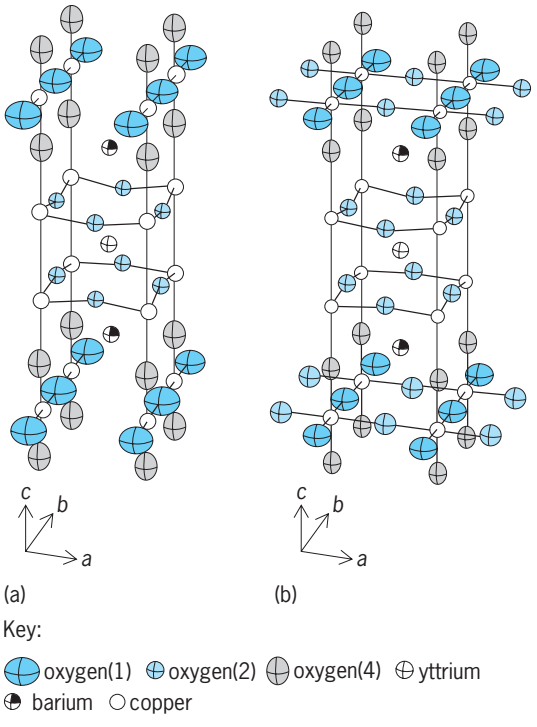


Fig. 5. Perovskite structures $YBa_2Cu_3O_{7-\delta}$ of (*a*) orthorhombic and (*b*) tetragonal. In the tetragonal structure the ellipsoidal symbols for the O(1) site is used to indicate that this site is not fully occupied. (*After J. D. Jorgensen et al., Structural properties of oxygen-deficient $YBa_2Cu_3O_{7-\delta}$, Phys. Rev. B, 41:1863–1877, 1990*)

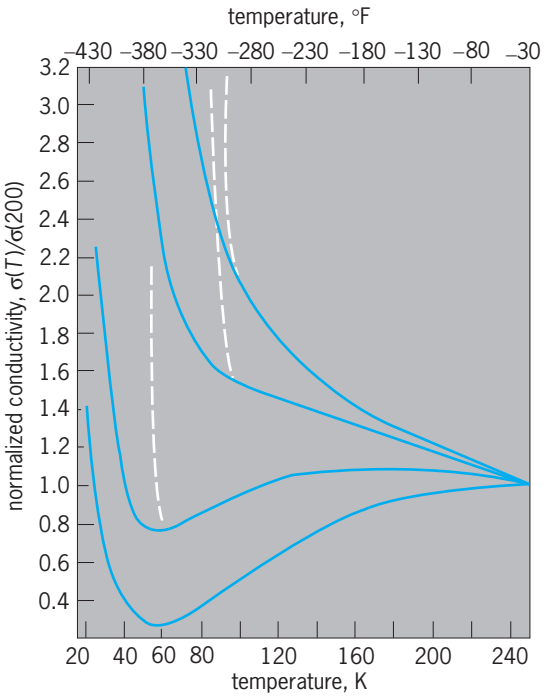


Fig. 6. Normalized electrical conductivity versus temperature for various concentrations of oxygen in $YBa_2Cu_3O_{7-\delta}$. Oxygen content decreases from the topmost curve to the lowest. (*After R. J. Thorn and C. E. Thorn, Semiconductive and semimetallic components in superconducting oxides, J. Phys. Chem. Sol., 50:153–161, 1989*)

reaction in terms of mass and charge transfer is feasible. However, in a number of reactions the molar volumes differ sufficiently to destroy the integrity of the product, so that much of the reaction then proceeds at interfaces, microcracks, and fissures. Thus the total net mechanism becomes a composite process so complex that a comprehensive theory is lacking. The theories that have been developed are generally discussed under topics such as corrosion. Thus, no formal classification of reactions in solids is universally recognized, or, in fact even cited, in the literature. *See* DIFFUSION; ELECTRICAL CONDUCTIVITY OF METALS.

Condensation. Although it is well recognized that condensation to form solids occurs when the translational energy of the gaseous reactants is dissipated into vibrational and electronic states, the processes involved are sufficiently complex that they are difficult to study and to describe in detail. In general, two kinds of condensation need to be described. In a supersaturated vapor, condensation is a stepwise process in which molecules form in successively larger clusters if the rate of growth of the clusters exceeds the rate of dissociation. At some critical size, the process results in condensation, because the translational energy can be dissipated into the cluster. This process occurs in shock tubes and jets which produce super-saturated vapors and in hydrodynamically flowing gaseous beams in which adiabatic expansion occurs. Another kind of condensation occurs if atoms or molecules in the gas are sufficiently accommodated thermally on a colder substrate. In this case, the growth occurs primarily through the flow of translational energy into the substrate. However, the atoms or molecules bound to the surface can be mobile on the surface so that growth can occur through a stepwise process therein. Usually the process is nonequilibrious, and an amorphous, metastable product is formed. For instance, gaseous silicon monoxide can be condensed on a cold substrate to produce solid, amorphous, metastable silicon monoxide. This is the only known way of producing it. Condensation to produce epitaxial layers on substrates is used to produce various kinds of solid-state devices.

An elementary condensation which can be used to illustrate the processes involved and which represents what could be called the first step in a solid-state chemistry is the formation of a metal. Because the process is one in which the neutral gaseous atom is converted to a collection of cations and free electrons, it involves internal oxidation and reduction, represented by reaction (3). Thus, in this reaction

$$M(---np^6) \rightarrow M^{q+}(---np^{(6-q)})e_q \qquad (3)$$

the gaseous atom with the electronic configuration of the core electrons (- - -) and the outer six *np* electrons condenses to form something resembling a cation of charge $q+$ associated with q free electrons. In the supersaturated vapor, the critical step probably is the one in which the cluster is large enough to have free or conduction electrons to serve to dissipate the translation energy. When gaseous metal-forming atoms impinge on a metal substrate, the free electrons in the latter serve to transfer the energy, so that the accommodation and condensation coefficients of metal on metal are generally unity. On nonmetallic insulators, however, the coefficients are not unity, and condensation occurs through nucleation on the surface. The significance of recognizing that the condensation to form a metal is a redox process is contained in what is involved upon further oxidation. When the solid metal undergoes further reaction with an anion-forming element, it is the free electrons which are used first, and the cation is subsequently oxidized. *See* CRYSTAL GROWTH; OXIDATION-REDUCTION.

Internal processes. One of the simplest reactions that occurs in solids is a phase transition. These transitions may be either first- or second-order. In the first case, a change occurs in the crystal structure, sometimes with the higher-temperature form having an apparently higher symmetry. Second-order transitions are usually order-disorder transitions such as the λ-type transitions which occur on only one of the sublattices. *See* PHASE TRANSITIONS.

The primary internal reactions which involve a chemical reaction are those associated with site occupancy, changes in valency, and clustering. Because it has been extensively studied, uranium dioxide (UO_2) serves as a useful illustration. In this phase, the interstitial position (*is*) of the fluorite lattice can accommodate an oxygen ion. Thus, an equilibrium exists between the occupancy of this site and the

regular anion sites (*as*), as shown in reaction (4).

$$O^{2-}as + is \rightarrow O^{2-}is + as \quad (4)$$

Whenever uranium dioxide is oxidized, the valence of uranium is increased from 4+ to 5+ so that reaction (5) is the one occurring primarily. Reactions

$$2U^{4+}cs + O(g) + is \rightarrow 2U^{5+}cs + O^{2-}is \quad (5)$$

similar to these are responsible for the existence of nonstoichiometric phases. The dual valency on the cation also is involved in hole conductivity in *p*-type semiconductors.

Another nonstoichiometric oxide that has been extensively studied is $YBa_2Cu_7O_{7-\delta}$ which, for δ approximately equal to 0.1, is a superconductor. In this phase the internal equilibrium is Eq. (6). In

$$Cu^{3+}cs + O^{2-}as = Cu^{2+} + O^{1-}as \quad (6)$$

the range of homogeneity $O \leq \delta \leq 1$, the structure changes from orthorhombic to tetragonal as oxygen is added in the basal plane (Fig. 5). For intermediate compositions, the neutron diffraction patterns are analyzed in terms of fractional occupancies of the sites in the basal plane, so the real unit cells are much larger. Electron diffraction reveals superstructure and, consistently, high-resolution electron microscopy (HREM) reveals that the phase can be represented as microdomains (such as in Fig. 2) in a matrix, so that the phase can be represented as $(YBa_2Cu_3O_7)_{1-\delta}$ $(YBa_2Cu_3O_6)_\delta$. Photoelectron spectroscopy reveals that the energy for this reaction is small; it occurs at low temperature. *See* CHEMICAL EQUILIBRIUM.

A comparison of the energies associated with these reactions in solids and in gases illustrates the differences in the two cases. In the gas phase, the O^{2-} ion is unstable; in the solids, it is stabilized by the lattice self-potential. The energy associated with reaction (6) in the gas phase equals the difference between the two ionization potentials. Usually, this difference is of the order of tens of electronvolts. However, in the solid, this difference is compensated by the difference in the lattice self-potentials. Consequently, the energy associated with reaction (6), in the solid phase, is only a few electronvolts. *See* HOLE STATES IN SOLIDS; SEMICONDUCTOR.

Interphase reactions. Reactions between phases accompanied by product formation with integrity maintained occur via chemical diffusion. This situation occurs when the reactants and the product have the same structure and comparable ionic radii. But even in this case, there generally is a volume change, the concentration of defects changes, and the initial interface between the reactant moves. A classical example is the reaction (that is, diffusion) between copper and copper-zinc alloy. The zinc diffuses more rapidly out of the alloy than the copper diffuses in, and the original interface moves toward the alloy. This phenomenon is known as the Kirkendall effect; it is one of the evidences that diffusion occurs via vacancies and interstitials rather than through positional exchange. *See* ALLOY STRUCTURES; INTERFACE OF PHASES; PHASE EQUILIBRIUM.

Two well-studied solid-state chemical reactions are sintering and corrosion. In the first reaction, mass is transferred between particles in a solid so that densification and frequently plastic deformation or creep occurs. The details of the transfer between particles is not well understood, but self-diffusion is generally involved. In corrosion, diffusion through the product layer generally controls the rate of reaction, although electromigration may also be significant. *See* CREEP (MATERIALS); PLASTIC DEFORMATION OF METAL.

In both cases, the mechanism can involve either the migration of a reacting component from the substrate through the layer or the migration of a corroding component from the external surface. When the layer becomes sufficiently thick and its density is sufficiently different, the layer develops fissures and cracks. Then the mechanism occurs along these, but the primary chemical driving force may still be the same as before. *See* CORROSION; SINTERING.

The oxides and other phases of the transition and lanthanide elements are catalysts for heterogeneous reactions at a gas–solid interface. It seems plausible that the lattice defects or the *d* and *f* electronic configurations serve as sites for adsorption of a gaseous reactant so that an energy barrier for the reaction is reduced. *See* CATALYSIS; HETEROGENEOUS CATALYSIS.

Electrochemical reactions. Electrolytic cells with solid electrodes and electrolytes have been used to determine chemical potentials, especially in nonstoichiometric phases. An example is a cell composed of nonstoichiometric oxides, UO_{2+x} and $Fe–Fe_{1-y}O$, as electrodes and stabilized zirconia as the electrolyte: $Fe–Fe_{1-y}O|ZrO_2|UO_{2+x}$. The electrolyte is an oxygen ion conductor, so that the voltage developed is produced by the difference in the chemical potential of oxygen in the two solid electrodes. Rechargeable batteries in which either the electrolyte or the electrodes are solids have been investigated. *See* ELECTROCHEMISTRY.

Photochemical reactions. Many of the photoinduced processes which occur in solids can be imagined to be solid-state chemical reactions. The classical ones, of course, are those involved in photographic plates and films. *See* PHOTOCHEMISTRY.

Sublimation. At sufficiently high temperatures, the rate of sublimation of all solids can be measured. The partial vapor pressures of most of the inorganic fluorides, chlorides, and oxides and several of the carbides and sulfides have been determined. In a number of cases, the vapor contains complex molecular species, even though the temperature required is high. The presence of these molecules is, in a sense, a consequence of the relatively high vapor pressure and not the high temperature, which of course at the constant pressure is a degradative factor. The chemical potentials and the partial molar enthalpies and entropies of sublimation are derived from the partial vapor pressures and their dependencies on temperature. These are direct measures of the bonding in

the solid. *See* HIGH-TEMPERATURE CHEMISTRY; SUBLIMATION.

R. J. Thorn

Bibliography. D. M. Adams, *Inorganic Solids: An Introduction to Concepts in Solid State Structural Chemistry*, 1974; P. A. Cox, *The Electronic Structure and Chemistry of Solids*, 1987; W. Hayes and A. M. Stoneham, *Defects and Defect Processes in Nonmettalic Solids*, 1985; S. Holt et al. (eds.), *Solid State Chemistry: A Contemporary Overview*, 1980; W. Jost (ed.), *Physical Chemistry*, vol. 10: *Solid State*, 1970; P. Kofstad, *Nonstoichiometry, Diffusion, and Electrical Conductivity in Binary Metal Oxides*, 1972, reprint 1983; G. G. Libowitz, *The Solid State Chemistry of Binary Metal Hydrides*, 1965; *J. Phys. Chem. Solids*, monthly; *J. Solid State Chem.*, monthly; D. L. Nelson and T. F. George (eds.), *Chemistry of High Temperature Superconductors II*, 1988; D. L. Nelson, M. S. Wittingham, and T. F. George (eds.), *Chemistry of High Temperature Superconductors*, 1987; H. Schmalzried, *Solid State Reactions*, 2d ed., 1981; O. T. Sorensen, *Nonstoichiometric Oxides*, 1981; A. R. West, *Solid State Chemistry and Its Applications*, 1987.

Solid-state physics

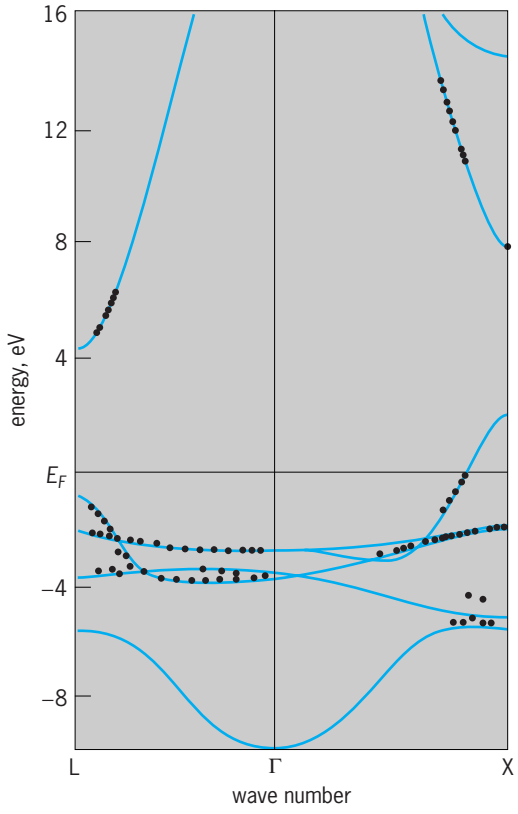
The study of the physical properties of solids, such as electrical, dielectric, elastic, and thermal properties, and their understanding in terms of fundamental physical laws. Most problems in solid-state physics would be called solid-state chemistry if studied by scientists with chemical training, and vice versa. Solid-state physics emphasizes the properties common to large classes of compounds rather than the dependence of properties upon compositions, the latter receiving greater emphasis in solid-state chemistry. In addition, solid-state chemistry tends to be more descriptive, while solid-state physics focuses upon quantitative relationships between properties and the underlying electronic structure. *See* SOLID-STATE CHEMISTRY.

Many of the scientists who study the physics of liquids identify with solid-state physics, and the term "condensed-matter physics" is increasingly replacing "solid-state physics" as a division of physics. It includes noncrystalline solids such as glass as well as crystalline solids. *See* AMORPHOUS SOLID; GLASS.

Electronic structure of solids. In solid-state physics it is generally assumed that the electronic states can be described as wavelike. The individual electronic states, called Bloch states, have energies which depend on the wave number (a vector equal to the momentum divided by $\hbar$, where $\hbar$ is Planck's constant divided by 2π), and the wave number is restricted to a domain called the Brillouin zone. This energy given as a function of wave number is called the band structure. It is shown for two lines (Γ to X and Γ to L) in the Brillouin zone for metallic copper in the **illustration**. There are several curves, called bands, for each line. *See* BRILLOUIN ZONE.

These bands can be calculated by using quantum theory; the results of such a calculation are given as the continuous lines in the illustration. It is also possible to measure them directly by using the photoelectric effect, as indicated by the points in the illustration. In any solid the electrons occupy the lowest-energy electronic states, one electron of each of the two allowed spin states in each electronic state, or two electrons per atom in each band in a simple structure. In copper there are 11 electrons per atom to occupy the bands shown, and they fill the states up to the Fermi energy E_F, indicated in the illustration. One of the bands is partly occupied (that is, it crosses the energy E_F), so that this solid is a metal; under an applied field or temperature gradient, the electrons shift from state to state, causing currents to flow and making copper a good conductor of electricity and heat and making it shiny. Light at the blue end of the spectrum can also lift electrons from the nearly horizontal bands to the empty states; the absorption of the blue light gives the reddish color to copper. *See* CRYSTAL ABSORPTION SPECTRA; FREE-ELECTRON THEORY OF METALS; METAL; NONRELATIVISTIC QUANTUM THEORY; PHOTOEMISSION.

If copper had an additional electron per atom (as does zinc), there would be just enough electrons to fill the lower bands (the bottom five curves). The illustration suggests that there would then be no partially filled bands, and thus an energy gap separating the full bands from the empty ones. Actually, in copper and in zinc some of the upper bands dip below the peak energy of the lower bands, as would be the



Energy bands of copper. (*After J. A. Knapp et al., Experimental energy band dispersions and lifetimes for valence and conduction bands of copper using angle-resolved photoemission, Phys. Rev., B19:4952–4964, 1979*)

case if the upper band at L dropped below 2 eV in the illustration. Thus copper would still be a metal, as is zinc, even if it had an additional electron.

When such energy gaps between occupied and empty bands do occur, as in silicon, the solid is no longer metallic. Applied fields only shift electrons from state to state within individual bands, and since the same states remain occupied, there is no current generated by, and no heat flow due to, the electrons. If the gaps are small, as is the 1.13-eV gap of silicon, a small number of electrons are thermally shifted at room temperature to the empty bands, called conduction bands, and electrical conductivity is introduced; it can also be accomplished by substituting impurities, called dopants. Phosphorus, for example, has one more electron than silicon and contributes that electron to the conduction band. Such systems in which conductivity can be introduced are called semiconductors. When the energy gaps are large, as is the 8.5-eV gap for rock salt, there are ordinarily no empty states in the lower bands and no electrons in the conduction bands, and the crystals are insulators. *See* BAND THEORY OF SOLIDS; ELECTRIC INSULATOR; SEMICONDUCTOR.

Cooperative phenomena. Each state in the copper bands shown in the illustration can accommodate two electrons of opposite spin. The two spin states have the same energy, so that they are both filled if the energy is less than E_F, or both empty if the energy is greater than E_F; there is no net electronic spin in the system and no magnetism. In nickel, less than copper in atomic number by one, the energies of the two spin states shift from each other. Electrons with spin of one sign become more numerous, and the metal becomes ferromagnetic; each of the bands shown in the illustration has become split into two. Such a transition is called a cooperative phenomenon since it is caused by the mutual lowering of each other's energies by electrons of the same spin. The theoretical and experimental study of such cooperative transitions is an active area of solid-state physics research. *See* FERROMAGNETISM; PHASE TRANSITIONS.

Superconductivity. Another cooperative phenomenon is superconductivity, in which pairs of electrons of opposite spin and moving in opposite directions form Cooper pairs that condense into a cooperative electronic state. There is again an energy gap between the occupied Cooper pair states and the empty states from broken pairs, but in this case the entire cooperative state can drift to carry a current without exciting individual electrons from the cooperative state. Such a flow is called supercurrent since it flows without any electrical resistance whatsoever; it occurs only in some metals and until 1986 it was observed only at very low temperatures.

In 1986 J. G. Bednorz and K. A. Müller discovered a copper oxide compound ($Ba_xLa_{2-x}CuO_4$, containing a small ratio x of barium atoms to copper atoms) that was superconducting at 30 K (−406°F). Soon, other copper oxide compounds were found ($YBa_2Cu_3O_{7-x}$) that were superconducting at 90 K (−298°F), well above the temperature of liquid nitrogen. The development of these so-called high-temperature superconductors may prove very important since it is much easier to cool materials with nitrogen than with the helium required for other superconductors. In spite of the enormous experimental and theoretical study of these materials, the origin of this new kind of superconductivity is not yet understood. *See* SUPERCONDUCTIVITY.

Small supercurrents can even flow through very thin insulators by quantum-mechanical tunneling, an effect predicted by B. D. Josephson. It was predicted and observed after the discovery by I. Giaever that individual electrons can tunnel through thin oxides; Giaever used this tunneling to make a direct measurement of the energy gap arising from the Cooper pairs in superconductors. *See* JOSEPHSON EFFECT; TUNNELING IN SOLIDS.

Crystal structure and lattice vibrations. The total energy of a solid includes a sum of the energies of the occupied electronic states. Since the energy bands depend upon the positions of the atoms, so does the total energy, and the stable crystal structure is that which minimizes this energy. The theory has proven adequate to predict the crystal structure of various solids. For some years it has been possible to predict the changes in energy under various distortions of the lattice. There are in fact three times as many independent distortions, called normal modes, as there are atoms in the solid. Each has a wave number, and the frequencies of the normal vibrational modes, as a function of wave number in the Brillouin zone, form vibrational bands in direct analogy with the electronic energy bands shown in the illustration. These can be calculated directly from quantum theory or measured by using neutron or x-ray diffraction. *See* CRYSTAL STRUCTURE; LATTICE VIBRATIONS; NEUTRON DIFFRACTION; X-RAY DIFFRACTION.

In quantum theory any normal mode can be excited only to discrete quantized vibrational energies. In solids a mode in the nth state of excitation is said to contain n phonons. These phonons interact with the electrons and cause the reduction in conductivity with increasing temperature in metals. This same electron-phonon interaction also provides the attraction that holds electrons together in Cooper pairs to form the cooperative superconducting state. *See* ELECTRICAL CONDUCTIVITY OF METALS; ELECTRICAL RESISTIVITY; PHONON.

Intrinsic and extrinsic properties. Solid-state physics includes the study of all of these properties of ideal crystals, called intrinsic properties. It also includes the study of defects in the structure, impurities, and surfaces, all of which are called extrinsic properties. The study of extrinsic properties can be made with very much the same experimental tools, and they can be understood by using quantum theory with very much the same concepts which are useful for studying intrinsic properties. *See* CRYSTAL DEFECTS.

Walter A. Harrison

Bibliography. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, 2d ed., 2000; H. Ehrenreich et al. (eds.), *Solid State Physics*, vols. 1–48, 1955–1994; W. A. Harrison, *Electronic Structure and the Properties of Solids*, 1980; C. Kittel, *Introduction to Solid*

State Physics, 7th ed., 1996; S. A. Wolf and V. Z. Kresin (eds.), *Novel Superconductivity*, 1987.

Solids pump

A device used to move solids upward through a chamber or conduit. It is able to overcome the large dynamic forces at the base of a solids bed and cause the entire bed to move upward.

Solids pumps are used to cause motion of solids in process-type equipment in which treatment of solids under special conditions of temperature, oxidation, and reduction can be combined with upward motion and discharge of the spent solids overhead from the reacting vessel.

Solids pumps are inherently of the positive displacement type. One practical method uses a reciprocating piston mounted on a trunnion, permitting it to swing into an inclined position for filling and then to swing back into vertical position for discharge. **Figure 1** shows a mechanically driven solids pump in four positions through its cycle of operation. **Figure 2** shows a large hydraulically operated pump used in units of capacity exceeding 1000 tons/day (900 metric tons/day). Hydraulic activation permits very precise control of the feeder mechanism and good efficiency in operation.

The solids pump has found its principal application in the operation of oil-shale retorts. Here it is used to feed crushed shale into the bottom of a conical vessel, and as the shale moves upward through this vessel, air is drawn downward countercurrent.

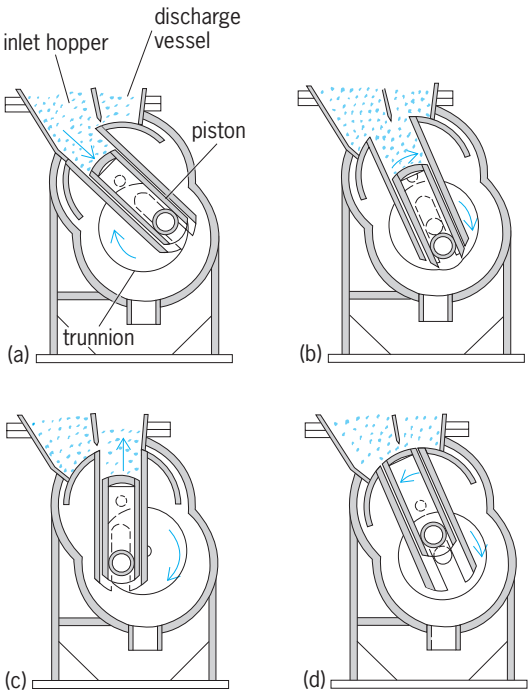


Fig. 1. Operation cycle of mechanically driven solids pump. (*a*) Filling with solids from inlet hopper. (*b*) Piston rotating on a trunnion toward its discharge position. (*c*) Discharge position, with piston pushing charge of solids upward. (*d*) Piston rotating back toward original filling position.

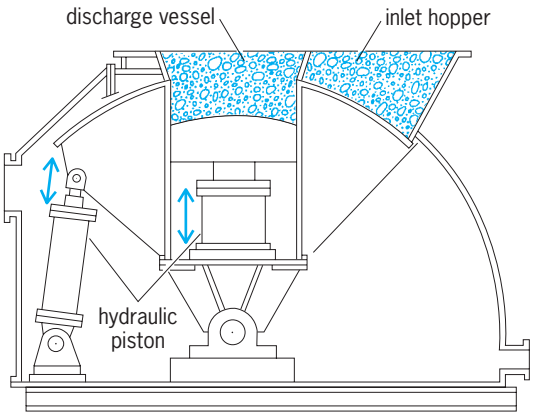


Fig. 2. Hydraulically operated solids pump.

At the top of the retort, the air burns the residual carbon on the shale ash. The hot flue gas so produced contacts shale in the midpoint of the reactor, educting the shale oil. These vapors together with the flue gas are cooled, and the oil condensed on the shale at the bottom of the retort. The oil flows out the bottom countercurrent to the upgoing bed of shale. *See* BULK-HANDLING MACHINES; PUMP; UNIT OPERATIONS.

Clyde Berg

Solifugae

An order of nonvenomous, spiderlike, predatory arachnids found chiefly in arid and semiarid, tropical, and warm-temperate regions worldwide. Solifugids can range in size from a few centimeters to almost 3 in. (7.5 cm). Most of the 1100 known species are nocturnal. They are commonly known as sun spiders, camel spiders, or wind scorpions (see **illustration**). The last name stems from the fact that they exhibit the most rapid sprint speeds for any known group of terrestrial arthropods.

Solifugids possess relatively large anterior appendages, known as chelicerae, that are used for grasping and crushing prey. Ounce for ounce, these jaws produce one of the most powerful bites in the animal kingdom. The second pair of appendages, the leglike and nonchelate palpi, are tactile in

Wind scorpion. (*Glenn and Martha Vargas © California Academy of Sciences*)

function and end in a structure (palpal sucker) that is adhesive in function and helps in procuring prey and mating. The first pair of legs is tactile, whereas the other three are ambulatory.

Most species have a life span ranging from 1 to 3 years. Life cycle stages include the egg, several nymphal instars, and adult. After laying her eggs, the female guards them within the nest until they hatch, and remains with the newly hatched nymphs for several days. Depending on their size, solifugids feed on a variety of arthropods and small vertebrates. These arachnids are most frequently found within burrows or rock crevices, or under rocks, logs, or surface debris. Many are excellent climbers, and some tropical species are arboreal. *See* ARACHNIDA; ARTHROPODA.

Fred Punzo; C. Clayton Hoff

Bibliography. M. Moffett, Wind scorpions, *Nat. Geog.*, 206:94–101, July 2004; F. Punzo, *The Biology of Camel-Spiders (Arachnida, Solifugae)*, Kluwer Academic, Norwell, MA, 1998; F. Punzo, *Desert Arthropods: Life History Variations*, Springer, New York, 2000.

Soliton

An isolated wave that propagates without dispersing its energy over larger and larger regions of space. In most of the scientific literature, the requirement that two solitons emerge unchanged from a collision is also added to the definition; otherwise the disturbance is termed a solitary wave.

There are many equations of mathematical physics which have solutions of the soliton type. Correspondingly, the phenomena which they describe, be it the motion of waves in shallow water or in an ionized plasma, exhibit solitons. The first observation of this kind of wave was made in 1834 by John Scott Russell, who followed on horseback a soliton propagating in the windings of a channel. Scott Russell's account was published in 1845; 50 years later D. J. Korteweg and H. de Vries proposed an equation for the motion of waves in shallow waters which possesses soliton solutions, and thus established a mathematical basis for the study of the phenomenon. Interest in the subject, however, lay dormant for many years, and the major body of investigations began only in the 1950s. Researches done by analytical methods and by numerical methods made possible with the advent of computers gradually led to a complete understanding of solitons. *See* DIFFERENTIAL EQUATION; WAVE MOTION; WAVE MOTION IN LIQUIDS.

Eventually, the fact that solitons exhibit particle-like properties, because the energy is at any instant confined to a limited region of space, received attention, and solitons were proposed as models for elementary particles. However, it is difficult to account for all of the properties of known particles in terms of solitons. More recently it has been realized that some of the quantum fields which are used to describe particles and their interactions also have solutions of the soliton type. The solitons would then appear as additional particles, and may have escaped experimental detection because their masses are much larger than those of known particles. In this context the requirement that solitons emerge unchanged from a collision has been found too restrictive, and particle theorists have used the term soliton where traditionally the term solitary wave would be used. *See* ELEMENTARY PARTICLE; QUANTUM FIELD THEORY.

Sin ϕ or sine-Gordon model. The sin ϕ or sine-Gordon model is formulated in terms of a field $\phi(x,t)$ depending on one spatial coordinate x and the time variable t, and describes one-dimensional propagation of waves. The energy density, defined by Eq. (1), leads to Eq. (2) as the equation of motion,

$$\mathscr{E} = \frac{\alpha}{2}\left(\frac{\partial\phi}{\partial t}\right)^2 + \frac{\beta}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 + \frac{\gamma}{2}\left(1 - \cos\frac{2\pi\phi}{\phi_0}\right) \tag{1}$$

$$\alpha\frac{\partial^2\phi}{\partial t^2} - \beta\frac{\partial^2\phi}{\partial x^2} = -\frac{\pi y}{\phi_0}\sin\frac{2\pi\phi}{\phi_0} \tag{2}$$

which is one of the simplest equations having soliton solutions. Here α, β, γ, and ϕ_0 are positive constant parameters. The first term in the energy density is sensitive to the rate of variation in time of the field at a definite location and becomes larger as this rate increases. The second term depends on the rate of variation of the field along the x-coordinate axis and also becomes larger as this rate of variation increases. Thus, the first two terms in the energy will be minimal (and indeed will equal zero) if the field is constant in space and time. The third term depends on the actual value of the field. When $\phi = 0, \pm\phi_0, \pm 2\phi_0$, or any integer multiple of ϕ_0, $\cos(2\pi\phi/\phi_0)$ [argument expressed in radians; otherwise 2π is replaced by 360°] becomes equal to 1 and the third term in the energy vanishes. When $\phi = \pm(\phi_0/2), \pm(3\phi_0/2), \pm(5\phi_0/2), \ldots$, the third term in the energy is maximal and equal to γ, and for all other values of ϕ it varies between 0 and γ. The state of minimum energy, or state of rest or vacuum state, is therefore achieved when the field takes a value constant in space and time and equal to $0, \pm\phi_0, \pm 2\phi_0, \ldots$. Thus there is not a single vacuum, but many states of minimal energy (multiple vacua).

Solitons occur when the field approaches different vacuum values in the two opposite directions of the x-coordinate axis. Simple and intuitive reasoning shows that if different vacua are reached as $x \to +\infty$ and $x \to -\infty$, a region of transition must exist where the energy density is nonvanishing. If these two vacuum values differ precisely by ϕ_0, the system exhibits a single soliton. The field profile, constant in time, which minimizes the energy gives rise to the static soliton solution (**Fig. 1**). The energy is thus confined to a definite region of space.

A more gradual passage of ϕ between the two vacuum states (which would spread out the energy) would lead to an increase of the contribution of the last term to the energy not compensated by the decrease in the second term. A more abrupt transition

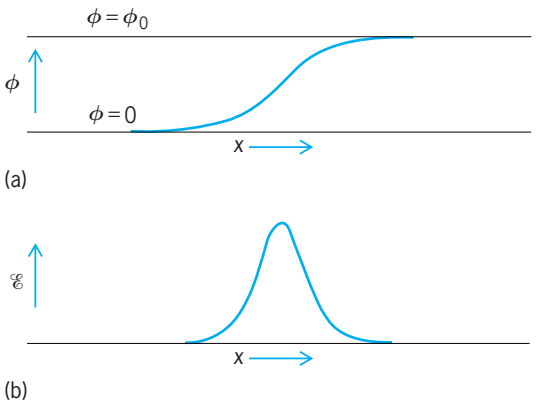


Fig. 1. Static soliton solution of the sin ϕ mode. (*a*) Field ϕ. (*b*) Energy density◁.

would also increase the total energy. Configurations in which the field undergoes more than one transition between neighboring vacuum states exhibit multiple solitons.

Solutions with moving solitons also exist. In an encounter of two solitons moving in opposite directions (**Fig. 2**), the two waves emerge from the collision with shape and speed unchanged. This remarkable feature of the solitary waves results from special properties of the sin ϕ equation (and of other equations having soliton solutions) and, as has already been mentioned, is often considered part of the definition of a soliton.

Other systems with solitons. Many equations describing one-dimensional propagation of waves give origin to soliton solutions. The equation of Korteweg and de Vries, the nonlinear Schrödinger equation, and the Boussinesq equation belong to this category. The preservation of shape and velocity of the waves after a collision has been related to the existence of conservation laws. Any isolated dynamical system obeys conservation laws: the total energy of the system and its total momentum, for instance, do not change in time. In a perfectly elastic frontal collision of two billiard balls of the same mass, these two conservation laws are sufficient to predict the outcome of the collision: the two balls just exchange their speeds. The systems exhibiting solitons are characterized by obeying a very large number of conservation laws, well beyond what would be expected of an ordinary dynamical system. The existence of these conservation laws implies, and in turn follows from, the fact that the system is integrable; by rather elaborate analytical techniques, a new set of variables may be found for the description of the motion which evolve in time independently and according to very simple rules. *See* CONSERVATION LAWS (PHYSICS); CONSERVATION OF ENERGY; CONSERVATION OF MOMENTUM; SYMMETRY LAWS (PHYSICS).

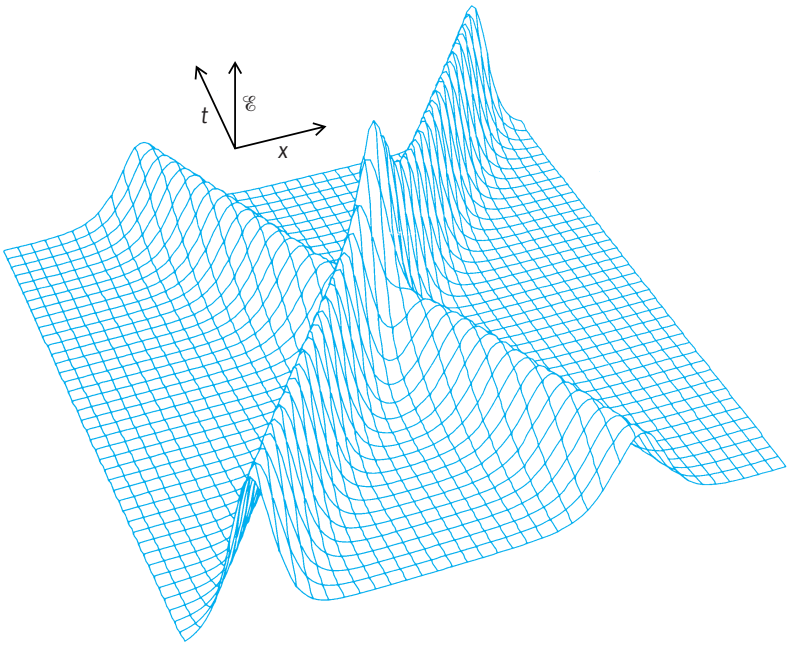


Fig. 2. Space and time variation of the energy density in a collision between two solitons moving in opposite directions. The wave with the higher peak energy is moving faster. The two waves emerge from the collision with shape and speed unchanged.

Derivation of solutions. Equations with soliton solutions have been widely studied; various mathematical methods have been devised to investigate their properties and to derive their solutions. In the method of inverse scattering transforms, an auxiliary problem is set up, which consists of studying the diffusion of a set of plane waves by a potential well. The shape of the potential well varies with time and is related to the profile of the solitons. From the initial soliton configuration, the scattering data, that is, the parameters that totally characterize the diffusion process, can be determined. These scattering data evolve in time according to some simple, soluble equations. Thus, to obtain the soliton profile at a later time, the evolution of the scattering data is followed, and eventually the shape of the diffusing potential is reconstructed from them. It is this last step, where the normal process of deriving the scattering data from a potential is inverted to determine the latter from the former, that gives the name to the method. Other mathematical techniques rely on special sets of transformations, called Bäcklund transformations, which convert given soliton solutions into other soliton solutions, frequently exhibiting a different number of solitons; or rely on embedding the original equations into larger sets of partial differential equations (prolongation schemes), which possess special algebraic symmetries. *See* INVERSE SCATTERING THEORY.

Two- and three-dimensional systems. While one-dimensional solitons are very interesting for a variety of applications, an isolated wave must be three-dimensional to be interpreted as a particle. The existence of systems exhibiting solitary waves in two and three spatial dimensions has been demonstrated. These confined waves are not expected to preserve their dynamical state in a collision (although at present it cannot be excluded that some have this property), but particle theorists have called them solitons as well. In some cases, the confinement of the wave can be related to the existence of multiple states of minimum energy (multiple vacua), as for the sin ϕ model; such waves are called topological solitons because of the special geometrical

properties of the arrangement of vacua around the wave. The topological features are not universal, however, and nontopological solitons exist as well.

Quantum phenomena. A quantum-mechanical formulation of the dynamics of the solitons has been developed, and it has been discovered that topological solitons are often associated with formerly unexpected, remarkable quantum phenomena. Some of the topological solitons are magnetic monopoles; that is, they carry an isolated pole of magnetic charge. Magnetic monopoles cannot be introduced in the ordinary theory of electromagnetism without allowing for some type of singularity in the fields, but no singularity is present in the soliton-monopole solutions, which exhibit everywhere regular field configurations. Topological solitons may violate the normal relation between the statistics of the fields and the spin of the particles. In quantum field theories, fields are classified as bosonic or fermionic, according to the algebraic properties of their quantization formulas. The elementary quantum excitations of bosonic fields are then particles with integer spin (in units of the fundamental constant $\hbar$), such as the photon and π-meson, whereas quantum excitations of fermionic fields are particles with half-integer spin, such as the electron, proton, and neutron. Yet, theories can be devised where all the fields are bosonic but where solitons behave as particles of half-integer spin. Finally, it has also been demonstrated that in some theories with topological solitons a phenomenon of fermion fractionalization occurs. Fermionic particles, which are normally indivisible units, can find themselves split in the presence of a soliton-antisoliton pair, and each soliton itself or each antisoliton carries a fraction of a fermion. In the domain of particle physics, such phenomena have only a theoretical basis, since solitons, as particles, have not been experimentally observed. However, certain solitons have found experimental verification in condensed matter systems. There is, for instance, evidence of fermion fractionalization in connection with the solitons that occur in long, dimerized molecular chains. *See* MAGNETIC MONOPOLES; QUANTUM STATISTICS; SPIN (QUANTUM MECHANICS).

Massive particles. Many field theories used to describe particles have solutions of the soliton type. These solutions, properly quantized, should not be identified with the known atomic, nuclear, or subnuclear particles, which are rather associated with elementary quantum-mechanical excitations of the fields or bound states of such excitations. The solitons, if indeed realized as particles in nature, would have much larger masses, to the point of making their production in present or conceivable future experiments impossible. They might have been formed, however, in the extreme conditions of temperature and energy density in the very early stages of the evolution of the universe, and their experimental detection would have remarkable implications for particle physics and cosmology. *See* COSMOLOGY.

Claudio Rebbi

Hydrodynamic solitons. A hydrodynamic soliton is simply described by the equation of Korteweg and de Vries, which includes a dispersive term and a term to represent nonlinear effects. Easily observed in a wave tank, a bell-shaped solution of this equation balances the effects of dispersion and nonlinearity, and it is this balance that is the essential feature of the soliton phenomenon.

Among the largest and most energetic solitons are those that appear in hydrodynamic systems. Tidal waves in the Firth of Forth were found by Scott Russell to be solitons, as are internal ocean waves and tsunamis. At an even greater level of energy, it has been suggested that the Great Red Spot of the planet Jupiter is a hydrodynamic soliton. *See* JUPITER; OCEAN WAVES; TSUNAMI.

Optical solitons. The most significant technical application of the soliton is as a carrier of digital information along an optical fiber. First observed in 1980, the optical soliton is governed by the nonlinear Schrödinger equation, and again expresses a balance between the effects of optical dispersion and nonlinearity that is due to electric field dependence of the refractive index in the fiber core. If the power is too low, nonlinear effects become negligible, and the information spreads (or disperses) over an ever increasing length of the fiber. At a pulse power level of about 5 milliwatts, however, a robustly stable soliton appears and maintains its size and shape in the presence of disturbing influences.

Because of unavoidable optical losses, it is necessary to amplify the pulse every 25 mi (40 km) or so. This requirement leads to the most important advantage of solitonic data transmission: the soliton retains its standard shape under optical amplification (using distributed Raman amplifiers). A linear pulse, however, cannot be optically reshaped. It must first be converted into an electronic signal for reashaping and amplification and then reconverted into a light pulse for transmission over the next 25-mi (40-km) leg of its journey, greatly reducing the bandwidth of the system. Present designs for data transmission systems based on the optical soliton have a total length of 4500 mi (7300 km) and a data rate of 4×10^9 bits per second at an error rate of 10^{-9}, which is the telecommunications industry standard. *See* OPTICAL COMMUNICATIONS.

Fluxons on long Josephson junctions. A carefully studied soliton system is the transverse electromagnetic (TEM) wave that travels between two strips of superconducting metal separated by an insulating layer thin enough (about 2.5 nanometers) to permit transverse Josephson tunneling (**Fig. 3**). In this case the relevant nonlinear wave equation is the sine-Gordon model given in Eq. (2) with added terms to represent bias current and electrical losses. Since each soliton carries one quantum of magnetic flux, it is also called a fluxon if the magnetic flux points in one direction, and an antifluxon if the flux points in the opposite direction.

A bias current flowing through the Josephson junction (Fig. 3) forces a fluxon to move transversely in one direction and an antifluxon in the other direction. Because a moving fluxon is reflected from the end of a junction waveguide as an antifluxon, steady bias current accelerates the fluxon (or antifluxon) until its speed becomes limited by electrical losses.

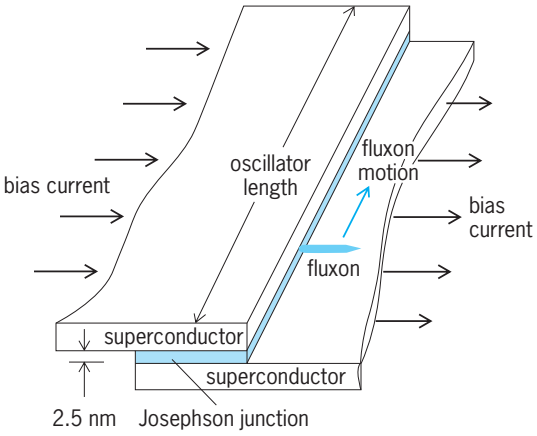


Fig. 3. Fluxon oscillator. Steady bias current flows down through the long Josephson junction and forces a fluxon to move in one direction and an antifluxon to move in the other direction. The fluxon is reflected from the end of the structure as an antifluxon, so the fluxon-antifluxon oscillates back and forth in the long direction.

(The limiting velocity of electromagnetic waves on the waveguide is about $1/\sqrt{20}$ the speed of light.) At this point, the system is a fluxon oscillator with its frequency under the control of the bias current. Since the oscillator length can be less than 100 micrometers, and since the maximum frequency is the limiting velocity divided by twice this length, such oscillators reach into the submillimeter wave region of the electromagnetic spectrum (frequencies greater than 10^{11} Hz). *See* JOSEPHSON EFFECT; WAVEGUIDE.

Biological solitary waves. The all-or-nothing action potential or nerve impulse that carries a bit of biological information along the axon of a nerve cell shares many properties with the soliton. Both are solutions of nonlinear equations that travel with fixed shape at constant speed, but the soliton conserves energy, while the nerve impulse balances the rate at which electrostatic energy is released from the nerve membrane to the rate at which it is consumed by the dissipative effects of circulating ionic currents. The nerve process is much like the flame of a candle. *See* BIOPOTENTIALS AND IONIC CURRENTS.

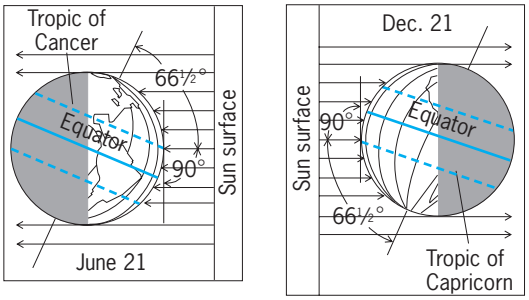
Theories for true (energy-conserving) solitons have been proposed for a variety of elongated biomolecular structures. These structures include the alpha-helix regions of protein molecules, where solitons may participate in the phenomenon of muscular contraction, carry electronic and protonic charge by polaronic mechanisms, or transport ions across semipermeable membranes; deoxyribonucleic acid (DNA), where solitons may help to open and close the double helix during transcription of the genetic code; and protein crystals of the cytoskeleton, where the solitons may serve as a basis for information processing in single-celled organisms such as the ameba, paramecium, and vorticella. Each of these biological solitons is closely related to one of the lattice solitons that have been observed in condensed matter, and describe the propagation of dislocations, ferroelectric and ferromagnetic domain walls, polarons, charge-density waves, and orientations of chemical bonds in a variety of crystalline structures. *See* CHARGE-DENSITY WAVE; CRYSTAL DEFECTS; CYTOSKELETON; DEOXYRIBONUCLEIC ACID (DNA); DOMAIN (ELECTRICITY AND MAGNETISM); POLARON; PROTEIN.

Alwyn Scott

Bibliography. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect*, 1982; G. Chaohao (ed.), *Soliton Theory and Its Applications*, 1995; A. S. Davydov, *Solitons in Molecular Systems*, 2d ed., 1991; G. L. Lamb, Jr., *Elements of Soliton Theory*, 1980, reprint 1994; A. C. Newell and J. V. Moloney, *Nonlinear Optics*, 1991; C. Rebbi, Solitons, *Sci. Amer.*, 240(2):92–116, 1979; C. Rebbi and G. Soliani (eds.), *Solitons and Particles*, 1984; M. Remoissenet, *Waves Called Solitons: Concepts and Experiments*, 3d ed., 1999.

Solstice

The two days during the year when the Earth is so located in its orbit that the inclination (about $23\frac{1}{2}^\circ$, or 23.45°) of the polar axis is toward the Sun. This occurs on June 21, called the summer solstice, when the North Pole is tilted toward the Sun; and on December 22, called the winter solstice, when the South Pole is tilted toward the Sun (see **illus.**). The adjectives summer and winter, used above, refer to the Northern Hemisphere; seasons are reversed in the Southern Hemisphere.



The Earth at the time of the summer and winter solstices. The dates may vary because of the extra one-fourth day in the year.

At the time of the summer solstice the Sun's rays are vertical overhead at the Tropic of Cancer, $23\frac{1}{2}^\circ$ north. At the North Pole the Sun will then circle $23\frac{1}{2}^\circ$ above the horizon; and at the Arctic Circle, $66\frac{1}{2}^\circ$ north, the noon Sun will be 47° above the horizon and the setting Sun will touch the horizon to the north. Thus, on this day every place north of the Arctic Circle will have 24 h of sunlight and the length of day at all places north of the Equator will be more than 12 h, increasing in length with increasing latitude.

Identical conditions are found in the Southern Hemisphere at the time of the Northern Hemisphere's winter solstice when the Sun is vertical above the Tropic of Capricorn, $23\frac{1}{2}^\circ$ south, and the South Pole is tilted toward the Sun. *See* MATHEMATICAL GEOGRAPHY.

Van H. English

Solubility product constant

A special type of simplified equilibrium constant (symbol K_{sp}) defined for, and useful for, equilibria between solids (*s*) and their respective ions in solution, for example, reaction (1). For this relatively

$$AgCl(s) \rightleftharpoons Ag^+ + Cl^- \quad (1)$$

simple equilibrium, Eqs. (2) and (3) apply.

$$[Ag^+][Cl^-] \cong K_{sp} \quad (2)$$

$$(Ag^+)(Cl^-)/(AgCl) = K_{sp} \quad (3)$$

It can be demonstrated experimentally that a small increase in the molar concentration of chloride ion $[Cl^-]$ (produced, for example, by the introduction of NaCl, HCl, or other soluble chloride) causes a reduction in the concentration of silver present as Ag^+. Similarly, an increase in $[Ag^+]$ reduces $[Cl^-]$. The product of the two concentrations is approximately constant as indicated by Eq. (2) and equal to the K_{sp} of Eq. (3). Equation (3) is exact since the variables are activities instead of concentrations. In accordance with the choice of standard state usually made for a solid, the activity of solid AgCl is unity; hence Eq. (4) holds.

$$(Ag^+)(Cl^-) = K_{sp} = 1.8 \times 10^{-10} \text{ mole}^2 \text{ liter}^{-2} \quad (4)$$

In practice, various complications arise: addition of too much of either ion produces more complicated ions and hence actually increases the apparent concentration of the other ion. Addition of a salt without a common ion (that is, a salt supplying neither Ag^+ nor Cl^-) either may react with Ag^+ or Cl^- or may merely increase the concentration of both ions by a lowering of the mean ionic activity coefficient. (Sodium nitrate at a concentration 0.01 molar increases each concentration by about 10% and the product by 20%.)

It is usually assumed that an aqueous solution saturated with silver chloride contains only Ag^+, Cl^-, and the solvent. Some recent work indicates, however, that about 2.5% of the solute is present as undissociated AgCl. In practice, such effects are usually neglected. Equation (2) is especially useful in the explanation of analytical procedures in which it is desired to add a sufficient quantity of one ion to ensure (virtually) complete precipitation of the other.

An example of a salt of a different charge type is lead iodate $Pb(IO_3)_2$. The solubility equilibrium is represented by reaction (5).

$$Pb(IO_3)_2(s) \rightleftharpoons Pb^{2-} + 2IO_3^- \quad (5)$$

At about 77°F (25°C), the solubility of this salt is 0.024 g/liter. The mass of 1 gram-formula weight of the salt is 557 g. In the saturated solution, therefore, the concentrations are as in notation (6). Because

$$[Pb^{2+}] = \frac{0.024}{557} = 4.3 \times 10^{-5} \text{ mole/liter}$$
$$[IO_3^-] = \frac{2(0.024)}{557} = 8.6 \times 10^{-5} \text{ mole/liter} \quad (6)$$

the solid is in its standard state, its activity is unity. Hence Eq. (7) applies.

$$[Pb^{2+}][IO_3^-]^2 = K_{sp} = 3.2 \times 10^{-13} \text{ mole}^3/\text{liter}^3 \quad (7)$$

It is important to note that the concentration, 8.6×10^{-5} mole/liter, is the concentration of IO_3^- ion in the solution; it is not twice the concentration of the IO_3^- ion, although it happens to be twice the concentration of Pb^{2+}. It is also important to observe that the square of 8.6×10^{-5} mole/liter enters this product because the coefficient of IO_3^- in reaction (5) is 2. *See* GRAVIMETRIC ANALYSIS; IONIC EQUILIBRIUM; PRECIPITATION (CHEMISTRY). Thomas F. Young

Solubilizing of samples

The process by which difficultly soluble samples are converted into different chemical compounds which are soluble. The sample may be heated in air to evolve volatile components or to oxidize a component to a volatile higher oxidation state with the formation of an acid-soluble form, as in the roasting of a sulfide to form the oxide and sulfur dioxide. Most frequently, the sample is treated with a solvent which reacts with one or more constituents of the sample. The dissolution of organic materials often follows the generalizations suggested by the solubility-based schemes of qualitative organic analysis.

Solvents. The choice of solvent is determined by the chemical reactions which are required. Reactions used include solvation, neutralization, complex formation, metathesis, displacement, oxidation-reduction, or combinations of these. Most water-soluble salts dissolve by solvation. Basic oxides such as ferric oxide dissolve in aqueous hydrochloric acid by neutralization followed by complex formation. This is shown by reaction (1). Amphoteric oxides

$$Fe_2O_3 + 8H^+ + 8Cl^- \rightarrow 2FeCl_4^- + 2H^+ + 3H_2O \quad (1)$$

dissolve in either acids or bases. Reaction (2) rep-

$$Al_2O_3 + 2Na^+ + 2OH^- \rightarrow 2AlO_2^- + 2Na^+ + H_2O \quad (2)$$

resents the dissolving in a base. Aqueous ammonia converts insoluble silver chloride into a soluble complex as shown by reaction (3). Carbonates are solubi-

$$AgCl + 2NH_4OH \rightarrow Ag(NH_3)_2^+ + Cl^- + 2H_2O \quad (3)$$

lized by treatment with hydrochloric acid to displace carbon dioxide and to form soluble chlorides. Reaction (4) represents this reaction. Many substances are

$$CaCO_3 + 2H^+ + 2Cl^- \rightarrow Ca^{2+} + 2Cl^- + H_2O + CO_2 \quad (4)$$

converted to easily dissolved mixtures by metathesis. An example is the boiling of barium sulfate with aqueous sodium carbonate to form barium carbonate and sodium sulfate. The insoluble barium carbonate dissolves easily in hydrochloric acid.

Insoluble metal salts have been transposed to soluble ethylenediaminetetraacetic acid complexes.

Insoluble phosphates, sulfates, and even silicates have also been transposed by batch digestions with cationic ion-exchange resins as shown by reaction (5).

$$MSO_4 + 2HR \rightarrow MR_2 + H_2SO_4 \qquad (5)$$

The metal which is bound to the resin can then be eluted or displaced into solution by treatment with an excess of an indifferent salt solution or with an acid whose anion is compatible with the cation present.

Dissolution of metals and alloys. Metals above hydrogen in the electrochemical series will dissolve in a nonoxidizing acid by reduction of hydrogen ion. Other metals such as copper and lead require an oxidizing acid, usually nitric acid. The treatment of alloys is determined by the constituents present. All of the components of brass and bronze are usually dissolved by nitric acid except tin, arsenic, and antimony, which precipitate as hydrated oxides. Alloy steels are usually dissolved by combinations of hydrochloric, nitric, phosphoric, and hydrofluoric acids, depending on the elements present. Aluminum-base alloys are treated with sodium hydroxide solution and any residues are dissolved in acid. Either total or component-selective anodic dissolution of metals and alloys may be achieved by applying a suitable anodic potential to the massive sample, weighed before and after attack, which is immersed in an electrolyte bath.

Fusions. Many substances do not dissolve at temperatures obtainable in the presence of liquid water. However, many refractory silicates, strongly ignited oxides of beryllium, cerium, plutonium, and so on, are completely attacked by aqueous hydrochloric acid at 300°C (570°F). A sealed quartz-tube procedure was developed which permits this widely useful type of sample attack.

However, fused salt reactions employing temperatures of 400–1100°C (750–2000°F) are necessary for the attack and decomposition of many types of samples. The material used as the solvent is called a flux, and the process of melting the mixture of dry, solid flux with the sample is called a fusion. Fusion is done in a crucible (usually platinum or nickel) which is not attacked by the flux or the sample constituents. The same types of reactions are used as with aqueous solvents. Sodium carbonate is used to attack acid materials such as silicates and for metathesis reactions with sulfates, as in reaction (6). Potassium bisulfate

$$CaSiO_3 + Na_2CO_3 \rightarrow CaCO_3 + Na_2SiO_3 \qquad (6)$$

on fusion yields potassium pyrosulfate, an acid flux, which attacks basic oxides such as alumina and ferric oxide and metals such as chromium and tungsten. This flux must be used in porcelain crucibles. Oxidizing fluxes such as sodium peroxide and mixtures of sodium carbonate with potassium nitrate are used with sulfides, chromium, and tin ores and some silicon samples. These usually require nickel crucibles. Calcium carbonate plus ammonium chloride is used to free the alkali metals from silicates. *See* ANALYTICAL CHEMISTRY.

Charles L. Rulfs

Bibliography. W. F. Hillebrand and E. F. Lundell, *Applied Inorganic Analysis*: *With Special Reference to the Analysis of Metals*, *Minerals, and Rocks*, 1980; D. A. Skoog, *Analytical Chemistry: An Introduction*, 7th ed., 2001.

Solution

A homogeneous mixture of two or more components whose properties vary continuously with varying proportions of the components. A liquid solution can be distinguished experimentally from a pure liquid by the fact that during transfers into other single phases at equilibrium (freezing and vaporizing at constant pressure) the temperature and other properties vary continuously, whereas those of a pure liquid remain constant. For an apparent exception *see* AZEOTROPIC MIXTURE; SOLVENT.

Gases, unless highly compressed, are mutually soluble in all proportions.

A solid solution is, similarly, a single phase whose composition and other properties vary continuously with changing composition of the liquid phase with which it is in equilibrium. *See* SOLID SOLUTION.

Types of intermolecular force. The extent to which substances can form solutions depends upon the kind and strength of the attractive forces between the several molecular species involved. It is necessary to consider the attractive forces exerted by molecules of the following types: (1) nonpolar molecules; (2) polar molecules, that is, those containing electric dipoles; (3) ions; and (4) metallic atoms.

London forces. The theory of attraction between nonpolar molecules, developed by F. London in 1930, is based upon the quantum-mechanical interaction between pairs of electron systems. For two molecules with electrons having frequencies ν_1 and ν_2, polarizabilities α_1 and α_2, and separated by the distance r between centers, the attraction potential is shown as Eq. (1), where h is the Planck constant. For molecules of the same species, this reduces to Eq. (2). The frequency ν_0 is that corresponding to $h\nu_0$, the zero-point energy, of the molecule in its unperturbed state. The perturbation by another molecule is related to its perturbation by light of varying frequencies, as seen in the variation of refractive index n with the frequency of light, that is, the dispersion. For this reason London designated these forces as dispersion forces. It is equally appropriate to speak of London forces by analogy with the nearly equivalent term, van der Waals forces. In the case of gases the dispersion n_ν is related to the frequency as in

$$\epsilon_{12} = -\frac{3\alpha_1\alpha_2 h}{2r^6} \cdot \frac{\nu_{0,1}\nu_{0,2}}{\nu_{0,1} + \nu_{0,2}} \qquad (1)$$

$$\epsilon_{11} = -\frac{3\alpha_1^2}{4r^6} h\nu_0 \qquad (2)$$

Eq. (3), where C is a constant. The polarizability α

$$n_\nu - 1 = \frac{C}{\nu_0^2 - \nu^2} \tag{3}$$

can be determined from the refractive index with the aid of the Lorentz-Lorenz formula. As a substitute for zero-point energy, London proposed the ionization potential.

The model upon which these relations are based is much simpler than the polyatomic molecules in most solutions of interest. In these the potential field is not central and radial; the interaction is between the electrons in the peripheral bonds. A striking example is octamethylcyclotetrasiloxane, whose core of alternating silicon and oxygen atoms is so buried within the eight methyl groups that it behaves toward other molecules essentially as an aliphatic hydrocarbon. The normal paraffins themselves are not symmetrical spherically. Moreover, the electrons in the molecules have many frequencies.

Although attempts have been made to extend London's basic concept to take account of such complexities, only the general implications of the concept as to the characteristics of the London forces are necessary here. These forces are (1) of very short range; (2) additive and nonspecific; (3) temperature independent; (4) operative between molecules of all types, whether nonpolar, polar, or metallic; (5) dependent in magnitude upon the number and "looseness" of the electrons; and (6) ordinarily less than average between molecules of different species. This last property can be seen by comparing ϵ_{12} with ϵ_{11} and ϵ_{22} as given by Eqs. (1) and (2). Eliminating the α terms one obtains Eq. (4). Most component pairs

$$\epsilon_{12} = \frac{(\nu_1\nu_2)^{1/2}}{(\nu_1 + \nu_2)/2} \cdot (\epsilon_{11}\epsilon_{22})^{1/2} \tag{4}$$

differ much less in ionization potential than in polarizability, and the factor representing frequencies in this expression is not far from unity. Thus, relation (5) can be written. This means that the interaction

$$\epsilon_{12} \approx (\epsilon_{11}\epsilon_{22})^{1/2} \tag{5}$$

potential between unlike molecules is less than the arithmetic mean of the like potentials.

The pair potentials can be integrated over all the molecules in the pure components as well as the solution to obtain approximate attraction constants, a's, corresponding to the attraction constant of the van der Waals equation. M. P. E. Berthelot proposed the relation $a_{12} = (a_{11}a_{22})^{1/2}$ between the attraction constants of like and unlike species. J. Hildebrand and H. M. Carter found the Berthelot relation to be valid within 1% for seven liquid pairs, for example, $a = 31.21$ for liquid CCl_4 and 64.79 for $SnBr_4$. The calculated geometric mean is 46.46, and a_{12}, observed, is 46.86.

The consequence of this geometric mean relation is that the cohesion in a mixture of two liquids having different cohesion is less than their average. This results in expansion in volume, absorption of heat, and vapor pressures greater than additive upon mixing.

This geometric mean relation is usually adhered to rather well in cases of unlike molecules whose outer electrons are of similar types, such as (1) "N-electrons," nonbonded, as in halogenated paraffins and halogens; (2) π electrons, as in olefins and aromatics; (3) bonding electrons only, as in H_2, CH_4, and other aliphatics; and (4) fluorochemicals. But deviations are found between molecules whose outer electrons are of different types. Illustrations will be found below in the sections on regular solutions and on solubility of gases.

Dipole interaction. This is the attraction between molecules containing permanent electric dipoles; it includes both the London forces and an electrostatic interaction of the dipoles. The latter depends upon the dipole moments of the molecules; it is temperature dependent because thermal agitation opposes the antiparallel orientation in which the interaction is greatest. Its magnitude depends also upon the geometry of the molecules because it is related to the distance of approach of the dipoles, not the molecular centers; the dipoles of some molecules are buried more deeply than those of others. This is the case with chloroform, which has solvent properties similar to those of carbon tetrachloride, except in a few specific cases. J. G. Kirkwood has expressed the degree of interaction between the dipoles of pure liquids by a g factor. For pyridine, the dipole moment μ is 2.20×10^{-18} cgs units, the dielectric constant ϵ is 12.5, and the dipole interaction g is 0.9. For water, $\mu = 1.84 \times 10^{18}$, $\epsilon = 78.5$, and $g = 2.7$; for ethyl alcohol, $\mu = 2.80 \times 10^{-18}$, $\epsilon = 24.6$, and $g = 3.0$.

It is the g factor, not the dipole moment or the dielectric constant, that is most significant for understanding solubility relations. Furthermore, in the case of molecules having more than one polar bond, it is the separate polar bonds, not their vector sum of the overall dipole moment, that determine solubility relations. The three isomeric dinitrobenzenes all affect the vapor pressure of benzene virtually to the same extent, even though their dipole moments are quite different.

The substances with the largest g factors are those that form hydrogen bonds. These have exceptionally high boiling points and are poor solvents for nonpolar substances. These liquids resist penetration by nonpolar molecules. The best-known pairs of incompletely miscible liquids are composed of a nonpolar liquid and water.

Electron donor-acceptor interaction. In the modern theory of generalized acids and bases, initiated by G. N. Lewis, a base is a substance having electrons that may be "accepted" into the vacant orbitals of other molecules, termed acids. This acceptance of electrons takes place reversibly and with little or no activation energy. Typical bases or donors are pyridine, acetone, ether, alkyl bromides, alkyl iodides, alkyl sulfides, iodide ion, thiocyanate ion, and aromatic hydrocarbons. Typical acids are the pure and mixed halogens, sulfur dioxide and trioxide, boron trichloride and trifluoride, aluminum halides, and stannic chloride. R. S. Mulliken and his co-workers have pointed out the close relationship between base

strength and ionization potential and elaborated a theory of charge transfer complexes. H. A. Benesi and Hildebrand discovered the strong absorption in the ultraviolet characteristic of such complexes. They found that the basic strength increases in the order benzene < toluene < xylene < mesitylene. R. L. Scott determined that acid strength increases in the order $Cl_2 < Br_2 < I_2 < BrI < BrCl < ICl$.

This type of interaction is specific and saturating, and it reduces the escaping tendencies of the components. It corresponds to $\epsilon_{12} > (\epsilon_{11}\epsilon_{22})^{1/2}$. In cases where it is weak it may reduce but not overcome the opposite effect of unequal London forces.

Ion-ion interaction. The ions in a solid or liquid salt attract and repel electrostatically according to Coulomb's law, but there is also a London force component, and large ions are polarized by smaller ones. This last effect is illustrated by solid silver bromide, which is colored although its ions in aqueous solutions are colorless and whose crystals have the sodium chloride structure. The evidence is that in both the solid and the fused salt the electron cloud of the bromide ion is distorted equally by each of its six neighboring silver ions.

Ion-dipole interaction. In order to dissolve a solid salt, its lattice energy must be supplanted by the ion-ion action of another salt already in the liquid state or by the predominantly electrostatic attraction of a polar solvent or by the specific chemical interaction represented by complex ions.

Ideal solution. It is profitable to deal with actual solutions in terms of their departure from a simple idealized model—a mixture of components having the same attractive fields, which mix without change in volume or heat content. This is analogous to an ideal gas mixture, which is formed with no heat of mixing and in which the total pressure is the sum of the partial pressures. In such a solution the escaping tendency of the individual molecules is the same, whether they are surrounded by similar or by different molecules. Therefore, the combined escaping tendency of all the molecules of species 1, f_1, is given by Eq. (6), where x_1 is the mole fraction of species 1 and f_1^0 is the escaping tendency of the molecules from the pure liquid. For a binary mixture, $x_1 + x_2 = 1$. If the gas imperfections of vapors are disregarded, vapor pressures may be substituted for fugacities to give Raoult's law (1886) in its usual form, $p_1 = p_1^0 x_1$. The total pressure is $P = p_1 + p_2 = p_1^0 x_1 + p_2^0 x_2$.

$$f_1 = f_1^0 x_1 \tag{6}$$

A more sophisticated derivation than the foregoing requires one to postulate molecules of the same size and shape. The gross structure of the solution containing n_1 plus n_2 molecules of two components is identical with that of the pure liquid components. The number of configurations of the components in the mixture within this structure is $(n_1 + n_2)!/n_1!n_2!$, and the configurational entropy of mixing is, by Stirling's formula, Eq. (7). The substitution of a number of moles of each, N_1 and N_2, gives Eq. (8). The partial derivative $(\partial \Delta S/\partial N_1)_{N_2}$ represents the partial molal entropy of transfer of component 1 from pure liquid into an ideal solution of mole fraction x_1, Eq. (9). Because the model postulates no change in enthalpy,Eq. (10) applies, and $f_1/f_1^0 = x_1$, which is Raoult's law, where f is fugacity. *See* FUGACITY.

$$\Delta S = k\left[n_1 \ln \frac{n_1 + n_2}{n_1} + n_2 \ln \frac{n_1 + n_2}{n_2}\right] \tag{7}$$

$$\Delta S = R\left[N_1 \ln \frac{N_1 + N_2}{N_1} + N_2 \ln \frac{N_1 + N_2}{N_2}\right] \tag{8}$$

$$\overline{S}_1 - S_1^0 = -R \ln x_1 \tag{9}$$

$$T(\overline{S}_1 - S_1^0) = -RT \ln \frac{f_1}{f_1^0} \tag{10}$$

But to arrive at this conclusion one must assume identical structures in the solution and the pure liquid components. This is very far from the case in solutions of high polymers in ordinary solvents, even though, as with polystyrene in benzene, the heat of mixing is practically zero.

Moderate difference in molal volume between components of high symmetry has little effect, as might be expected from the fact that the radius varies only with the cube root of the volume.

As a foundation for dealing with actual solutions in terms of the deviations of their properties from those of the model, it is necessary to derive other equivalent thermodynamic relationships.

Solubility of a crystalline solid. The fugacity of a crystalline substance f^s at temperature T is less than that of its supercooled liquid f_0 to an extent depending upon its melting point T_m and heat of fusion ΔH^F, as given by Eq. (11). If ΔH^F is assumed constant, this gives upon integration Eq. (12). If the heat capacities of the solid and liquid forms are known, the variation of ΔH^F with temperature can be taken into account. This is hardly necessary for the present purpose because the deviations from ideal solubility involve factors that are more uncertain than this.

$$\frac{d \ln(f^s/f^0)}{dT} = \frac{\Delta H^F}{RT^2} \tag{11}$$

$$\ln \frac{f^s}{f^0} = -\frac{\Delta H^F}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) \tag{12}$$

Molecular weight measurements. If a solid dissolves to form an ideal solution, its heat of solution is the same as its heat of fusion ΔH^F and $f_1^s/f_1^0 = x_1$. Therefore, Eq. (13) is formed. This is the approximate equation for solubility of a solid that forms an ideal solution. It can be transformed into one much used for determining the molal weight of a solute by the depression of the freezing point of the solvent, here

$$-\ln x_1 = \frac{\Delta H^F}{R}\left(\frac{T_m - T}{T_m T}\right) \tag{13}$$

component 1. For $-\ln x_1$, one can write $\ln (1 + N_2/N_1)$. Expanding in powers of N_2/N_1 gives Eq. (14).

$$\ln\left(1+\frac{N_2}{N_1}\right) = \frac{N_2}{N_1}\left[1-\frac{1}{2}\frac{N_2}{N_1}+\frac{1}{3}\left(\frac{N_2}{N_1}\right)^2-\cdots\right] \quad (14)$$

When $N_2 \ll N_1$, the higher powers may be neglected to give Eq. (15), where $\Delta T = T_m - T$. By measuring ΔT for a known weight of solute in N_1 moles of solvent, the molal weight of the solute can be calculated. Because of the approximations made and the fact that even good solvents for the solid in question are seldom ideal, the resulting molal weights are not very exact unless extrapolated to $x_2 = 0$ from a series of values.

$$\frac{N_2}{N_1} \approx \frac{\Delta H_1^F}{RT_m^2}\Delta \quad (15)$$

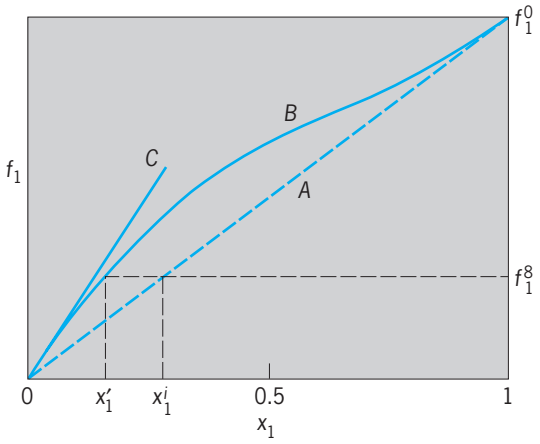
The lowering of the vapor pressure of a solvent upon the addition of a nonvolatile solute (component 2) may be offset by raising the temperature to restore the pressure of the solvent. These changes are related as in Eq. (16). This relation is far less useful than that for the freezing point depression because the heat of vaporization is much greater than the heat of fusion; therefore, the elevation of the boiling point is much smaller than the depression of the freezing point and also is harder to determine.

$$x_2 = \frac{\Delta H_1^{\text{vap}}}{RT_b^2}\Delta T \quad (16)$$

Osmotic pressure. One mole of a solvent can be removed from a large quantity of a solution in which its mole fraction is x_1 in two reversible, and hence equivalent, ways. If it is distilled from the solution into pure liquid, the gain in (Gibbs) free energy is $\Delta G_1 = RT \ln (f_1^0/f_1)$. If it is pressed out through a semipermeable membrane against the hydrostatic pressure difference, osmotic pressure ΔP, the gain in free energy is $\Delta P\overline{V}_1$, where $\overline{V}_1$ is the partial molal volume of the solvent. In an ideal solution this is the molal volume. Equating the free energy of these processes gives Eq. (17). Expanding as before and neglecting the higher powers gives $\Delta P\overline{V}_1 \approx (N_2/N_1)RT$ or $PV = RT$, where $V = N_2\overline{V}_1/N_1$, the volume containing 1 mole of solute.

$$\Delta P\overline{V}_1 = RT\ln\frac{f_1^0}{f_1} = RT\ln\frac{N_1+N_2}{N_1} \quad (17)$$

This is the van't Hoff law for osmotic pressure, put forth in 1887. The theoretical basis of Raoult's law, discovered at almost the same time, was not yet appreciated. The formal correspondence between the van't Hoff law and the perfect gas law seemed to lend unique significance to osmotic pressure and elevated the van't Hoff law to the status of an ideal solution law. It is a limiting law only and not valid at high concentrations; it neglects the specific nature of the solvent. The solvent is regarded as furnishing space for a quasi-gas solute. Thus the law cannot cover molecular states in solutions of finite concentration.



Fugacity and mole fraction. Line *A*, ideal. Line *B*, typical nonideal. Line *C*, Henry's law.

The determination of osmotic pressure offers a valuable means for determining molal weights of high polymers in solution, where high weight concentrations correspond to mole fractions so low as to have only minute effects upon the vapor pressure and the freezing point of the solvent. For example, consider a solution of 0.001 mole of solute in 1 mole of benzene; $T_m = 279$ K and $\Delta H^F{}_1 = 2370$ cal/mole, ΔT by Eq. (15) would be only 0.065°, but ΔP by Eq. (17) would be 194 mm. The latter is large enough to be measured with some precision.

Nonideal solutions. Unlikeness of the components of a binary mixture leads, as explained earlier, to fugacities in excess of ideal values. The excess is largest when the molecules of one species are surrounded mostly by those of the other, as shown in the **illustration**. They approach Raoult's law at the upper end and Henry's law, $p_1 = kx_1$ (or $f_1 = k_1x_1$), at the lower end, where k is an experimentally determined constant.

An important relation between the two components is given by the Gibbs-Duhem equation (18).

$$\left(\frac{\partial \ln f_1}{\partial \ln x_1}\right)_T = \left(\frac{\partial \ln f_2}{\partial \ln x_2}\right)_T \quad (18)$$

If Raoult's law holds in the limit when $x_1 = 1$, since $(\partial \ln f_1)/(\partial \ln x_1) = 1$, then also $(\partial \ln f_2)/(\partial \ln x_2) = 1$. Integrating gives $\ln f_2 = \ln x_2 + \ln k_2$ or $f_2 = kx_2$, where k is a constant of integration that cannot be evaluated unless Raoult's law holds for component 1 over the whole range. In that case it also holds for component 2.

The activity in the case of nonelectrolytes is defined as $a_1 = f_1/f_1^0$ and so on. In an ideal solution, $a_1 = x_1$ and $a_2 = x_2$. The activity coefficient is $\gamma_1 = a_1/x_1$ and so on. Alternate, equivalent forms of the Gibbs-Duhem equation include Eq. (19), and

$$\frac{\partial \ln a_1}{\partial \ln x_1} = \frac{\partial \ln a_2}{\partial \ln x_2} \quad (19)$$

$N_1 d\overline{G}_1 + N_2 d\overline{G}_2 = 0$, where $\overline{G}$ denotes partial molal Gibbs free energies.

If one component is a crystalline solid, its activity a_s is less than that of the liquid, which is 1, as given by Eq. (12); its maximum solubilities would be x^i_1 in the illustration, if in an ideal solvent, and x'_1, if in a real solution.

Regular solutions. There are many mixtures of nonpolar components in which, except in the immediate neighborhood of a critical mixing temperature, thermal agitation suffices to neutralize tendencies to segregate and yields virtually complete randomness of mixing, with a close approach to ideal entropy of mixing, Eqs. (8) and (9).

The enthalpy of mixing can be calculated as the difference between the potential energy of the mixture and the sum of the potential energies of the liquid components. The potential energy of a mole of liquid may be related to the potential between a pair of molecules $\epsilon(r)$. The lattice energy of a crystal is obtained by summation of $\epsilon(r)$ over all of the lattice distances; that of a liquid is obtained by integration over the continuous distribution function $\rho(r)$. The expression for a pure liquid is Eq. (20).

$$\Delta E_{\text{vap}} = -\frac{2\pi N_{Av}^2}{V} \int \rho(r)\epsilon(r) r^2 dr \qquad (20)$$

Here N_{Av} is the Avogadro number and V is the molal volume of the liquid. The corresponding expression for the potential of the mixture of N_1 and N_2 moles of the pure components involves their relative sizes. With certain simplifying assumptions, including the geometric mean for ϵ_{12} (r), Eq. (21*a*) was

$$\Delta E_M = \frac{N_1 V_1 N_2 V_2}{N_1 V_1 + N_2 V_2} \times \left[\left(\frac{\Delta E_1}{V_1} \right)^{1/2} - \left(\frac{\Delta E_2}{V_2} \right)^{1/2} \right]^2 \qquad (21a)$$

$$\overline{E} - E_2^0 = V_2 \phi_1^2 (\delta_1 - \delta_2)^2 \qquad (21b)$$

obtained for the energy of mixing N_1 and N_2 moles of two nonpolar liquids. The corresponding partial molal energy of transferring pure liquid to solution, for component 2, is Eq. (21*b*). Here ϕ_1 denotes volume fraction, neglecting expansion on mixing, and the δ's are $(\Delta E_{\text{vap}}/V)^{1/2}$, designated solubility parameters. Because energy and enthalpy are virtually identical for liquids, Eq. (21*b*) may be combined with the entropy of transfer as given by Eq. (9) to give the free energy of transfer, as in Eqs. (22*a*) and (22*b*).

$$\begin{aligned} \overline{G}_2 - G_2^0 &= \overline{H}_2 - H_2^0 - T(\overline{S}_2 - S_2^0) \\ RT \ln a_2 &= V_2 \phi_1^2 (\delta_1 - \delta_2)^2 + RT \ln x_2 \end{aligned} \qquad (22a)$$

$$\text{or} \quad RT \ln \gamma_2 = V_2 \phi_1^2 (\delta_1 - \delta_2)^2 \qquad (22b)$$

The quantity in square brackets in Eq. (21*a*) is the excess of the arithmetic mean, $\frac{1}{2}[(\Delta E_1/V_1) + (\Delta E_2/V_2)]$, over the geometric mean, $[(\Delta E_1/V_1) \times (\Delta E_2/V_2)]^{1/2}$. As mentioned earlier, the geometric mean assumption is amazingly valid for many species of similar electronic structure but fails for others.

Solubility parameters and molal volumes, 77°F (25°C)

Liquid	Formula	Molal volume, ml	Solubility parameter
Perfluoroheptane	C_7F_{16}	225	5.9
Perfluorotributylamine	$(C_4F_9)_3N$	360	6.0
Perfluoromethylcyclohexane	c-$C_6F_{11}CF_3$	196	6.1
n-Heptane	n-C_7H_{16}	147	7.4
Silicon tetrachloride	$SiCl_4$	115	7.6
Cyclohexane	C_4H_{12}	109	8.2
Carbon tetrachloride	CCl_4	97	8.6
Chloroform	$CHCl_3$	81	9.3
Benzene	C_6H_6	89	9.2
Carbon disulfide	CS_2	60	10.0
Bromine	Br_2	52	11.5
Iodine	I_2	59	14.1

Representative values of solubility parameters at 77°F (25°C) are given in the **table**. Parameters for substances solid at 77°F (25°C) have been calculated by Eq. (14*a*) from solubilities of these substances in solvents whose parameters are well determined. With paraffins the solubility data give concordant δ-values a little greater than $(\Delta E_{\text{vap}}/V)^{1/2}$.

Joel H. Hildebrand

Bibliography. *Annual Review of Physical Chemistry*, annually; E. A. Boucher, *Properties of Liquids and Solutions*, 2d ed., 1994; R. G. Mortimer, *Physical Chemistry*, 2d ed., 2000; J. H. Noggle, *Physical Chemistry*, 3d ed., 1997; O. Popovych and R. P. Tompkins, *Nonaqueous Solution Chemistry*, 1981, reprint 1990.

Solution mining

The extraction of valuable components from a mineral deposit using an aqueous leaching solution. In its original sense, solution mining refers to evaporite mining, the dissolution of soluble rock material such as salt by using borehole wells to pump water into the deposit and remove the resulting saturated brine. In its current usage, solution mining also includes ore leaching, the in-place (in-situ) leaching of valuable metal components from an orebody, and the mine-site procedures of heap leaching and dump leaching. Often included is the Frasch process of using superheated water to melt sulfur in its deep deposits and recover the molten sulfur in borehole wells. *See* ORE AND MINERAL DEPOSITS.

Evaporite Mining

Evaporites represent a broad class of water-soluble minerals (salts). Commercially important evaporites include halite (sodium chloride), sylvite and silvinite (potash), bischofite (magnesium chloride), nahcolite (sodium bicarbonate), trona (raw soda ash), and langbeinite and carnallite [both sources of potash and magnesia (magnesium oxide)]. Most of these

evaporite deposits are being mined in North America and Europe. *See* SALINE EVAPORITES.

The geologic setting (shape, attitude, and extent) and nature of each evaporite resource affect the mining method used. Sometimes conventional mechanical mining is more economical. In the Wyoming trona operations, either underground room-and-pillar or solution mining techniques are used, depending on localized conditions. Alternatively, natural brines may be processed directly for their salts; for example, soda ash is recovered from subsurface brines at Searles Lake, California.

Solution mining involves injecting a solvent into the pay zone of the deposit through a cased borehole. For evaporites, the solvent is hot water, which forms brine as the soluble minerals dissolve. The brine is brought to the surface via the casing system in the same or another borehole and sent to a processing facility for recovery by the controlled crystallization of the desired product, followed by dewatering and drying. Some minerals may require additional purification steps, such as the flotation of potash crystals. The depleted brine is chemically reconditioned and injected back into the deposit. Often a bleed stream, typically only 5-10% of the total brine flow, is withdrawn and treated to control the buildup of deleterious impurities in the solvent. Any impurity products, usually small in volume, woud be handled in the appropriate manner. Thus, solution mining creates minimal surface disturbance and little waste, compared to conventional mining. *See* WELL.

Boreholes are drilled to form a grid across the mining zone, and the injection pumping systems are located on the surface. Normally the injection pressure is sufficient to lift brine to the surface without submersible pumps in the recovery bores. To provide constant feed to the process plant, new wells are drilled periodically and old ones taken off-line and plugged as areas are leached out, until the entire evaporite bed is depleted.

Although solution mining is simple in principle, there are several key issues for successful operation. One is to maintain close control over the solvent parameters, such as pH, oxidation potential, temperature, and pressure. For example, a pressure loss would allow dissolved gas to escape, shifting the solubilities of the remaining constituents. *See* ELECTROCHEMICAL SERIES.

Another issue is isolation of the solution mining zone from the surrounding geologic structures. This requires effective well completion techniques that are compatible with the brine. Once the well is drilled and cased, the casing must be cemented into the formation. This prevents brine from migrating along the annular space outside the casing and

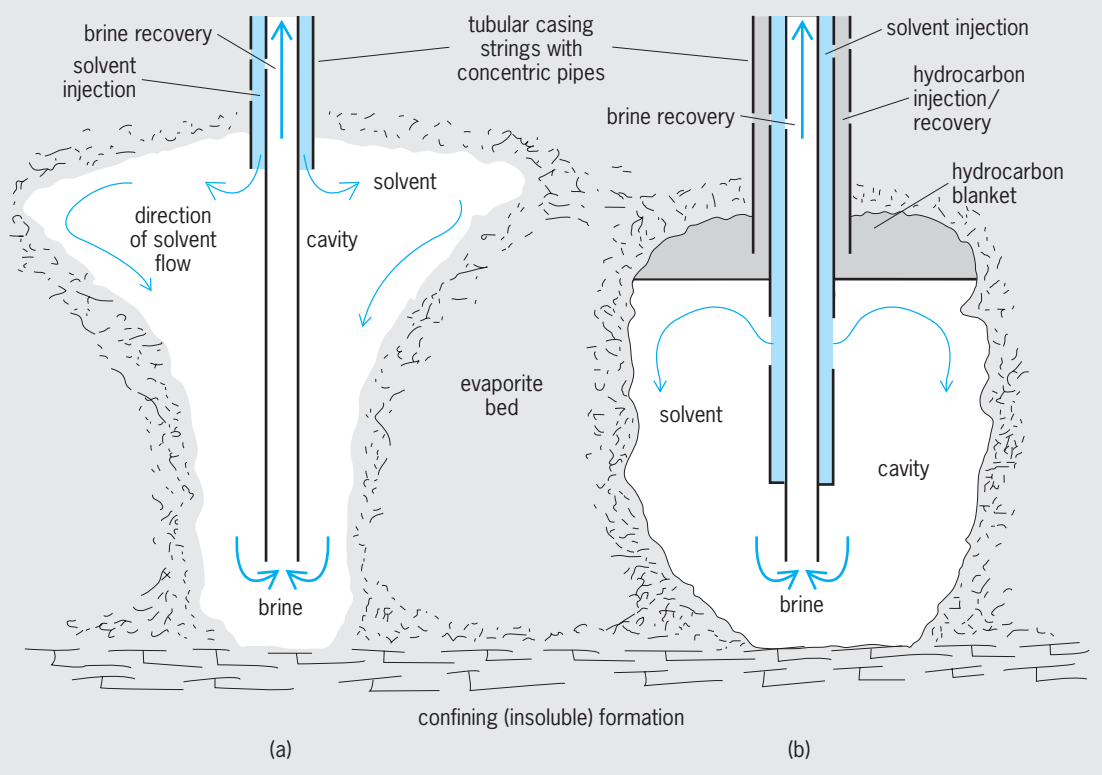


Fig. 1. Solution mining of a subsurface evaporite deposit. Mining is normally initiated at the bottom of the deposit and retreats vertically, leaving a mined-out cavity. Saturated brine is heavier (denser), and is recovered from the bottom of the cavity. (*a*) Early single-well solution mining system with uncontrolled solvent flow. Unsaturated solution (low-density) is injected at the top of the cavity, where most of the dissolution occurs, giving rise to a "morning glory"–shaped cavity. Roof span will continue to increase until it becomes unstable and collapses under the load of the overlying material. (*b*) Single-well system with hydrocarbon blanket. The blanket prevents excessive dissolution of salt at the top of the cavity. The hydrocarbon can be withdrawn periodically to allow for the controlled increase in cavity height.

contaminating adjacent aquifers. *See* OIL AND GAS WELL COMPLETION.

Proper brine flow patterns within the cavity control the shape of the cavity so that it is stable and not subject to excessive stresses, which might cause a ceiling collapse. An inert material (oil or gas) is often injected to form a blanket at the top of the cavity, retarding dissolution of the ceiling (**Fig. 1**).

Controlling the stability of solution mined cavities is particularly important in salt domes along the Gulf Coast. If the mined-out cavities are stable, they have great value for underground storage, as salt has very low porosity and permeability and is geotechnically stable. Cavities below 1000 ft (300 m) can withstand great pressures and are used for storing crude oil, liquefied petroleum products, and gases. Shallower cavities are used for climate-controlled storage, including isolation of waste materials. *See* SALT DOME.

W. Joseph Schlitt

Ore Leaching

Leaching large stockpiles and mine waste dumps (ore heaps) and, somewhat less important, modified in-situ leaching of caved underground orebodies annually accounts for about one-third of both new copper and gold production in the United States, the world's second largest source of both commodities. Ore leaching is an important contributor to the total world production of these commodities, along with silver. Ore leaching's very low unit processing cost, combined with low-cost earthmoving, allows profitable treatment of huge tonnages of low-grade material that could not otherwise be exploited.

Ore containing rock must first be fragmented to allow chemical leaching solutions to percolate through it. The metal-bearing minerals are liberated using an aqueous wetting fluid that penetrates and saturates the microfractures within the ore particles by capillary action. The ore minerals, originally deposited geochemically from hydrothermal solutions, are typically located in these micropores. Once dissolved, the metals are removed from the ore particles by diffusion through the solution-filled micropores and swept out of the ore heap by the flowing leachate. In effect, nature's ore deposition process has been reversed at a quicker pace. Nevertheless, commercial leaching times are measured in weeks, months, and even years. Leachates, or "pregnant liquors," are processed in surface plants to recover the metal and regenerate the leachant for reuse in the ore heap in a closed-loop system. *See* HYDROMETALLURGY; LEACHING.

In the case of gold and silver ores, the leachant is a basic sodium cyanide solution which forms soluble complex anions [$Au(CN)_2^-$]. Gold is recovered from solution either by adsorption on granular activated carbon or by reductive precipitation (cementation) with zinc dust. Ores amenable to heap leaching are usually fine-grained sedimentary rocks with highly disseminated, submicrometer, free gold grains. While low-grade, these deposits are typically massive and account for nearly all of the present primary gold production in the United States and several other countries. Coarse gold lode deposits and placer deposits are not good candidates for heap leaching because the gold dissolution rate is slow and unacceptably long times are required to dissolve coarse gold grains. Sealed rock microfractures will prevent extraction. A fairly common example of sealing is silicification, occurring after ore deposition. *See* ADSORPTION; GOLD METALLURGY; PRECIPITATION (CHEMISTRY); SEDIMENTARY ROCKS.

Oxidized copper ores contain minerals, such as chrysocolla and malachite, that are readily acid-soluble. Sulfuric acid is the universal leachant for these ores, with the copper dissolved as sulfate salts. However, in the more prevalent primary ores, chalcopyrite ($CuFeS_2$) is the dominant copper mineral, accompanied by relatively large amounts of pyrite (FeS_2). Metals are released from these and other sulfide minerals only upon biooxidation. In the case of pyrite, biooxidation generates sulfuric acid and large quantities of soluble iron within the ore heap. Consequently, primary copper ores generally do not require an external source of acid. *See* BIOLEACHING; COPPER METALLURGY.

Autotrophic bacteria, principally *Thiobacillus ferrooxidans*, are ubiquitous in sulfide ore leaching systems exposed to air and water. The primary role of these bacteria in sulfide leaching systems is to catalyze the oxidation of ferrous ions to ferric ions in the leachant while it is exposed to air inside the ore heap. At the necessarily modest (near-ambient) ore leaching temperatures, uncatalyzed iron oxidation is much too slow. Soluble ferric ions generated by biooxidation outside the ore particles diffuse under a concentration gradient through microfractures into the ore particles, causing sulfide mineral oxidation, and also causing generation and counterdiffusion of both ferrous ions and cupric ions out of the ore particle in concentration gradients. For air (oxygen) to be present, as required for sulfide leaching to occur, the ore heap must not be saturated with percolating leachant. However, if the ore fragmentation is too fine or if excessive amounts of clay are present, heap permeability will be inadequate and saturation may occur, thereby excluding air. *See* INDUSTRIAL MICROBIOLOGY; IRON; OXIDATION-REDUCTION.

Soluble iron and acid released by pyrite oxidation are eventually limited to steady-state concentrations by other consuming chemical processes. Acid attack of nonore minerals usually limits the pH to a range from 2.3 to about 2.9 depending on the individual ore mineral assemblage. Iron is precipitated by hydrolysis reactions, most commonly in ore heaps by generation of potassium jarosite. The concentrations of soluble iron species depend on solution oxidation potential and pH.

From process engineering and economical perspectives, the ore leaching time required to achieve a desirable extent of metal extraction is paramount. Leaching rates must be considered at different spatial scales. The microscale includes the heterogeneous dissolution kinetics of the ore minerals and the

behavior of bacteria. Transport of reactant and product species by diffusion in rock micropores depends on rock size, with the time to achieve equivalent fractional extraction increasing with the ore particle size squared (r_p^2). Heap leaching may involve ore fragmented only by blasting and excavation at the mine face, or it may involve one or more stages of crushing to reduce the rock size and substantially lower the required leaching time when rock micropore diffusion is rate-limiting. Often, crushed ore will be heap-leached for a fixed duration and removed to a spent-ore discard pile, followed by restocking the leaching pad with fresh ore (the on/off method). Alternatively, the leached ore may be covered with a new layer (a new ore lift). Generally, this approach is less expensive, and it is more common when the ore is not crushed, but it is not as well controlled as the on/off method. In both cases, a variety of leachant distribution methods using portable pipe and equipment are used to spread the leachant on the ore heap. These are similar to agricultural irrigation methods and include drip irrigation. Lightweight plastic materials are used for portability and avoidance of corrosion.

Many factors affect leaching extractions over time, including mineral grain size distribution and accessibility to open microfractures, rock microporosity (typically 1–6% in disseminated gold and copper ores), ore particle size and size distribution, and reactant concentrations. Because of the complexities and uncertainties of ore leaching, thorough amenability testing of each ore must be conducted. This is done by leaching representative ore samples in columns that mimic vertical sections through the planned ore heap.

In large ore heaps requiring oxygen for biooxidation of sulfide minerals, inadequate air flow into the heap usually limits the extraction rate. Heat from exothermic sulfide mineral oxidation reactions, and changes in the composition of oxygen and water vapor in the air cause some buoyancy, but this natural advection is very weak and usually inadequate for economic metal extraction rates. Forced-air ventilation with low-pressure fans using perforated air pipes, either underlying ore heaps or bored vertically into preexisting ore heaps, is usually affordable. This approach is shown in **Fig. 2**; a crushed ore with gold encapsulated in sulfide minerals (pyrite and arsenopyrite) is being treated by heap biooxidation prior to conventional cyanide heap leaching, which is conducted after removal of the ore and neutralization of its self-generated acid.

True in-situ leaching of permeable ore deposits, now the primary method of uranium production in the United States, is similar to production of brines from evaporite deposits, involving a field of alternate injection wells and production wells. However, the deposit permeability and porosity are generally very much lower than for evaporites, and a chemical leachant must be used. For uranium, the leachant is commonly an ammonium carbonate solution. True in-situ leaching of deep copper ore deposits (without fragmentation) has been attempted in major field tests but not yet successfully commercialized because of low formation permeability and poor sweep efficiency, coupled with high capital costs. *See* URANIUM METALLURGY.

Fig. 2. Forced-air ventilation of sulfide ore heaps during biooxidation.

Careful environmental planning, permitting, and control during ore leaching are essential. Containment to prevent spills and ground-water contamination, and adequate treatment of all discharged solutions to meet safe drinking water standards must be achieved. At a minimum, this involves geotechnically engineered, impermeable membrane leach pads and solution storage ponds. Ore leaching heaps are exposed to weather and cannot be turned off during maintenance shutdowns as conventional chemical plants are. Large sealed storage and overflow ponds must be provided to collect spills and handle variations in water balance through the year, including floods.

Reclamation of spent ore heaps, discard piles, and in-situ leached ore deposits is primarily directed at the soluble contaminants derived from both the ore and the chemicals introduced during leaching. Acid mine drainage, formed when rain water percolates through the spent ores, containing remnant sulfide minerals that continue to slowly oxidize, is a major problem that has required water treatment plants to be operated in perpetuity at closed mines. Biochemical reduction of sulfates using soluble organic media within the spent ore heap is a promising solution to this problem. As with leaching, environmental site reclamation is often a slow process, requiring months and years of treatment. *See* ENVIRONMENTAL ENGINEERING; LAND RECLAMATION; MINING.

Robert W. Bartlett

Frasch Process

Modern sulfur mining dates from the invention of the Frasch process in the late nineteenth century. H. Frasch accomplished in-place sulfur mining by using superheated water. The process has come down to the present with only minor modifications, and it is

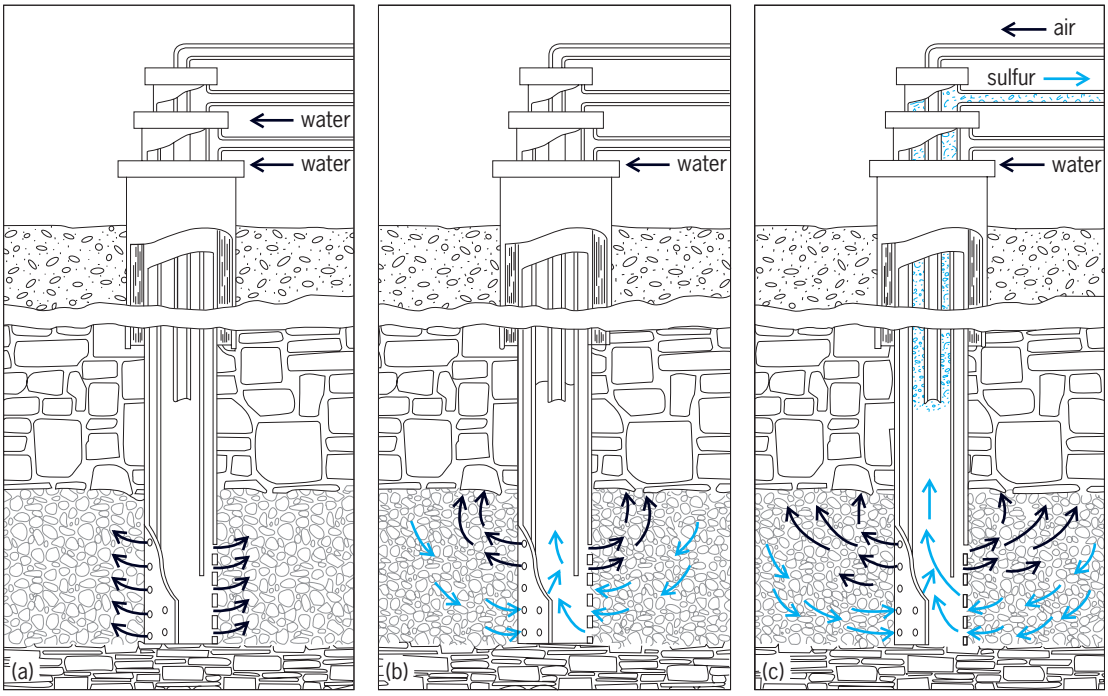


Fig. 3. Sulfur production well in a Gulf Coast salt dome deposit. (*a*) Superheated water is pumped down the inner and outer casing strings. (*b*) Molten sulfur enters inner casing string. (*c*) Sulfur is jetted to the surface as a result of compressed air being pumped down the inner line.

used for mining sulfur deposits in West Texas, offshore Louisiana, Iraq, Poland, and Russia.

Frasch mining begins by drilling wells into the sulfur deposit. Steel tubing (casing) is run into the drill hole to case off the barren overlying formation, and it is cemented in at the top of the sulfur deposit. Within this casing, three concentric strings of pipe are set within the sulfur deposit. The outermost string of pipe, 8 in. (20 cm) in diameter, is set to the bottom of the sulfur deposit. The lower 20 ft (6 m) of this pipe is perforated, providing access for drillhole fluids to the sulfur deposit. Within this string of pipe is another, 4 in. (10 cm) in diameter, with its lower 10 ft (3 m) perforated. In the annulus between the two strings of pipe, a seal (or packer) is placed about 10 ft (3 m) from the bottom of the two casing strings. Thus, superheated water at 325°F (163°C) can be pumped down the annulus between the two strings of pipe to leave the casing string through the perforations above the packer, and to circulate in the sulfur deposit (**Fig. 3***a*). Native sulfur melts at 275°F (135°C) in a cone-shaped area of influence around the well, and because liquid sulfur is denser than water, it settles to the bottom of the well.

The pressure from the heated water forces the molten sulfur into the lower set of perforations and into the inner string of casing (Fig. 3*b*). The molten sulfur rises to its hydrostatic head. This point is determined by practice, and a 0.75-in. (1.9-cm) titanium air line is set at this depth inside the 4-in. (10-cm) casing. Compressed air pumped through this inner tubing expands when it leaves the tubing, thus "jetting" the sulfur to the well collar, where it is collected in a tank (sulfur pan) [Fig. 3*c*]. The procedure to initiate sulfur production involves pumping hot water through inner and outer casing strings, followed by changing to the production setup with sulfur and air traveling through the inner casing strings. The water used to mine the sulfur cools and is removed from the sulfur-bearing zone by bleed-water wells.

Alan F. Edwards

Bibliography. R. W. Bartlett, Metal extraction from ores by heap leaching, *Metals Mater. Trans. B*, 28B:529–545, 1997; R. W. Bartlett, *Solution Mining: Leaching and Fluid Recovery of Materials*, 2d ed., Gordon & Breach, 1998; *Dictionary of Mining, Mineral and Related Terms*, 2d ed., American Geological Institute, Alexandria, VA, 1997; A. F. Edwards et. al., *The Geology of the Main Pass Sulphur and Oil Deposit, Amer. Inst. Mining Preprint*, no. 92–254, 1992; I. A. Given (ed.), *Mining Engineering Handbook*, vol. 2, sec. 21, SME-AIME, 1973; H. L. Hartman (ed.), *SME Mining Engineering Handbook*, 2d ed., sec. 15.2 and 15.3, SME, Littleton, CO, 1992; *In-Situ Leaching Mining, U.S. Bur. Mines Inform. Circ.*, no. 9216, March 1989; D. D. Rabb, Solution mining, *Min. Eng.*, 24(2):62–64, February 1972; D. D. Rabb, *Solution Mining: A Review*, SME-AIME, September 1986; W. J. Schlitt (ed.), *Salts & Brines '85*, SME-AIME, New York, 1985.

Solvation

The association or combination of a solute unit (ionic, molecular, or particulate) with solvent molecules. This association may involve chemical

or physical forces, or both, and may vary in degree from a loose, indefinite complex to the formation of a distinct chemical compound. Such a compound contains a definite number of solvent molecules per solute molecule.

Solvation occurring in aqueous solutions is referred to as hydration. In aqueous ionic solutions, the highly polar water molecules become oriented about the ions, forming spheres of hydration. As a result, the mobility of an ion under an applied voltage gradient is decreased. The extent of hydration depends upon the size and charge of the ion.

In certain colloidal suspensions, solvation is, to a large extent, responsible for the stability of the sol. Particles in lyophilic sols strongly adsorb on their surfaces one or more layers of solvent molecules. This protective layer prevents the particles from approaching so closely as to adhere. In water, starch and many proteins form suspensions which are highly solvated. Such systems exhibit a viscosity which is markedly higher than that of the pure solvent. *See* COLLOID; ELECTROLYTIC CONDUCTANCE; HYDRATION; SOLUTION.

Francis J. Johnston

Solvent

By convention, the component present in the greatest proportion in a homogeneous mixture of pure substances (solutions). Components of mixtures present in minor proportions are called solutes. Thus, technically, homogeneous mixtures are possible with liquids, solids, or gases dissolved in liquids; solids in solids; and gases in gases. In common practice this terminology is applied mostly to liquid mixtures for which the solvent is a liquid and the solute can be a liquid, solid, or gas. *See* SOLUTION.

Three broad classes of solvents are recognized—aqueous, nonaqueous, and organic. Formalistically, the nonaqueous and organic classifications are both not aqueous, but the term organic solvents is generally applied to a large body of carbon-based compounds that find use industrially and as media for chemical synthesis. Organic solvents are generally classified by the functional groups that are present in the molecule, for example, alcohols, halogenated hydrocarbons, or hydrocarbons; such groups give an indication of the types of physical or chemical interactions that can occur between solute and solvent. Nonaqueous solvents are generally taken to be inorganic substances and a few of the lower-molecular-weight, carbon-containing substances such as acetic acid, methanol, and dimethylsulfoxide. Nonaqueous solvents can be solids (for example, fused lithium iodide), liquids [sulfuric acid (H_2SO_4)], or gases [ammonia (NH_3)] at ambient conditions; the solvent properties of the first-named substances are manifested in the molten state, whereas the last-named substances must be liquefied to act as solvents.

Nonaqueous Solvents

The classification, often arbitrary, of nonaqueous solvents can be made on the basis of a variety of factors. Excluding utilitarian considerations and classifications based on chemical character, that is, the presence of distinctive groups such as carbonyl in the molecule, useful classification schemes involving nonaqueous solvents inevitably involve considerations of acid-base phenomena. Of the two major theories, the Brönsted-Lowry protonic concept has been the most useful because the early solvents of interest were invariably potential proton donors. However, the Lewis theory of acidity has become very useful for understanding solution phenomena in aprotic systems such as sulfur dioxide (SO_2). *See* ACID AND BASE.

The ability of a solvent to form a so-called onium species is an important factor in defining the nature of the solution phenomena it will support. Onium species can be formed by reaction with a potential proton donor or by self-ionization of the solvent. Four classes of solvents are generally recognized according to their ability to coordinate with the proton—basic, acidic, aprotic, and amphiprotic solvents. *See* COORDINATION CHEMISTRY.

Basic solvents. Such solvents form the onium species most readily and are generally derivatives of ammonia (amines, hydrazines, pyridine, and so on) or water (alcohols and ethers). Hydrogen-containing species in this classification can undergo self-ionization to form the onium species, for example, $2NH_3 \rightleftharpoons NH_4^+ + NH_2^-$, but the other substances in this class require the presence of a proton donor to do so, $R_2O + HX \rightleftharpoons R_2OH^+ + X^-$ (R = organic group; X = an anion).

Acidic solvents. These solvents exhibit a greater tendency to release protons than do basic solvents, and they form onium species with great reluctance. However, it is possible for acidic solvents to undergo self-ionization, for example, acetic acid, $2CH_3CO_2H \rightleftharpoons CH_3CO_2H_2^+ + CH_3CO_2^-$, to form the corresponding onium species. Even though a solvent may be classified as acidic, for example, CH_3CO_2H, it is possible to protonate such molecules with a more strongly acidic substance. Thus, the strong mineral acids such as hydrochloric acid (HCl) ionize in anhydrous acetic acid to form the onium species, $CH_3CO_2H + HCl \rightleftharpoons CH_3CO_2H_2^+ + Cl^-$.

Aprotic solvents. Solvents such as SO_2 and CO_2, also commonly called inert, have very little affinity for protons, and they are incapable of dissociating to give protons. Such aprotic solvents are also called indifferent, nondissociating, or nonionizing. *See* SUPERCRITICAL FLUID.

Amphiprotic solvents. These solvents are capable of either adding or donating protons. Thus, ammonia is amphiprotic because it can lose or accept a proton as shown in the following reactions:

$$(C_6H_5)_3C^- + NH_3 \rightarrow (C_6H_5)_3CH + NH_2^-$$

$$NH_3 + HX \rightarrow NH_4^+ + X^-$$

The classification of a substance as amphiprotic adds little to understanding solvent phenomena, because this designation depends only on the relative strengths of the acids or bases involved. It has always

been possible to produce a medium sufficiently acidic to protonate even acidic solvents which, of course, also would be expected to be good proton donors to bases. From this point of view, anhydrous acetic acid is considered amphiprotic. Whether a substance is an acid or base depends upon the character of the solvent in which it is dissolved. For example, urea, which is very weakly basic in water, is a weak acid in the more basic solvent ammonia and a strong base in the acidic solvent CH_3CO_2H.

Amphiprotic solvents exhibit a well-defined self-dissociation, of hydrogen sulfide, $2HS \rightleftharpoons H_2S^+ + S^-$ ($K_S = a_{H_2}S^+ \cdot a_{S^-}$), for which an autoprotolysis constant K_S can be measured (**Table 1**). It is generally difficult to measure K_S for nonaqueous solvents because apparent self-ionization can arise from the presence of traces of water or other amphiprotic contaminants. *See* IONIC EQUILIBRIUM.

Chemical behavior. Solvents intervene in chemical processes by producing species from solutes that are more reactive than if the solvent were not present. If the solute consists of ions, the energy for the dissolution process is supplied almost entirely from the solvation of ions; this is primarily an electrostatic process. However, two processes may occur when a covalent solute is dissolved: the solvation of molecules or the formation of ions. The solvation of molecular species usually involves dipolar interactions, specific interactions such as hydrogen-bond formation, and formation of covalent bonded species via coordinate covalent bond formation. In many instances it is possible for such intermediate species to undergo ionization. The formation of ions from covalent molecules is a measure of the ionizing power of the solvent, and attempts have been made to correlate this with the dielectric constant of the solvent. Unfortunately, there is a sufficient number of contradictions to make such relationships unreliable.

The extent to which a substance is ionized by neutral donor (solvent) molecules should increase with increasing stability of the cation resulting from nucleophilic attack, $S: + X\text{-}Y \rightarrow S\text{-}X^+ + Y^-$, compared with that of the unionized solute. The strength of the coordinate covalent bond found in the species SX^+ is related to the donor ability of S, the acceptor ability of X in the species XY, steric effects, and the magnitude of specific solvent-solute interactions such as hydrogen bonding. The donor strengths, expressed as donicity (D_n; **Table 2**) of solvent molecules have been defined relative to the reference acceptor antimony pentachloride ($SbCl_5$); the enthalpy of the reaction between $SbCl_5$ and a series of donors in an inert medium is taken as a measure of donicity. The donicity of a solvent has been interpreted as a measure of its donor intensity, nucleophilicity, or Lewis base strength and can be a useful guide in assessing the ionizing power of a solvent. There is no necessary correlation between the donicity of the solvent and its dielectric constant (ϵ). J. J. Lagowski

TABLE 1. Autoprotolysis constants (K_s) of some common solvents at 25°C

Solvent	log (K_s)	Dielectric constant (ϵ)
H_2O	14.0	78.4
H_2O_2	13	84.2
CH_3OH	16.9	32.6
H_2S	34.5 (–78°C; –61°F)	8.99 (–78°C; –61°F)
H_2SO_4	3.6	101
HCO_2H	6.2	58.5 (16°C; 61°F)
$H_2NC_2H_4OH$	5.1 (20°C; 68°F)	37.7
$(CH_3)_2SO$	32	46.6
$HCN(CH_3)_2$	>21	36.7
CH_3CN	28.5	36.0
C_2H_5OH	19.1	24.3
$H_2NC_2H_4NH_2$	15.3	14.2 (20°C; 68°F)
CH_3CO_2H	14.5	6.1 (20°C; 68°F)

TABLE 2. Donicity (D_n) and dielectric constant (ϵ) of some common solvents

Solvent	D_n	ϵ
1,2-Dichloroethane	0.0	10.1
Sulfurylchloride	0.1	10.5
Tetrachlorethylene carbonate	0.2	9.2
Thionyl chloride	0.4	9.2
Acetyl chloride	0.7	15.8
Benzoyl chloride	2.3	23.0
Nitromethane	2.7	35.9
Dichloroethylene carbonate	3.2	31.8
Nitrobenzene	4.4	34.8
Acetic anhydride	10.5	20.7
Phosphorus oxide chloride	11.7	14.0
Benzonitrile	11.9	25.2
Selenium oxide chloride	12.2	46.0
Monochloroethylene carbonate	12.7	62.0
Acetonitrile	14.1	38.0
Sulfolane	14.8	42.0
Propylene carbonate	15.1	69.0
Benzyl cyanide	15.1	18.4
Ethylene sulfite	15.3	41.0
Isobutyronitrile	15.4	20.4
Propionitrile	16.1	27.7
Ethylene carbonate	16.4	89.1
Phenylphosphorus oxide difluoride	16.4	27.9
Methyl acetate	16.5	6.7
Butyronitrile	16.6	20.3
Acetone	17.0	20.7
Ethyl acetate	17.1	6.0
Water	18.0	81.0
Phenylphosphorus oxide dichloride	18.5	26.0
Diethyl ether	19.2	4.3
Tetrahydrofuran	20.0	7.6
Diphenylphosphorus oxide chloride	22.4	—
Trimethyl phosphate	23.7	6.8
Tributyl phosphate	23.7	6.8
Dimethylformamide	26.6	36.1
Dimethylacetamide	27.8	28.9
Dimethyl sulfoxide	29.8	45.0
Diethylformamide	30.9	—
Dimethylacetamide	32.2	—
Pyridine	33.1	12.3
Hexamethylphosphoramide	(38.8)	30.0

Organic Solvents

Organic solvents are liquids with atmospheric boiling points lying in the approximate range 40–200°C (100–390°F). They are relatively low cost products of the petrochemical industry, with small quantities

TABLE 3. Properties of commonly used solvents

Solvent	Boiling point, °C (°F)	Specific gravity	Polarity (water = 1)	Solubility parameter*	Evaporation time (ether = 1)
n-Pentane	36.1 (97.0)	0.626	0.009	7.0	1.0
n-Hexane	69 (156)	0.659	0.009	7.3	1.4
n-Heptane	98.4 (209)	0.684	0.012	7.5	3.0
Cyclohexane	80.7 (177)	0.778	0.006	8.2	3.4
Toluene	110.6 (231.1)	0.867	0.099	8.9	6.1
Xylenes	139 (282)	0.871	0.074	8.8	13.5
Ethylbenzene	136 (277)	0.867		8.3	8.8
Methanol	64 (147)	0.792	0.762	14.5	6.3
Ethanol	78 (172)	0.789	0.654	13.4	8.3
i-Propanol	82 (180)	0.786	0.546	11.5	11
n-Butanol	118 (244)	0.810	0.602	11.4	33
Ethylene glycol	198 (388)	1.1135	0.790	17.1	1550
Ethyl glycol	136 (277)	0.929	0.627	9.9	43
Butyl glycol	171 (340)	0.901	0.602	9.5	119
Acetone	56 (130)	0.790	0.355	10.0	2.0
Me Et ketone	80 (180)	0.805	0.327	9.3	3.3
Me iBu ketone	116 (241)	0.801	0.269	8.4	7.0
Dichloromethane	40 (100)	1.326	0.309	9.7	1.4
Trichloroethylene	87 (190)	1.464	0.160	8.0	3.6
Perchloroethylene	121 (250)	1.622		4.5	11.0
Tetrahydrofuran	66 (150)	0.888	0.207	9.9	2.2
Dioxane	101 (214)	1.0336	0.164	10.0	7.3
Ethyl ether	34.5 (94.1)	0.715	0.117	7.4	1.0
Me tBu ether	55 (130)	0.740	0.148	7.4	1.6
Methyl acetate	57 (135)	0.927	0.287	9.0	2.1
Ethyl acetate	77 (170)	0.902	0.228	9.1	3.0
n-Butyl acetate	126 (259)	0.883	0.241	8.6	12
Dimethylformamide	153 (307)	0.945	0.404	12.1	120
Dimethylacetamide	166 (330)	0.959	0.657		172
Acetonitrile	81.6 (179)	0.782	0.460	11.9	2.04
Pyridine	115 (239)	0.983	0.302	10.7	12.7

*Hildebrand solubility parameter, $(\text{calories/cm}^3)^{1/2}$.

also made from coke oven by-products, from turpentine and by fermentation.

Organic solvents are used in a wide variety of domestic and industrial products and processes falling into three main categories. (1) paints, printing ink, adhesives, toileteries, and other coatings where the solvent allows the solute to be put into position and to adhere to the surface when the solvent evaporates; (2) cleaners of materials ranging from fabrics to metal surfaces and electronic parts, where the solvent carries away unwanted contaminants; and (3) reaction media, largely in the chemical and pharmaceutical industries, where the solvent is removed as a mother liquor or by evaporation, or is incorporated in the product as a means of delivery (for example, in agrochemicals).

For dissolving resins and for cleaning when no chemical change takes place, the polarity of a solvent, its solubility parameter, and its evaporation rate are useful ways of predicting a solvent's performance (**Table 3**).

Since many of the ways of recapturing solvents from the air result in the recovered solvent being water-saturated, its water solubility and the ease with which, if necessary, it can be dried are also important.

Solvents used in chemical reactions can be chosen to modify the course of a reaction and to remove, by forming an azeotrope or by a phase separation, water or other unwanted products of reaction.

Hazardous properties. Apart from the operational performance of a solvent, its choice for any application will be made with its various hazardous properties in mind.

Toxicity by inhalation. This is an important consideration. Many of the solvents that were once widely used are being rejected because of the toxicity of their vapor. Benzene and chloroform are examples of solvents that are so toxic that they are difficult to handle in a safe way. *See* BENZENE.

Other organic solvents are comparatively harmless and can be used in domestic products without serious risk (**Table 4**). A parameter known as the threshold limit value (TLV) is defined as the concentration in air considered safe for people to work in on a regular basis. These limits vary slightly country to country.

Only a few solvents have strong enough odors at their threshold limit values to be reliably detected by smell, and even then prolonged exposure tends to reduce an individual's sensitivity.

Much higher concentrations of solvent in air are those giving immediate danger to life and health (IDLH), at which 30 min as a single exposure might have a permanently harmful effect; and even this is lower than the level to which solvent abusers expose themselves, which is close to the vapor pressure of the pure solvent (Table 4).

Toxicity by absorption. Some solvents, such as dimethyl sulfoxide and dimethylacetamide, are easily

TABLE 4. Health and fire hazards of commonly used solvents

Solvent	Threshold limit value, ppm	Immediate danger to life and health, ppm	Odor threshold*, ppm	Saturated vapor at 21°C (70°F), ppm	Flash point °C (°F)	Auto-ignition temperature, °C (°F)
n-Pentane	600	5,000	400	720,000	−40 (−40)	260 (500)
n-Hexane	50	5,000	130	200,000	−22 (−7.6)	225 (437)
n-Heptane	400	4,250	220	51,800	−4 (25)	230 (446)
Cyclohexane	300	10,000	25	155,700	−7 (1.4)	260 (500)
Toluene	100	2,000	0.2	31,000	4 (39)	480 (896)
Xylenes	100	10,000	1.1	9,180	23 (73)	480 (896)
Ethylbenzene	100	2,000	3	9,960	21 (70)	460 (860)
Methanol	200	25,000	100	133,000	15 (59)	470 (878)
Ethanol	1,000	20,000	84	60,000	13 (55)	423 (793)
i-Propanol	400	20,000	90	46,000	12 (54)	456 (853)
n-Butanol	50	8,000	3	6,300	35 (95)	367 (693)
Ethylene glycol	50			53	116 (241)	400 (752)
Ethyl glycol	5	6,000	200	5,400	40 (104)	238 (460)
Butyl glycol	25	700	13	860	60 (140)	244 (471)
Acetone	750	20,000	100	342,800	−18 (−0.4)	465 (869)
Me Et ketone	200	3,000	10	112,000	−6 (21)	515 (961)
Me iBu ketone	100		25	21,700	13 (55)	459 (858)
Dichloromethane	100	5,000	250	500,000	None	605 (1121)
Trichloroethylene	50	500	30	80,000	None	420 (788)
Perchloroethylene	50	400	5	23,000	None	None
Tetrahydrofuran	200		30	230,000	−15 (5)	321 (607)
Dioxane	25	200	170	38,000	12 (54)	180 (356)
Ethyl ether	400	19,000	10	19,000	−45 (−49)	160 (320)
Methyl acetate	200	10,000	50	290,000	−10 (14)	500 (932)
Ethyl acetate	400	10,000	4	100,000	−4 (25)	484 (903)
n-Butyl acetate	150	10,000	10	14,200	22 (72)	421 (790)
Dimethylformamide	10	3,500	100	3,700	62 (144)	445 (833)
Dimethylacetamide	10	400	50		70 (158)	490 (914)
Acetonitrile	40	4,000	40	96,000	6 (43)	524 (975)
Pyridine	5	3,600	0.01	22,000	20 (68)	522 (972)

*Individuals' abilties to detect solvent vapors by odor vary widely.

absorbed through the skin. While this property can sometimes be turned to good use, great care is needed to avoid exposure of unprotected skin to them. *See* DIMETHYL SULFOXIDE.

Fire safety. Of the widely used organic solvents with an assured future, only the chlorohydrocarbons (methylene chloride, trichloroethylene, and perchloroethylene) are not flammable. Some organic solvents produce explosive vapors at ambient temperatures. This characteristic is measured by the solvent's flashpoint (Table 4).

The potential of generating an explosive atmosphere will usually accompany a high rate of evaporation and call for precautions in storage and handling, such as blanketing storage tanks with nitrogen and using specially protected electrical equipment. With the sole exception of methanol, all organic solvents have a vapor with a density substantially higher than air, and ventilation should be planned with this in mind.

Some solvents, mostly ethers and glycol ethers, have low autoignition temperatures (Table 4), meaning that their vapors mixed with air may explode without a spark or naked flame if brought into contact with a modestly hot surface.

Many solvents, particularly hydrocarbons, combine a high electrical resistance with immisciblity with water, and they generate static electricity when a mixture of solvent and water flows through a pipe. To combat the danger of a static spark causing an explosion in the head space in a tank, careful electrical grounding of the entire installation is necessary.

Environmental pollution. All used solvent finally ends its life in the atmosphere, in water, or in incineration. For many applications, particularly in domestic and workshop use, incineration is impractical. Thus for all used solvents consideration must be given to their eventual environmental impact.

Solvents have three major impacts on the atmosphere: damage to the ozone layer, contribution to the production of low-level ozone, and smell.

The use of chlorofluorocarbons (CFC 11 and CFC 131) as cleaning solvents is being rapidly phased out, and the less harmful but widely used 1,1,1-trichloroethane is also being withdrawn. *See* HALOGENATED HYDROCARBON.

Local air pollution from ozone formation in urban areas has as a vital component the presence of volatile organic compounds. Organic solvents form up to 40% of all volatile organic compounds, and it is desirable to use solvents whose photochemical ozone creation potential (POCP) is low (**Table 5**).

Some people with sensitive noses find solvent odors at well below the hazardous level unpleasant (Table 4). *See* AIR POLLUTION; STRATOSPHERIC OZONE.

Discharge of solvents into rivers or ground water is potentially harmful. The biological degradation of the solvents in wastewater plants requires special treatment and may give rise to offensive odors.

TABLE 5. Photochemical ozone creation potential (POCP) of commonly used organic solvents

Solvent	POCP
n-Pentane	
n-Hexane	40
Cyclohexane	25
Toluene	55
Xylenes	90
Ethylbenzene	—
Methanol	10
Ethanol	27
i-Propanol	15
n-Butanol	55
Ethylene glycol	—
Ethyl glycol	75
Butyl glycol	75
Acetone	20
Me Et ketone	42
Dichloromethane	0.9
Trichloroethylene	7
Perchloroethylene	1
Tetrahydrofuran	70
Dioxane	—
Ethyl ether	60
Methyl acetate	3
Ethyl acetate	20
n-Butyl acetate	45
Dimethylacetamide	—
Acetonitrile	—
Pyridine	—

Low concentrations of chlorine-containing solvents can make water unfit for drinking. *See* WATER POLLUTION.

Usage. Global restrictions are causing a change in use of organic solvents away from the less effective and more polluting hydrocarbon solvents toward the oxygenated ones. Since the latter cost more and have a lower heating value, there is more incentive to recover rather than to incinerate solvents used in industry. *See* INDUSTRIAL HEALTH AND SAFETY; TOXICOLOGY.

Ian McN. Smallwood

Bibliography. H. Ohtaki and Y. Yamatera, *Structure and Dynamics of Solutions*, 1992; P. Politzer and J. S. Murray, *Quantitative Treatments of Solute Solvent Interactions*, 1994; C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, 2d ed. 1988; D. J. De Renzo, *Solvents Safety Handbook*, 1986.

Solvent extraction

A technique, also called liquid extraction, for separating the components of a liquid solution. This technique depends upon the selective dissolving of one or more constituents of the solution into a suitable immiscible liquid solvent. It is particularly useful industrially for separation of the constituents of a mixture according to chemical type, especially when methods that depend upon different physical properties, such as the separation by distillation of substances of different vapor pressures, either fail entirely or become too expensive.

Industrial plants using solvent extraction require equipment for carrying out the extraction itself (extractor) and for essentially complete recovery of the solvent for reuse, usually by distillation. *See* DISTILLATION; EVAPORATION; STRIPPING (CHEMICAL ENGINEERING).

Applications. The petroleum refining industry is the largest user of extraction. In refining virtually all automobile lubricating oil, the undesirable constituents such as aromatic hydrocarbons, which have poor chemical and viscosity-temperature characteristics, are extracted from the more desirable paraffinic and naphthenic hydrocarbons. The principal solvents used are furfural, phenol, and a combination of phenol with propane and cresylic acid; nitrobenzene and 2,2′-dichloroethyl ether are used in minor amounts. Liquid propane is also used preferentially to extract the desirable constituents from unwanted asphaltic compounds.

By suitable catalytic treatment of lower boiling distillates, naphthas rich in aromatic hydrocarbons such as benzene, toluene, and the xylenes may be produced. The latter are separated from paraffinic hydrocarbons with such solvents as liquid sulfur dioxide, furfural, and ethylene glycol to produce high-purity aromatic hydrocarbons and high-octane gasoline.

Gasoline is "sweetened," or freed of its sulfur-containing compounds, by extraction with aqueous caustic solutions containing various naphthenic and aromatic acids, or methanol, to modify the solvent characteristics. Aqueous copper ammonium acetate is used to extract butadiene from other 4-carbon hydrocarbons in synthetic rubber production.

Vegetable oils are separated into relatively saturated and unsaturated glyceride esters with furfural or liquid propane as solvents. The former are edible products; the latter are drying oils used in paints. Fish oils are similarly treated and yield a high-vitamin fraction as well.

In by-product coke-oven plants, phenols and other tar acids are recovered from the ammoniacal liquors with benzene, tricresyl phosphate, butyl acetate, and other solvents in large installations. The pharmaceutical industry uses extraction to separate natural impurities or unwanted chemical by-products from products such as synthetic vitamins, penicillin, Aureomycin antihistamines, reserpine, and a host of others.

All uranium for atomic energy purposes is freed of its impurities in aqueous solution by extraction into diethyl ether, tributyl phosphate, and other solvents. The reprocessing of atomic energy fuels for the recovery of plutonium, and the separation of many of the other fission products such as the rare-earth metals, utilize solvent extraction extensively. The otherwise hard-to-separate metal pairs, zirconium-hafnium and niobium-tantalum, are separated in quantity with comparative ease by these methods.

Equipment. Extractors bring about direct contact of the feed (solution to be separated) and extracting solvent in order to permit diffusional transfer of the constituents from the feed to the solvent. The rate of transfer depends upon the contact area of the two liquids and the degree of turbulence developed within them. The extractor disperses one of the

liquids in the other to produce large surface area, and relative motion to produce turbulence. The extractor must also provide for the subsequent mechanical separation of the dispersion, based upon the different densities of the liquids, to permit withdrawal of the two effluent products, the extract (solvent containing the extracted constituents) and the raffinate (unextracted residue).

Mixer-settlers (**Fig. 1**) provide for these requirements in separate vessels. The feed and solvent flow continuously through the mixer, in which the rotating agitator disperses one of the liquids into small droplets immersed in the other. The size of this vessel must provide sufficient residence time for the liquids that the desired diffusional transfer occurs. The degree of agitation must be intense without, however, producing so fine a dispersion that subsequent settling is difficult. The dispersion flows to the settler, most simply a drum, in which low velocity and lack of agitation promote gravity settling and coalescence of the drops to provide clear effluents.

Since in such single-stage apparatus the extractable substance approaches a concentration equilibrium in the effluents, nearly complete extraction requires a multiplicity of stages. An arrangement for countercurrent interstage flow of the liquids reduces the amount of solvent needed (**Fig. 2**). The compact modification of **Fig. 3** has found particular favor in extraction of radioactive metals from aqueous solutions in processes associated with atomic energy operations.

To reduce the floor space and pump requirements for multistage extractors, a variety of vertical towers are also used. These involve countercurrent vertical flow, under gravity, of one of the liquids in dispersed form through a continuum of the other by virtue of the different liquid densities. A packed tower

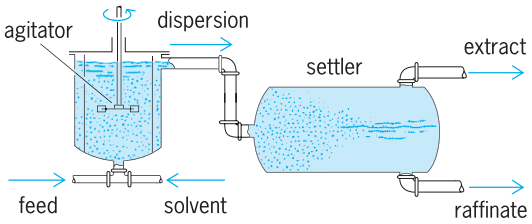


Fig. 1. Single-stage mixer-settler extractor.

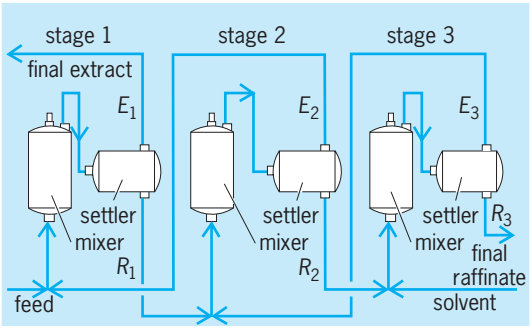


Fig. 2. Diagram of a three-stage countercurrent mixer-settler extractor. (*After R. E. Treybal, Mass Transfer Operations, 2d ed., McGraw-Hill, 1968*)

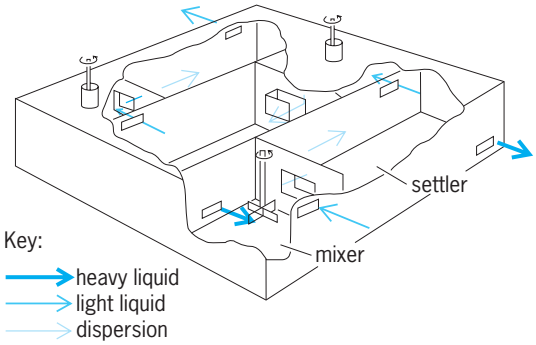


Fig. 3. Three-stage, box-type, mixer-settler extractor.

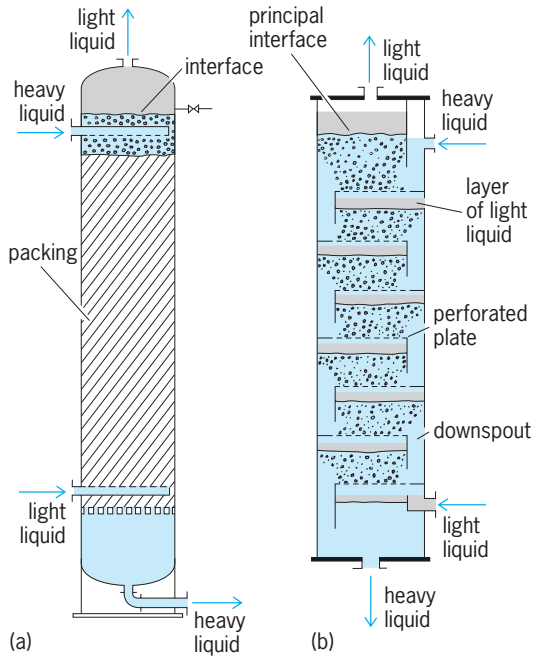


Fig. 4. Vertical tower extractors. (*a*) Packed-tower extractor. (*b*) Perforated-tray extractor. (*After R. E. Treybal, Mass Transfer Operations, 2d ed., McGraw-Hill, 1968*)

(**Fig. 4*a***) is a cylindrical shell, the bulk of which is filled with manufactured packing, such as rings or saddles, randomly arranged. The more dense liquid, introduced at the top, flows downward as a continuum. The less dense liquid enters at the bottom through small nozzles. The resulting small droplets rise through the heavy liquid, during which time extraction occurs, and then coalesce into a bulk and leave at the top. The packing serves to maintain the dispersion and provide moderate turbulence. The dispersed liquid may be either feed or solvent, light or heavy. If it is heavy, the droplets settle downward. Although the liquids are not repeatedly dispersed and settled as in the multistage mixer-settler, nevertheless multistage effects are obtained. Spray towers contain no packing and are not as effective. *See* GAS ABSORPTION OPERATIONS.

In perforated-tray towers (Fig. 4*b*) the light liquid collects in a layer under each tray and is dispersed into droplets by the small perforations. The drops rise through the heavy liquid, which flows across

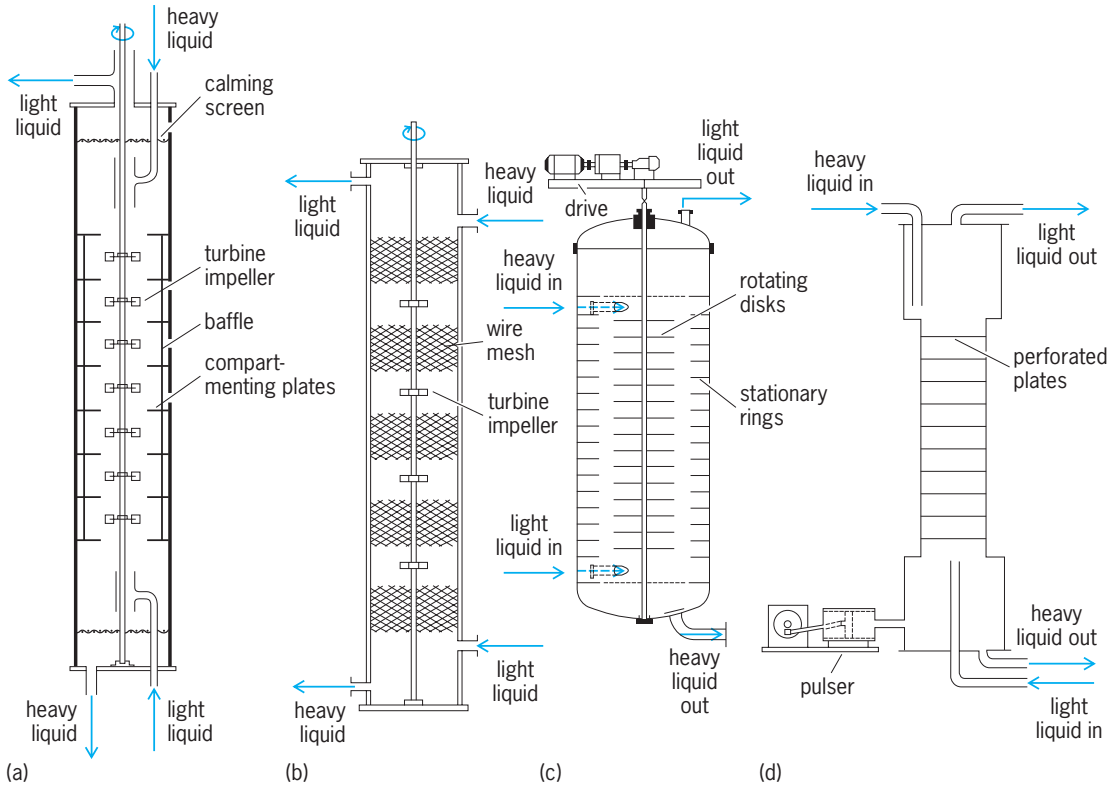


Fig. 5. Types of extractors with mechanical agitation. (*a*) Oldshue-Rushton extractor; (*b*) Scheibel-York extractor (*after R. E. Treybal, Mass Transfer Operations, 2d ed., McGraw-Hill, 1968*). (*c*) Rotating-disk extractor (*after G. H. Reman and R. B. Olney, Chem. Eng. Prog., 51:141–146, 1955*). (*d*) Pulsed extractor (*after T. B. Drew and J. W. Hoopes, eds., Advances in Chemical Engineering, vol. 1, Academic, 1956*).

each tray and through the downspouts. The frequent redispersion achieved makes these towers very effective. Alternatively, by turning the tower upside down, the heavy liquid may be dispersed.

Mechanical agitation, provided by rotating impellers as in the towers of **Fig. 5***a*, *b*, and *c*, is used to obtain finer dispersions and increased turbulence. The pulsed tower (Fig. 5*d*) provides the mechanical agitation by rapid (20–100 cycles/min), small-amplitude (0.25–2 in. or 0.64–5.0 cm), reciprocating motion superimposed upon the natural flow of liquids as they alternately pass through small perforations in the plates. This is particularly useful for handling radioactive liquids, since moving parts may be located in a place of safety.

In all these designs, the tower diameter is governed by the quantity of liquids to be handled, the height by the number of stages of extraction required. Towers up to 15 ft (4.8 m) in diameter and 125 ft (38 m) tall have been built. Auxiliary equipment may include pumps for movement of the liquids, motor drives for agitators, valves and flowmeters for control of flow rates, and liquid-level control instruments.

The centrifugal extractor (**Fig. 6**) consists of a series of perforated, concentric rings in a cylindrical drum, the whole rapidly rotated (2000–5000 rpm) on the horizontal shaft. Liquids enter and leave through the shaft; they flow radially and countercurrently in the rotating drum because the effects of density differences are increased by centrifugal force. The particular virtue of this machine is the low residence time of the liquids, which has made it especially useful in the extraction of antibiotic pharmaceuticals from fermentation broths. *See* COUNTERCURRENT TRANSFER OPERATIONS; EXTRACTION; MIXING; UNIT OPERATIONS.

Laboratory applications. Solvent extraction is carried out regularly in the laboratory by the chemist as a commonplace purification procedure in organic

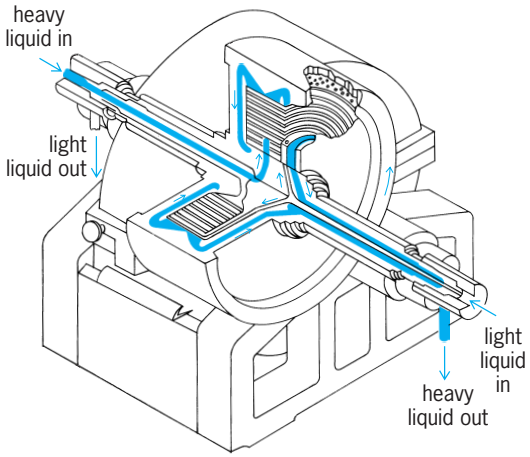


Fig. 6. Podbielniak centrifugal extractor, featuring low residence time of liquid. (*Podbielniak, Inc.*)

synthesis, and in analytical separations in which the extraordinary ability of certain solvents preferentially to remove one or more constituents from a solution quantitatively is exploited. Batch extractions of this sort, on a small scale, are usually done in separatory funnels, where the mechanical agitation is supplied by handshaking of the funnel.

Natural biological products, such as hormones, serums, vaccines, vitamins, and plant extracts, usually consist of many individual chemical compounds, frequently so similar and complex as to defy ordinary analytical procedures to separate them. Such substances, sometimes initially not even suspected of being mixtures, have been successfully separated into their constituents by the very ingenious laboratory device of L. C. Craig, which is capable of performing tens of thousands of extractions automatically.

Mechanism of extraction. Consider the single-stage extraction of a typical solute from a feed solution with a suitable immiscible solvent, as carried out in the device of Fig. 1. **Figure 7**, which is characteristic of the chemical nature of the system, shows the equilibrium distribution of solute between the two liquids as the curve *AGHB*. If two solutions whose respective solute concentrations plot as a point on this curve are brought into contact, they will remain at their initial concentrations indefinitely, and no solute transfer will occur. On the other hand, if their concentrations are such that they provide a point removed from the curve, a spontaneous transfer of solute will occur from one liquid to the other, in an effort to bring the system to equilibrium. If the feed solution of Fig. 1, flowing at rate F, is of concentration C_F (point D in Fig. 7), and the solvent, flowing at rate S, contains no solute, transfer of solute will occur from the feed to the solvent. The rate of extraction at any time is proportional to the degree of departure from equilibrium, the proportionality involving the interfacial area A and rate constants (mass-transfer coefficients k_E and k_R) characteristic of the circumstances. As the liquids enter the vessel, the initial rate of extractionis given by the equation below, where

$$\text{Rate} = k_E A \Delta C_E = k_R A \Delta C_R$$

ΔC_E and ΔC_R are the initial departures from equilibrium at G in the two liquids, as shown in Fig. 7. Because of the necessity of satisfying a material balance (amount of solute removed from the feed equals that taken up by the solvent), the concentrations within the liquids move, not toward G, but instead toward H, whose location is fixed by the relative rates of flow of liquids, as shown. As extraction proceeds, so that the liquids reach point E, the concentration departure from equilibrium is lessened ($\Delta C'_E$ and $\Delta C'_R$) and the rate of extraction is reduced. Theoretically, equilibrium for the effluents at H will never be reached since the rate will then have fallen to zero, but in practice a point J is ultimately reached which can be as close to equilibrium at H as desired, depending upon the time of residence (volume of vessel/volume rate of flow) in the vessel. The degree of approach to equilibrium actually attained is the stage efficiency of the extractor, which may be expressed as the ratio of line lengths DJ/DH.

The interfacial area A is the surface of all the droplets of dispersed liquid in the vessel. Since for reasons of cost it is desired that the rate of extraction be large and the time of residence and vessel volume small, A is made large by creating small droplets through vigorous agitation. In well-agitated vessels, A may be 1000 ft^2 or more for each cubic foot (or 3280 m^2 for each cubic meter) of liquid contents. Too vigorous agitation, which may produce such small droplets that permanent, unsettleable emulsions form, must be avoided, however. The droplets are smallest in the vicinity of the agitator owing to locally high turbulence intensity. They tend to coalesce and become larger in regions remote from the agitator.

The mass-transfer coefficients k_E and k_R are characteristic of the physical properties and the nature of the motion of the extract and raffinate liquids, respectively. If the liquids are entirely quiet, they are related to the molecular diffusion coefficient of

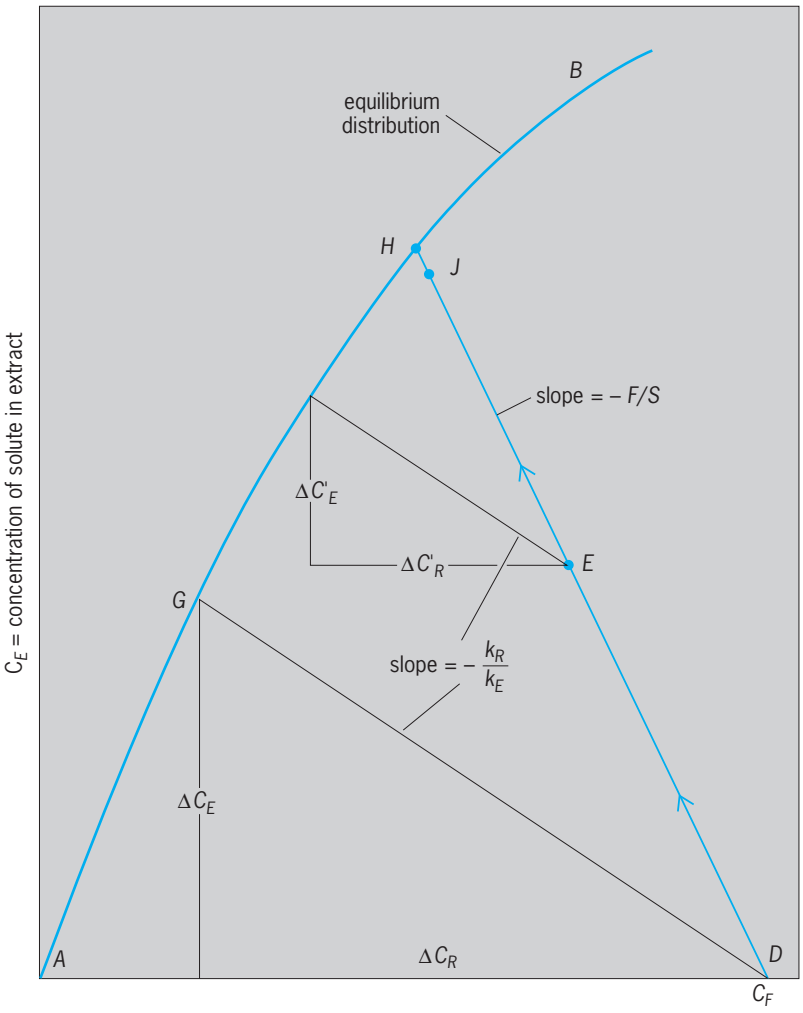


Fig. 7. Single-stage extraction. C_R = concentration of solute in raffinate.

the solute through the liquids. The transfer of solute molecules from the bulk of the raffinate, through the interface surface, and into the bulk of the extract, then depends upon the random, thermal motion of the molecules, and their passage is greatly hindered by their frequent collision with the more abundant solvent molecules. The resulting mass-transfer coefficients are then very small.

For liquids in motion, however, the coefficients are very much larger, to an extent depending upon the intensity and scale of turbulence within each liquid. Under these conditions, it is customary to ascribe the transport of solute from the bulk of the raffinate liquid to the interface surface, and ultimately into the extract, to eddies, or relatively large chunks of liquid which move rapidly, carrying with them their contained solute. The influence of the diffusion coefficients is then substantially less.

For a given degree of turbulence as produced by the agitator, a number of additional factors also have influence on the rate of extraction. These are discussed briefly below.

1. The relative motion of liquid droplets and the continuum in which they are immersed result in a transfer of momentum across the interface, and this in turn results in internal circulation of the droplet liquid, leading to relatively large mass-transfer coefficients for the dispersed liquid. The larger this internal circulation the larger the ratio of viscosities of continuous to dispersed liquid. In many systems of industrial importance, however, there are present trace concentrations of substances which tend to adsorb at the interface, producing a rigidity of the interface and a marked reduction in the internal circulation within the droplets, with consequent lowering of the mass-transfer coefficients. In one case, 6×10^{-5} g surface active agent/100 ml liquid reduced the expected rate coefficient by 68%. Other mechanisms have also been proposed to explain such a reduction. Blocking of portions of the interface by the adsorbed molecules and chemical interaction of the adsorbed substance and the extracting solute are two possible explanations.

2. The relative motion of the liquid droplets and the continuum result in unequal mass-transfer rates, and hence unequal solute concentrations, about the drop periphery. Since interfacial tension depends upon concentration, an interfacial tension gradient along the drop surface may result, which brings about a relatively violent surface movement (Marangoni effect). This in turn enhances the mass-transfer coefficient substantially.

3. When two droplets coalesce, the reduction of surface results in a substantial release of energy. This produces an increased relative motion of the droplets and the continuum, in turn leading to an increase in the mass-transfer coefficients.

4. Particularly in the extraction of metals from solution, extraction is frequently dependent upon a chemical reaction between the solute metal and a substance in the solvent, to make the metal soluble in the organic solvent. Such reactions may be relatively slow, and the rate of extraction will then be largely governed by the reaction rate.

While the mass-transfer coefficients and the stage efficiency of extraction may be computed with reasonable accuracy for ordinary circumstances, when any of the four conditions described above exist it is presently impossible to predict the rate of extraction with assurance. Resolution of these problems is the object of much current research. Robert E. Treybal

Bibliography. J. Marinsky (ed.), *Ion Exchange and Solvent Extraction*, 1995; T. Ritcey and A. W. Ashbrook, *Solvent Extraction*, pt. 1: *Principles and Applications to Process Metallurgy*, 1984; T. Sekine, *Solvent Extraction Chemistry*, 1977; R. E. Treybal, *Mass-Transfer Operations*: *Chemical Engineering*, 3d ed., 1979; A. Weissberger (ed.), *Techniques of Organic Chemistry*, 2d ed., vol. 3, 1957; J. A. Wesselingh and R. Krishna, *Mass Transfer*, 1991.

Somatic cell genetics

The study of mechanisms of inheritance in animals and plants by using cells in culture. In such cells, chromosomes and genes can be reshuffled by parasexual means rather than having to depend upon the genetic recombination and chromosome segregation that occur during the meiotic cell divisions preceding gamete formation and sexual reproduction. Genetics is concerned with the role of genes and chromosomes and their environmental interactions in the development and function of individuals and the evolution of species. Genetic analysis of complex multicellular organisms classically requires multigeneration families. Because fairly large numbers of progeny of defined matings have to be scored, genetic analysis of animals and plants with long generation times, small families, or lack of controlled matings is difficult and slow.

Somatic cell genetics circumvents many of these limitations. It has enhanced the scope and speed of genetic analysis in higher plants and animals, especially when combined with the powerful techniques of molecular biology. Cells in culture can be used to generate fertile plants and animals from single cells, including stem cells. With these methods, virtually every gene in any species of interest can be identified, its functions clarified, and the structural and functional changes it has undergone in particular phylogenetic lineages determined.

Cells in culture. Plant meristem cells can be grown readily in culture. They can be totipotent (capable of differentiating into all cell types) and can even grow into a mature plant. In mammals, the only totipotent cells, in addition to fertilized eggs, are the embryonic stem cells of early preimplantation embryos and some embryonal carcinoma cells. Even adults contain some stem cells, but these appear to be only pluripotent or even lineage-specific, able to differentiate into only a limited number of cell types. Totipotent cells appear to be immortal in culture. Other diploid cells senesce, dying after a limited

number of cell divisions. This is sufficient for chromosome studies, and hundreds of thousands of these are performed each year on normal and cancerous human cells. In addition, cells in early passage cultures of many species can be frozen and maintained in a viable state for years (cryopreservation). The cells can be thawed at any time and grown in culture. Thus, a single biopsy can provide cultured diploid cells for many years, including cells from endangered species. *See* STEM CELLS.

Normal cells in culture can be genetically and chromosomally stable, but genetic changes, or mutations, can be readily induced in them. Mutations provide the markers essential for genetic analysis. Mutations occur spontaneously, but they can be induced at much higher frequencies by a variety of physical, chemical, and living agents. Normal cells can be transformed into immortal cells (which can divide indefinitely) by exposure to certain chemicals or oncogenic viruses, such as SV40 or papillomavirus, and the transformed cells can form tumors in a suitable host. Transformed cells, like most cancer cells, are immortal. They are unstable, with both structural and numerical changes in the chromosomes. Somatic cell genetic methods have shown that many cancers are associated with specific chromosome rearrangements. *See* ANIMAL VIRUS; CHROMOSOME; CHROMOSOME ABERRATION; MUTAGENS AND CARCINOGENS; MUTATION; ONCOGENES; RECOMBINATION (GENETICS); TUMOR.

Somatic cell hybrids. Cells in culture can be induced to fuse when their cell membrane is modified by certain viruses (such as parainfluenza or measles), polyethylene glycol, or an electric shock (electroporation). After fusion of two cells, the nuclei are in a common cytoplasm. The fused cell is called a heterokaryon if the parental cells are of different origins. Heterokaryons are useful in complementation analysis, which is based on fusing cells from two individuals who have the same disease phenotype. If the two individuals have a mutation in the same gene, the heterokaryons will lack the gene product and will still show the abnormal phenotype. However, if the two individuals have mutations of different genes, each parental cell will provide the normal gene product missing in the other parent, and the heterokaryons will have a normal phenotype. Complementation analysis has shown that the human disease Fanconi's anemia, a potentially fatal blood disorder, is caused by a mutation in any one of at least six different genes.

The nuclei in a heterokaryon tend to become synchronized and enter cell division (mitosis) together. When the two nuclear membranes break down at metaphase of mitosis, a single mitotic spindle is formed, and each of the two daughter nuclei gets a complete set of chromosomes from each parental nucleus and is a true somatic cell hybrid.

Hybridomas are a special class of somatic cell hybrids. B lymphocytes, the antibody-secreting cells of the immune system, senesce very quickly in culture, but if they are hybridized with immortal malignant myeloma cells of the same species, the resultant hybridoma cells are also immortal, and they produce antibody molecules with the same unique specificity as the lymphocyte parent. These monoclonal antibodies are widely used in diagnostic studies and research. *See* ANTIBODY; ANTIGEN; MITOSIS; MONOCLONAL ANTIBODIES.

Interspecific cell hybrids, especially those involving distantly related species, tend to lose chromosomes during mitosis. Rodent–human hybrid cells tend to lose human chromosomes. In hybrids with a much reduced number of human chromosomes, any human gene still present must reside on one of the remaining human chromosomes. The analysis of rodent–human somatic cell hybrids has led to the mapping of thousands of genes and anonymous deoxyribonucleic acid (DNA) fragments. Interspecific somatic cell hybrids have been widely used to map genes in other mammalian species as well, including the owl monkey, mouse, cow, pig, and fox. More precise mapping has been achieved by in-situ hybridization of labeled DNA fragments or genes to chromosomes in normal cultured cells of the species being studied. *See* CHROMOSOME MAPPING.

Cell hybridization can lead to the extinction of phenotypic characteristics of one of the parental types. When cancer cells are fused with nonmalignant diploid cells, the resultant hybrid cells are nonmalignant. However, the cells regain their malignancy if they lose both copies of a particular chromosome derived from the normal parent. This suggests that the product of a particular gene on this chromosome is a tumor suppressor. Dozens of human tumor suppressor genes have been identified; mutations in them are important causes of cancers in many different tissues.

Extinction of one parental phenotype in somatic cell hybrids allows genetic dissection of processes such as cell differentiation and senescence. For example, rat hepatoma (liver tumor) cells express the liver-specific gene for albumin production, but when these cells are hybridized with mouse fibroblasts, albumin production is extinguished. Loss of both copies of mouse chromosome 1 from the hybrids leads to reexpression of albumin, indicating that a gene on this chromosome is responsible for blocking the expression of this liver-specific gene, and perhaps others. Fusion of mortal diploid cells with immortal chromosomally unstable cells produces hybrid cells that are mortal (senesce), perhaps because the immortal parental cells had lost both copies of a chromosome carrying a mortality gene. In fact, a gene has been identified by somatic cell genetic methods that reverses the immortal phenotype of cells into which it is inserted. *See* ALBUMIN; LIVER.

A wide variety of plant somatic cell hybrids have been made; a number of these, such as tomato-potato and carrot-parsley, are able to differentiate and form mature plants, although they have lost a few chromosomes and are sterile. Cell hybrids between more distantly related organisms, such as soybean and tobacco, lose the chromosomes of one parent quickly,

but may retain some genes derived from this parent (introgression).

Gene transfer. A gene of interest can be inserted into cells in culture by various techniques.

Viral vectors. One method is to insert the gene into an infectious virus and transfect this into cells. Cells of Fanconi's anemia complementation type E were transfected with a genomic complementary, or copy, DNA (cDNA) library in which each of the thousands of different cDNAs was present in a different copy of the viral vector. When the cells were grown in a selective medium in which survival required a functional copy of the gene, all the surviving cells had incorporated the same cDNA, thus identifying the gene. Such complementation cloning facilitates quick identification of important genes.

When the transfection vector is a retrovirus or an adenovirus, its DNA integrates into a random chromosomal site in the host cell and becomes a permanent part of the genome. If the gene of interest is expressed, it can provide long-lasting correction of a deficiency of that gene product in the host. Adenoviral vectors can enter nonproliferating cells by endocytosis (introduction of extracellular material into the cytoplasm). Because they remain extrachromosomal and thus do not replicate, they can provide gene therapy for only a limited period of time.

Nonviral methods. Genes can also be transferred into cultured cells by nonviral methods, such as precipitation of the DNA onto a monolayer of cells, fusion of DNA-containing lipid droplets with the cell membrane, insertion of an artificial chromosome containing the gene, or microinjection of the gene directly into the nucleus. When genes are introduced into plant cells from other plants, bacteria, yeast, or humans, they continue to function, reflecting the unitary origin of all living organisms and their virtually identical molecular machinery. *See* ANIMAL VIRUS; BACTERIAL GENETICS; PLANT VIRUSES AND VIROIDS.

Transgenic plants and animals. When a foreign gene (transgene) is introduced into a cell nucleus by most of the mechanisms previously described, it inserts at random into some chromosomal site, sometimes disrupting an important gene and causing deleterious effects. Fortunately, methods have been developed that target the transgene to the gene's normal chromosomal position. This has made it possible to replace one of the two copies of a gene in diploid cells with a normal or engineered copy of the gene from an external source.

Plants. Transgenic methods have tremendous potential for the development of plants resistant to insects, microorganisms, or environmental stress. Corn and cotton containing a transgene for a toxin produced by the bacterium *Bacillus thuringiensis* have greatly increased resistance to specific insect pests and led to higher crop yields. Tobacco or tomato callus cells containing a transgene for the antiviral protein of pokeweed (*Phytolacca americana*) have grown into mature transgenic plants that are resistant to a broad spectrum of viruses. Transgenic tobacco plants containing the gene for viral hepatitis B surface antigen produce large amounts of this protein; such transgenic plants may be useful for low-cost vaccine production. *See* IMMUNOGENETICS.

Animals. Transgenic animals are generated in ever-increasing numbers. Transgenic livestock can produce human proteins that may find broad use as medical drugs. Gene transfer techniques are being used to speed the growth of farmed tilapia fish and to reduce the threat to wild populations of tilapia by generating sterile variants. Increasing numbers of transgenic mice are being produced to study gene function, cell differentiation, and morphogenesis and to create animal models of human diseases whose study may speed the development of treatments.

Some transgenic animals are generated by microinjection into the early embryo, but a growing number are produced by targeted gene transfer into cultured embryonic stem cells. The role of a specific gene in development can be studied by replacing one copy of the normal gene with a nonfunctional copy by using gene targeting in embryonic stem cells, and introducing the resultant heterozygous cells into the blastocyst cavity of a developing embryo of the same species. Upon implantation into a pseudopregnant uterus, the preimplantation embryo can develop into a chimeric adult containing both normal host cells and heterozygous transgenic cells. If transgenic cells are present in the germ line, they can produce haploid sperm or ova containing only the nonfunctioning transgene, and progeny produced by these will be heterozygous for the transgene. One fourth of the zygotes produced by a mating of two individuals heterozygous for the transgene should be homozygotes with two copies of the nonfunctional transgene. The effect of such gene knockouts can be assessed at any stage of embryogenesis or at birth, or even later, if the homozygotes are viable. *See* CHIMERA.

Transgenic mice, including gene knockouts, have been produced in an attempt to understand the roles of many different kinds of genes in normal and abnormal development, including cancer. For example, the connexin proteins are important structural components of gap junctions, the channels between some adjacent cells that facilitate the passage of important low-molecular-weight molecules. Knockout mice missing one of these connexins die soon after birth because of the unexpected occurrence of a congenital heart malformation. Transgenic studies have shown that gap junctions are essential for the body's immunological defense against viral and other infections.

Another use of transgenic animals is to introduce a gene modified to express a reporter molecule that is easily detected by its color or fluorescence. This enables the identification and mapping of the cells that express the gene at specific stages in embryonic development or in the adult organism. Transgenic mice with a single copy of a rat heat shock protein gene that was engineered to be highly expressed where the actin gene is expressed normally produce a high level of the heat shock protein in heart muscle and

are protected against heart damage due to ischemia (reduced blood flow).

Implications. Somatic cell genetics has revolutionized genetic analysis in animals and plants. It is providing fundamental insights into developmental biology, and is increasingly important in the identification and characterization of certain genes in key species.

The complete nucleotide sequence of the DNA of a number of animals and plants has been determined in the last few years, and the number is growing rapidly. Knowledge of the complete DNA sequence of several organisms has led to considerably improved estimates of the total number of genes in each [for example, roughly 14,000 genes in the fruit fly, 18,000 genes in the silkworm (*Bombyx mori*), and 22,000 genes in humans]. Somatic cell genetic methods will play a major role in determining the functions of the enormous number of hitherto unknown genes revealed by sequencing complete genomes. In clinical medicine, this will lead to better methods of diagnosis and treatment of genetic diseases. The implications for agriculture and animal husbandry are equally exciting. *See* GENE; GENE ACTION; GENETIC ENGINEERING; GENETICS; HUMAN GENOME PROJECT.

Orlando J. Miller

Bibliography. M. J. Bertram et al., Identification of a gene that reverses the immortal phenotype of a subset of cells and is a member of a novel family of transcription factor-like genes, *Mol. Cell. Biol.*, 19:1479–1485, 1999; M. R. Capecchi, Targeted gene replacement, *Sci. Amer.*, 270(3):52–59, 1994; N. McLean et al., Transgenic tilapia and the tilapia genome, *Gene*, 295:265–277, 2002; J. E. Meija et al., Functional complementation of a genetic deficiency with human artificial chromosomes, *Amer. J. Human Genet.*, 69:315–326, 2001; M. Qaim and D. Zilberman, Yield effects of genetically modified crops in developing countries, *Science*, 299:900–902, 2003.

Somatostatin

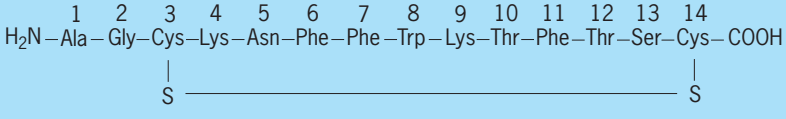
A naturally occurring regulatory peptide that carries out numerous functions in the human body, including the inhibition of growth hormone secretion from the anterior pituitary gland. Somatostatin consists of 14 amino acids (see **illus.**); the two cysteine residues are joined by a disulfide bond so that the peptide forms a ring structure. A larger variant of this peptide, called somatostatin-28, is produced in some cells and has an additional 14 amino acids attached at the amino-terminal end of normal somatostatin (somatostatin-14).

Somatostatin acts primarily as a negative regulator of a variety of different cell types, blocking processes such as cell secretion, cell growth, and smooth muscle contraction. This peptide was first discovered in the hypothalamus, a region of the brain known to control the release of hormones from the anterior pituitary gland. Somatostatin is secreted from the hypothalamus into the portal circulation and travels to the anterior pituitary gland, where it inhibits the production and release of both growth hormone and thyroid-stimulating hormone.

Many tissues other than the hypothalamus contain somatostatin, suggesting that this peptide has numerous roles. Somatostatin is present in widespread areas of the nervous system, including the cerebral cortex, basal ganglia, brainstem, spinal cord, and spinal sensory ganglia. There is evidence that it acts as a neurotransmitter or neuromodulator, providing chemical communication between nerve cells involved in the perception of sensory stimuli. Somatostatin is also present in high concentrations in the δ cells of the pancreas and has the ability to inhibit the release of both pancreatic insulin and glucagon. Somatostatin is widely distributed in the gastrointestinal tract, where it inhibits the secretion of several gut hormones, gastric acid, and pancreatic electrolytes and digestive enzymes. It inhibits the contraction of the stomach, intestine, and gallbladder, and reduces absorption of nutrients. Somatostatin also has a role in regulating the immune system, where it inhibits inflammatory processes.

Each of the functions of somatostatin is initiated by the binding of the peptide to one or more of five different cell-surface receptor proteins. These five receptor proteins have similar structures; each passes through the cell membrane seven times. Binding of somatostatin to these receptors activates one or more intracellular G-proteins and initiates biochemical signaling pathways within the cell. *See* SIGNAL TRANSDUCTION.

Analogs of somatostatin have been synthesized that are smaller, more potent, longer-lasting, and more specific in their biological effects than natural somatostatin. Some of these analogs, for example octreotide and lanreotide, have become useful as drugs. They are used in treating acromegaly because they can inhibit the excessive secretion of growth hormone by pituitary tumors (which causes this disease) and may even shrink the tumors in some patients. Somatostatin analogs are also used to treat malignant carcinoid syndrome, vasoactive intestinal



Amino acid sequence of somatostatin, including alanine (Ala), glycine (Gly), cysteine (Cys), lysine (Lys), asparagine (Asn), phenylalanine (Phe), tryptophan (Trp), threonine (Thr), and serine (Ser).

peptide (VIP)–secreting tumors, and other endocrine and neuroendocrine tumors that oversecrete hormones. Radiolabeled somatostatin analogs can be used to both localize and destroy tumors that produce an excessive number of somatostatin receptors. Newer analogs that specifically bind to just one of the five different somatostatin receptors should have more selective therapeutic responses with reduced side effects. *See* ENDOCRINE SYSTEM (VERTEBRATE); HORMONE; NEUROSECRETION; PITUITARY GLAND.

William H. Simmons

Bibliography. P. Chanson, Somatostatin analogs in the treatment of acromegaly: The choice is now possible, *Eur. J. Endocrinol.*, 143:573–575, 2000; J. Epelbaum et al., The neurobiology of somatostatin, *Crit. Rev. Neurobiol.*, 8:25–44, 1994; E. P. Krenning et al., Radiolabelled somatostatin analogue(s) for peptide receptor scintigraphy and radionuclide therapy, *Ann. Oncol.*, 10(suppl. 2):S23–S29, 1999; Y. C. Patel, Somatostatin and its receptor family, *Frontiers Neuroendocrinol.*, 20:157–198, 1999; Z. Tulassay, Somatostatin and the gastrointestinal tract, *Scand. J. Gastroenterol.*, 33(suppl. 228):115–121, 1998.

Somesthesis

A general term for the somatic sensibilities aroused by stimulation of bodily tissues such as the skin, muscles, tendons, joints, and the viscera. Six primary qualities of somatic sensation are commonly recognized: touch-pressure (including temporal variations such as vibration), warmth, coolness, pain, itch, and the position and movement of the joints.

These basic sensory qualities exist because each is served by a different set of sensory receptors (the sensory endings of certain peripheral nerve fibers) which differ not only in their sensitivities to different types of stimuli, but also in their connections to structures within the central nervous system. Information is transmitted in the form of nerve impulses from the receptor endings to the central nervous system over the peripheral nerve fibers—the latter converging as spinal nerves and entering the central nervous system via the dorsal roots. The cell bodies of these peripheral nerve fibers are in dorsal root ganglia, located just outside the central nervous system. Fibers from regions in the limb or trunk enter the spinal cord, while those from the head and face enter the brainstem. The area of skin and deep tissues supplied by fibers from a single dorsal root is called a dermatome (**Fig. 1**). There is a pair of left and right dorsal roots for each cord segment. There is considerable overlap between cutaneous areas supplied by adjacent dermatomes, since a peripheral nerve projects into several roots. Thus the sensory loss resulting from transection of a peripheral nerve may be clearly outlined and complete in the tissue supplied by the nerve, but the loss following transection of a dorsal root is usually less noticeable. Usually three or more contiguous roots must be cut before there is a region of complete anesthesia. *See* CENTRAL NERVOUS SYSTEM.

Somatic pathways. The somatic sensory pathways are dual in nature. One major part, the lemniscal system, receives input from large-diameter myelinated peripheral nerve fibers (for example, those serving the sense of touch-pressure). Those originating in the trunk or limbs have branches that terminate on cells located within the dorsal horn of the spinal cord, and also branches that ascend within the dorsal columns to terminate on cells in the dorsal column nuclei (**Fig. 2**). Cells originating in these nuclei send fibers (axons) across the midline to the other side to project via the medial lemniscus to the ventrobasal nuclear complex in the thalamus. Thalamic cells send their axons to the primary somatosensory cortex, located in the postcentral gyrus and in the depths of the central sulcus. Axons of certain cells in the primary somatosensory cortex project to other cortical areas, including the secondary somatic sensory cortex, located in the upper bank of the lateral sulcus. Peripheral nerve fibers serving touch-pressure in the face and head terminate in the main sensory nucleus in the brainstem. Cells there project across the midline to ascend with the medial lemniscus to the thalamus.

The second major somatic pathway is called the anterolateral system. It receives input from small-diameter myelinated and unmyelinated peripheral nerve fibers carrying pain and temperature information. Those fibers from the limb and trunk terminate on cells within the dorsal horn of the spinal cord. Cells of origin there send their axons across the midline to ascend within the ventral (anterior) lateral part of the spinal cord called the anterolateral tract. These axons project upon areas in the brainstem, midbrain, and thalamus. Peripheral nerve fibers from the head and face serving pain and temperature terminate in the nucleus caudalis in the brainstem. Cells of origin there send their axons across the midline to project to areas in the brainstem, midbrain, and thalamus.

There are several differences between the anterolateral and lemniscal systems besides differences in anatomical projections and the sensory modalities they serve. These are revealed by electrophysiologically recording nerve impulses from single cells or by electrically stimulating local regions in the brain and spinal cords of animals or humans. The results of such studies indicate, first, that there is a precise topographical representation of the body form at each anatomical stage within the lemniscal system (for example, for the primary somatosensory cortex see **Fig. 3**), whereas that for the anterolateral system is less well organized. Second, the peripheral area of tissue projecting to each cell is small and continuous for the lemniscal system, but often large and discontinuous for the anterolateral system, indicating a better degree of spatial resolution in the lemniscal system. Third, the capacities of cells to respond to temporal variations in stimulation applied to the body (for example, the dynamic responses of certain cells to

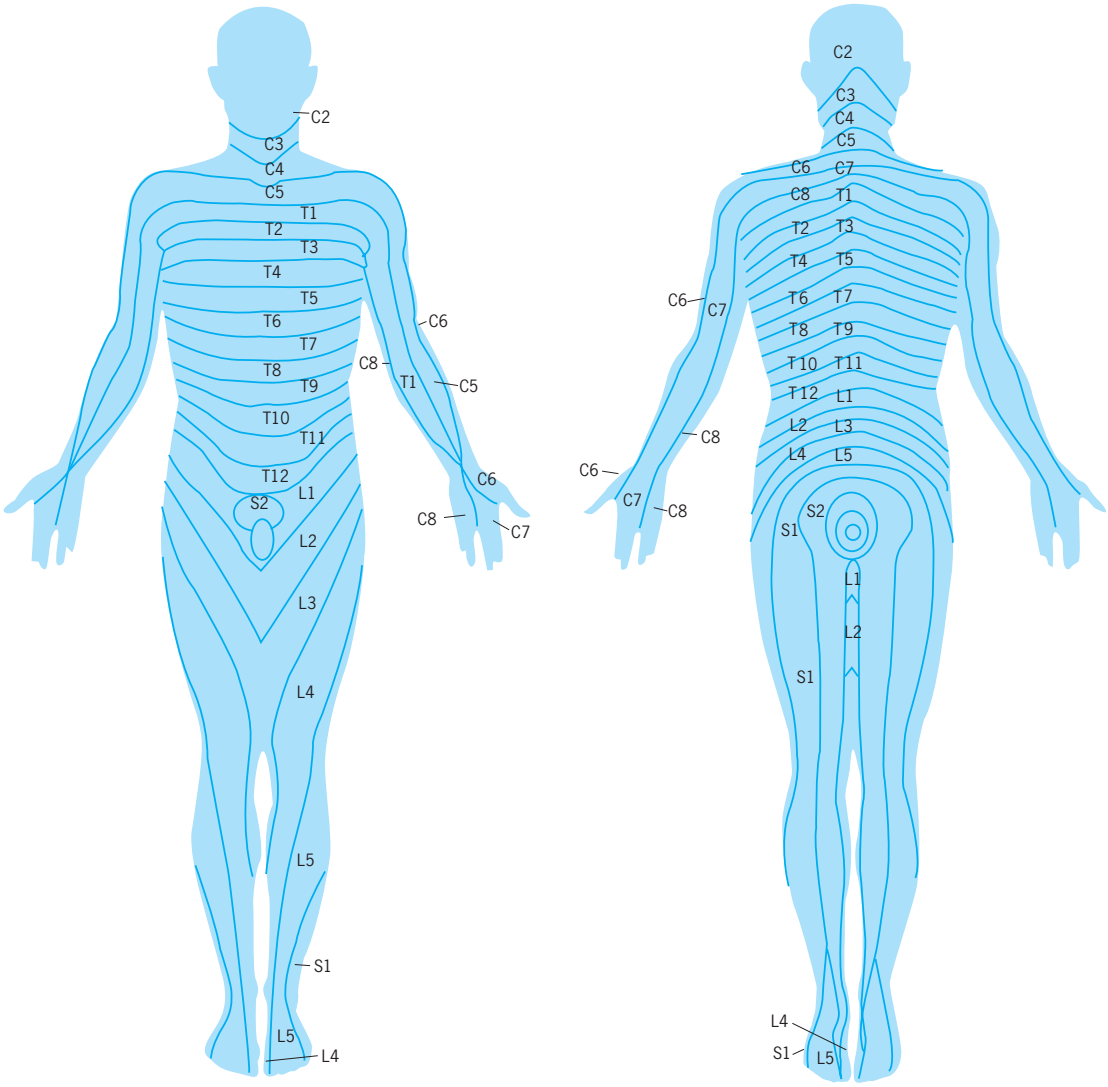


Fig. 1. Human dermatomes. The letters indicate the regions of the spinal cord: C, cervical; T, thoracic; L, lumbar; S, sacral. The numbers refer to segments of the spinal cord. (*After F. H. Netter, The CIBA Collection of Medical Illustrations, vol. 1: Nervous System, CIBA, 1967*)

a vibratory stimulus) are much more precise in the lemniscal system. These properties may reflect the greater requirement for temporal and spatial resolution in the touch-pressure sense, such as that needed to feel the spatial pattern of a finely textured cloth, than in the pain or temperature senses.

Somatosensory cortex. The functional organization of the somatosensory cortex occurs not only in the form of a representation of body geometry along its surface (Fig. 3), but also in the form of a vertical or columnar organization. All neurons encountered in a recording microelectrode penetration perpendicular to the cortical surface respond to stimulation of the same restricted area of the body. Further, all cells within a given column respond to the same type of stimulus, such as steady pressure applied to the skin or to movement of a joint or to warming the skin. Thus each column serves to process information of a particular stimulus modality arising from a specific location within or on the body.

Studies of the somatic sensory cortex have revealed that the somatotopic maps of the body are not static but can change as a result of sensory deprivation or stimulation. For example, after the amputation of the middle finger of one hand, the cells in the primary somatosensory cortex that had originally responded only to tactile stimulation of this digit become responsive to stimulation of the adjacent digits. Conversely, if an intact middle finger receives greater tactile stimulation than the adjacent fingers over an extended period of time, the cortical representation of this finger expands while that of each adjacent digit becomes smaller. Such mechanisms might contribute to the recovery of sensory function after a stroke (loss of blood supply to a local area resulting in cell death).

The reorganization of the sensory cortical map after sensory deprivation provides one possible explanation of why sensations from a body part, particularly a limb, may occur even in the absence of

the part. Sensations of the presence of a phantom limb in the place of a limb that was amputated, was traumatically deprived of its nerve supply, or was missing at birth are not uncommon. The phantom limb can appear quite lifelike and actually behave in a normal way, for example moving in coordination with an intact limb. Possibly, cells in primary and secondary somatosensory cortex that are denied the sensory input from the missing limb become endogenously active and are activated by input from adjacent, intact regions of the body, thereby giving rise to the phantom. It is not uncommon after the amputation of a limb for the phantom limb to be chronically painful. This medical condition is difficult to treat.

Lesions. Lesions (tissue destruction) of the anterolateral system may result in defects in pain and temperature sensibilities, with only mild losses in touch-pressure. A full surgical transection of the anterolateral column on one side of the spinal cord produces

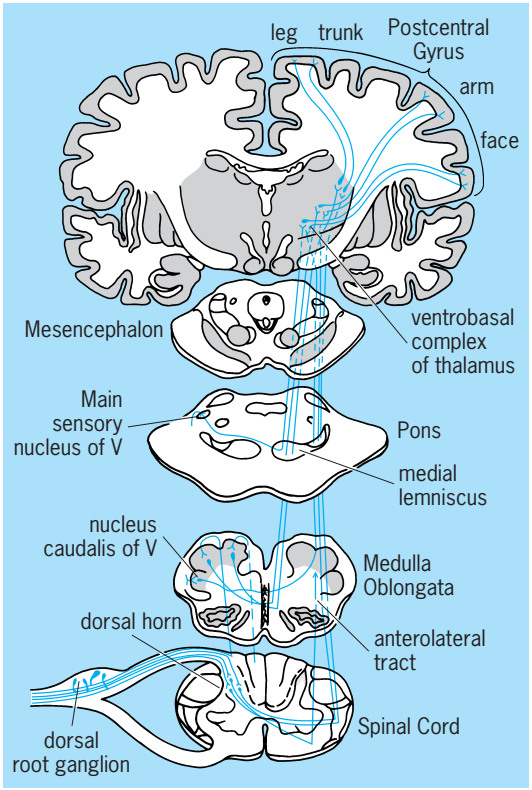


Fig. 2. Somatic afferent pathways. (*After A. Brodal, Neurological Anatomy in Relation to Clinical Medicine, Oxford University Press, 1969*)

a loss in sensitivity to pain and temperature stimuli on the opposite side of the body (whereas injury to the dorsal columns does not produce any such loss). Lesions of the lemniscal system, such as removal of SI, result in sensory defects in tactile discrimination, tactile localization, stereognosis (the perception of the shapes and sizes of objects), and the sense of

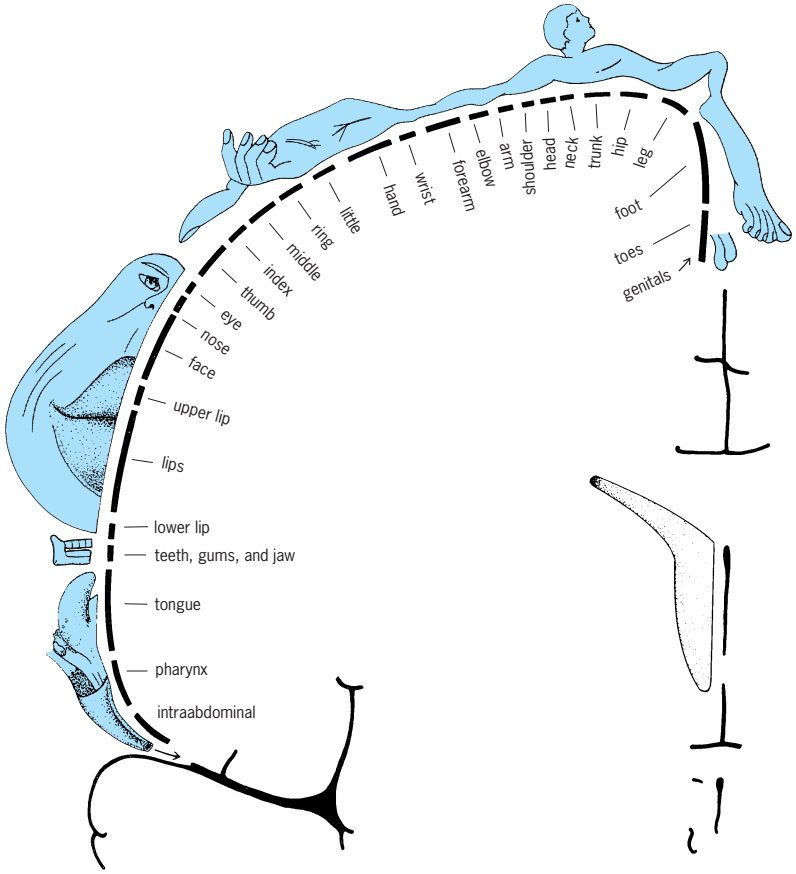


Fig. 3. Map of somesthetic areas in the human postcentral gyrus as determined by electrical stimulation (without anesthesia) and recording the subject's verbal description of his or her sensory experiences. The size of each part of the figurine is drawn approximately in proportion to the extent of sensory cortex devoted to it. (*After W. Penfield and T. Rasmussen, The Cerebral Cortex of Man, Hafner, 1968*)

position and movement of the limbs. *See* CUTANEOUS SENSATION; PAIN; PROPRIOCEPTION; PAIN.

Robert La Motte

Bibliography. E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 4th ed., 2000; M. M. Merzenich et al., Somatosensory cortical map changes following digit amputation in adult monkeys, *J. Comp. Neurol.*, 224:591–605, 1994; R. F. Schmidt and G. Thews (eds.), *Human Physiology*, 2d ed., 1989.

Sonar

A remote sensing technique or device that uses sound waves to detect, locate, and sometimes identify objects in water. The term is an acronym for sound navigation and ranging. There are many applications, using a wide variety of equipment. Naval uses include detection of submarines, sea mines, torpedoes, and swimmers; torpedo guidance; acoustic mines; and navigation. Civilian uses include determining water depth; finding fish; mapping the ocean floor; locating various objects in the ocean, such as pipelines, wellheads, wrecks, and obstacles to navigation; measuring water current profiles; and

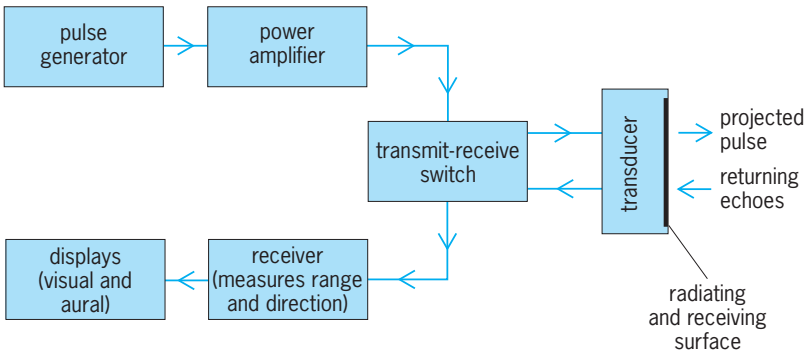


Fig. 1. Components of an elementary single-beam pulse sonar.

determining characteristics of ocean bottom sediments. Sound waves rather than electromagnetic waves (for example, radar and light) are used in these applications because their attenuation in seawater is much less. Some marine mammals use sound waves to find food and to navigate. *See* ACOUSTIC MINE; ACOUSTIC TORPEDO; ANTISUBMARINE WARFARE; ECHOLOCATION; MARINE GEOLOGY; MARINE NAVIGATION; UNDERWATER NAVIGATION; UNDERWATER SOUND.

There are two generic types of sonar: active (echolocation) and passive. An active sonar projects a signal (typically a short pulse of sound) into the water in a narrow beam, which propagates at a speed of about 1500 m/s (5000 ft/s). If there is an object (target) in the beam, it reflects a fraction of the sound energy to the sonar, which detects the echo. By measuring the elapsed time between projection and reception, the range to the target can be computed (range = sound speed × travel time ÷ 2). Direction to the target is determined from the orientation of the sound beam at the time of reception. Passive sonar does not radiate sound but depends on detecting sounds radiated by targets such as submarines and ships. Passive sonar determines direction to a target in the same manner as active sonar, but range determination is more difficult.

Active sonar. In an elementary active pulse sonar (**Fig. 1**), a pulse signal of certain frequency and duration is generated, amplified, and sent to an electroacoustic transducer, which converts the electrical signal into a sound signal, which then radiates into the water. If the transducer is reciprocal in character (typically the case), it also can be used to sense (detect) the returning echoes. The receiver amplifies the weak echoes and measures the range to each target, as well as the orientation of the receiving beam at the time of reception. This information is displayed in some form of range-direction plot.

Most active sonar transducers are mounted on the hulls of submarines or near the keels of surface ships. Sometimes, transducers are towed at a water depth that provides better operation. There are three basic transducer orientations. In the conventional depth sounder, the sound beam is directed downward. Echoes are reflected from the ocean bottom (and from fish that may be in the beam), and the depth of the ocean beneath the sonar can be determined. In the side-scan sonar configuration, the beam is oriented to the side of the ship (normal to the direction of travel) and (usually) slightly downward (**Fig. 2**). As the ship moves forward, a volume of water to the side of the ship is searched. Generally, two sonars are used, one searching to the right and one to the left. Side-scan sonars are well suited to search at a constant speed and along straight lines, such as in mapping the ocean bottom and in general searches of an area. The third, and most popular, sonar configuration involves rotating the sound beam about the vertical axis to search (scan) a sector of the water centered on the sonar platform. *See* ECHO SOUNDER.

Active sonar parameters. Sonar parameters such as operating frequency, bandwidth, beam widths, pulse duration, projected power level, and receiver sensitivity are interdependent. Ideally, the design goals are to maximize search rate, detection probability, accuracy of target location, and correctness of target classification. However, since these are conflicting goals, the parameters that optimize overall sonar performance for a specific application must be selected.

A restraint on all sonars is the speed of sound in the water. For example, if an elementary single-beam pulse sonar, known as a searchlight sonar, has a maximum search range of 6000 m (20,000 ft) and a beam width of 5°, then it takes 8 s to search a 5° sector (one beam width), and the beam cannot be rotated to search an adjacent volume until that

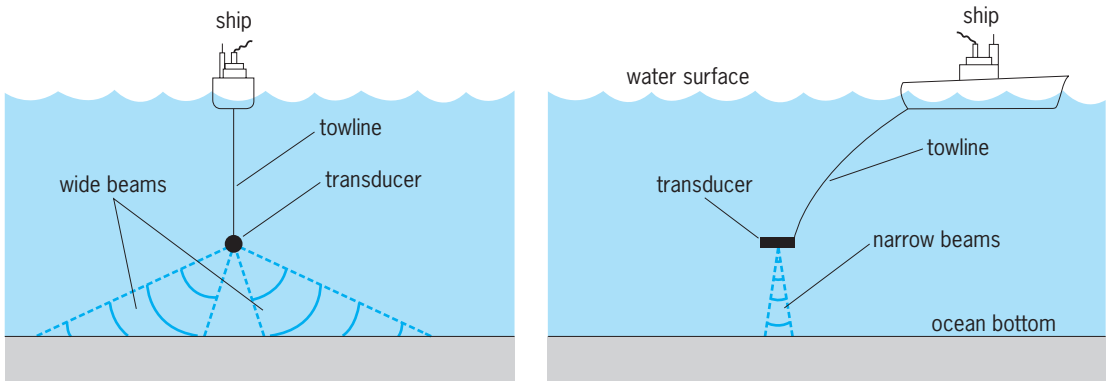


Fig. 2. Configuration for a two-beam side-scan sonar for bottom mapping. (*a*) End-on view. (*b*) Side view.

time has elapsed. About 9.6 min is required for a full 360° search in the horizontal plane. Most early active sonars were of the single-beam type, and many current small sonars still are. However, during and since World War II, much work has been done to increase search rates. The most popular technique is to project sound into a wide sector and to cover this sector with a fan of adjacent narrow receiving beams. In the example, 72 beams would be required. Rather than have 72 separate receivers, the output of each receiver beam is sampled sequentially by an electronic scanning switch and passed to a single receiver. In order not to leave gaps in the searched area, all 72 beams must be sampled during the pulse duration. If the sonar frequency is 15 kHz, the pulse needs to be at least 5 milliseconds or 75 cycles in duration in order to have at least one full cycle for each sample. An equivalent method is to electronically rotate (using phase-shift modulation) a single receiving beam through the search sector. Modern sonars, including advanced side-scan and depth sounders, employ large numbers of beams. *See* PHASE MODULATION.

Transducers. For radiation the transducer is called a projector (equivalent to an underwater loudspeaker), and for reception it is called a hydrophone (equivalent to an underwater microphone). Most electroacoustic transducers are reciprocal, meaning that they can be used for either function. The most popular type of electroacoustic transducer consists of a piezoelectric ceramic, such as barium titanate. For projection an alternating electric voltage is applied to electrodes on the ceramic, causing it to vibrate at the same frequency. For reception the alternating pressure of the echo causes the ceramic to generate an alternating electrical signal. Most practical transducers are formed from a number of small piezoelectric elements rather than a single large element and are called transducer arrays. By using many small elements, very large arrays can be fabricated in a variety of shapes (planes, cylinders, and spheres). Equally important, the elements can be connected in a variety of ways to form multiple beams simultaneously, some for projection and some for reception. *See* PIEZOELECTRICITY; TRANSDUCER.

The size and shape of the transducer, together with the sonar frequency or wavelength, determine the size and shape of the sound beam, both for projection and reception. For a given frequency the beam width decreases as the transducer size increases, and for a given transducer size the beam width decreases as the frequency increases. If the sonar is designed to search simultaneously in all directions in the horizontal plane, a suitable transducer might be a cylindrical array in which all elements are used to project, while groups of elements are connected to form multiple narrow beams for reception. Inhomogeneities in the ocean cause fluctuations in the speed of sound, which result in distortion of both the outgoing and returning sound waves and place an ultimate lower limit on the beam width; however, sonars with beams as narrow as 0.1° are generally feasible.

Frequency. The primary factor in selecting an operating sonar frequency is absorption of the sound energy by the water medium. This effect increases rapidly with increasing frequency (approximately as the frequency squared). Sonars used to locate submarines operate from about 3.5 to 35 kHz in order to achieve detection ranges of the order of 10 km (6 mi), and sometimes much greater depending on propagation conditions. Mine-hunting sonars must locate much smaller targets and operate at frequencies from about 100 kHz to 1.0 MHz, with ranges of a few hundred meters or less. Most other active sonars operate at frequencies within these extremes. *See* SOUND ABSORPTION.

Range. The range at which a target can be detected depends on the strength of the projected signal (source level), propagation losses to the target and return, reflection characteristics of the target (target strength), and sensitivity of the receiver. Also, the target echo must be stronger than various masking signals (noise and reverberation), which also are received by the sonar.

The ocean medium places an upper limit to the power that can be projected by a transducer. At low frequencies and shallow water depths, the limit likely will be due to cavitation bubbles, which essentially destroy the sound beam through absorption and scattering near the transducer. At high frequencies, the limit likely will be set by nonlinear characteristics of the water, which serve to convert any increase in power level into harmonics and heat. There is also a limit to the power that a transducer can handle without overheating or fracturing. *See* CAVITATION; NONLINEAR ACOUSTICS.

Propagation losses include absorption (discussed above) and scattering of the sound out of the beam because of inhomogeneities in the water. Finally, the sound beam spreads spherically as it travels from the projector and on the return path. This loss in intensity increases with the square of range for each path or as the fourth power of range for the round trip. For some portions of some propagation paths, the spreading is cylindrical rather than spherical, resulting in a loss proportional to the square of round-trip range. *See* SOFAR; SOUND.

Only a small fraction of the sound energy that strikes a target is reflected to the sonar. The rest is scattered in other directions or absorbed by the target. The ratio of reflected to incident intensity, known as target strength, generally increases with increasing target size, but also depends on target shape and composition, and is usually strongly dependent on target orientation, sometimes varying by a factor of 1000 depending on viewing angle.

The larger the receiving transducer, the more energy it intercepts from the returning wave, and thus the higher the sensitivity. However, the transducer also receives unwanted signals that tend to mask the desired ones. Ambient noise is a composite due to breaking surface waves, marine animals, seismic activity, and shipping. In general, ambient noise decreases with increasing frequency until thermal molecular noise dominates. Self-noise is

ship-generated noise, including turbulence as water flows past the transducer and ship.

Finally, the often dominant source of masking is the composite of unwanted echoes, known as reverberation, from inhomogeneities in the ocean volume such as animals, plants, and thermoclines; from the sea surface; and from the ocean floor. Generally, reverberation increases with frequency, but there are important exceptions. A reduction in receiver beam width results in a reduction in ambient noise and reverberation and an increase in receiver sensitivity, while a reduction in projector beam width results in an increase in source level.

Resolution. An important sonar performance characteristic is its ability to locate a target accurately and determine whether an echo is from a single target or from several targets close together. The uncertainty in the direction to the target is approximately the width of the receiving beam. Also, two targets at the same range can be resolved only if the receive beam width is less than the angle between the targets. Range accuracy and range resolution depend on the duration of the projected pulse. Two targets in the same direction at different ranges can be resolved if the separation is greater than one-half of the range extent of the pulse. There is a lower limit on pulse duration because the pulse must contain one or more full cycles of the sonar frequency (and many cycles for electronic scanning sonars). Also, the bandwidths of the receiver must equal the reciprocal of the pulse duration in order to receive the echoes without distortion. A wider bandwidth results in greater ambient noise. *See* BANDWIDTH REQUIREMENTS (COMMUNICATIONS).

Both range resolution and angular resolution are constant with range, but because of spherical spreading the beam width (in linear measure) increases linearly with range. At most operational ranges, range resolution is much better than cross-range resolution. If the resolution cell (range and cross-range) is small compared to the target, it is feasible to display a crude image of the target. Further, a target resting on or near the ocean floor blocks the sound beam and produces a reverberation-free acoustic shadow in back of the target. Some high-resolution sonars use target and shadow imaging for target classification.

Signal processing. There are many techniques to improve target detection against the masking of noise and reverberation, to increase detection ranges, and to improve target classification or identification. These include using complex pulses (for example, with shape and frequency modulation) and a wide variety of spatial and temporal frequency filters. Computer-aided detection, tracking, and target classification systems are often used. An important technique uses the Doppler phenomenon, whereby the echo from a moving target is distinguished from reverberation by its frequency shift relative to the incident frequency, and the approximate speed of target approach or recession can be determined. *See* ACOUSTIC SIGNAL PROCESSING; DOPPLER EFFECT.

Passive sonar. Passive sonars are used primarily to detect submarines and, to a lesser extent, surface ships. Because passive sonar does not radiate any sound that would reveal its location, it is the primary sensor used by submarines. The major weakness is that it cannot directly measure range to a target. To determine target location, the sonar must take bearings on a target from different locations. At short ranges (less than d^2/λ, where d is the length of the transducer and λ is the wavelength), the approximate range can be determined by measuring the curvature of the received sound wave.

Passive sonars depend on detecting noise radiated by targets, a mixture of sounds generated by propellers and hull vibrations (caused by motors, engines, pumps, and hydrodynamic forces). The noise has a continuous spectrum and discrete tones related to rotational speeds of propellers, engines, and so forth. By analysis of the received signals the sonar often can identify the type of target. Most of the radiated energy is in the audible frequency band and decreases in intensity with increasing frequency.

Most passive sonars use large receiving transducer arrays in order to achieve high sensitivity, discriminate against ambient noise, and determine precisely the direction to a target. In submarines, these arrays may be recessed into the structure or mounted on the hull. Submarines also may tow long slender line arrays. Typically, many beams are formed simultaneously to increase the search rate. Normally, surface combat ships make little use of passive sonars because of high self-noise levels. However, some special surface ships tow very long line arrays that maintain surveillance over large ocean areas. A number of very large fixed receiving arrays, placed on the ocean floor with cables running to shore stations, constantly observe strategically important ocean areas. Detection ranges for large passive sonars vary from hundreds of kilometers against noisy targets under good conditions to a very few kilometers against quiet targets.

The sonobuoy consists of a small surface buoy with a hydrophone array suspended beneath the water. Sounds received by the array are telemetered by radio link to an aircraft overhead. (There are also active sonobuoys.) Sonobuoys have relatively short detection ranges and are used primarily in tactical situations. *See* SONOBUOY; ULTRASONICS.

Chester M. McKinney

Bibliography. P. Blondel, *The Handbook of Sidescan Sonar*, Singer Praxis Books, 2004; H. E. Edgerton, *Sonar Images*, Prentice Hall, 1986; M. P. Hayes and P. T. Gough, Synthetic aperature sonar: a maturing discipline, in *Proceedings of the Seventh European Conference on Underwater Acoustics*, Delft, The Netherlands, pp. 1101-1104, 2004; H. T. Loeser, *Sonar Engineering Handbook*, Peninsula Publishing, 1984, reprint 1993; R. O. Nielsen, *Sonar Signal Processing*, Artech House, 1991; R. J. Urick, *Principles of Underwater Sound*, 3d ed., Peninsula Publishing, 1983, reprint 1996; O. B. Wilson, *Introduction to the Theory and Design of Sonar Transducers*, Peninsula Publishing, 1985, reprint 1989.

Sonic boom

An audible sound wave generated by an object that moves faster than the speed of sound (supersonic object). The sonic boom forms because the air is pushed away faster than the air molecules can move. The displaced air becomes highly compressed and creates a very strong sound wave, referred to as a compressional head shock or bow shock. At the back of the supersonic object the air has to fill the void left as the object moves forward; in this case, the gas becomes rarefied and a rarefractional tail shock develops. These shock waves are the main components of a sonic boom, and they are generated the entire time that an object flies faster than the speed of sound, not just when it breaks the sonic barrier. *See* SHOCK WAVE.

Sonic booms may be natural or generated by human activity. A natural sonic boom is thunder, created when lightning ionizes air, which expands supersonically. Meteors can create sonic booms if they enter the atmosphere at supersonic speeds. Human sources of sonic booms include aircraft, rockets, the space shuttle during reentry, and bullets. There are no reports of supersonic travel underwater; water induces a large drag that makes the concept impractical. However, a class of torpedoes is under development which may be able to travel supersonically; the torpedo creates a supercavitation bubble around itself which significantly reduces drag. Explosions can create shock waves; however, these shock waves spread spherically, like ordinary sound waves, and are not considered to be sonic booms. *See* METEOR; NAVAL ARMAMENT; THUNDER.

Supersonic aircraft. Sonic booms are commonly associated with supersonic aircraft. The most important parameter is the Mach number of the airplane, named after the Austrian physicist Ernst Mach (1838–1916). The Mach number is defined as $M = c_{plane}/c_0$, where c_{plane} is the speed of the airplane and c_0 the speed of sound of the air. At ground level, $c_0 = 343$ m/s (1230 km/h or 770 mi/h). The first human to fly supersonically was Chuck Yeager, in a Bell X1 aircraft on October 14, 1947. The sonic barrier was broken on land by a British jet car, ThrustSSC, on October 13, 1997, just 50 years later. Two supersonic passenger airplanes have been built, the Anglo-French Concorde and the Soviet Tu-144, which is currently used only for experimental purposes. Both aircraft fly at Mach 2, that is, twice the speed of sound. The world's fastest airplane is the Lockheed SR-71 Blackbird, which was built in the early 1960s and can fly faster than Mach 3 (3220 km/h or 2000 mi/h) at an altitude over 26,000 m (85,000 ft). [The speed of sound decreases with altitude, principally because the upper atmosphere is colder. For example, the sound speed is 290 m/s (1050 km/h or 650 mi/h) at an altitude of 10,000 m (33,000 ft). Therefore, if an SR-71 were flying at 3220 km/h at sea level, it would attain only Mach 2.6.] *See* MACH NUMBER.

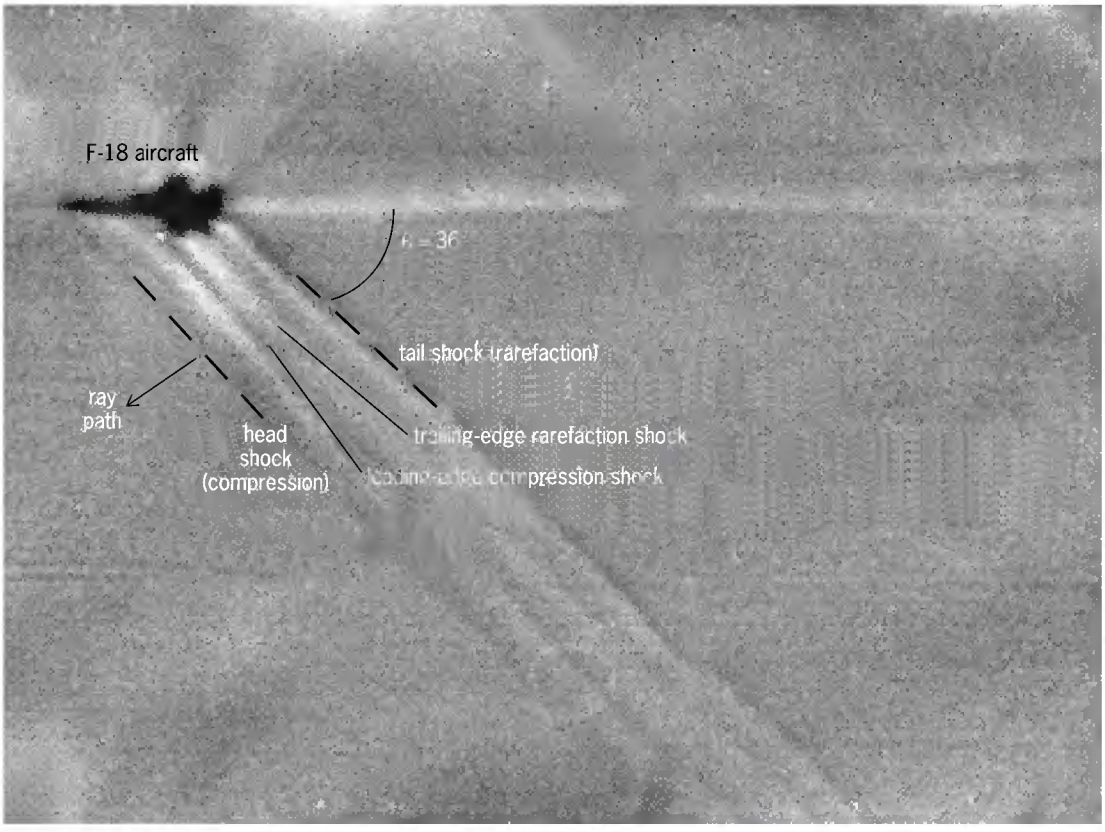


Fig. 1. Schlieren image of the shock waves generated by an F-18 aircraft flying at Mach 1.4. The Mach cone generated by the head and tail shocks can be seen as well as the shocks generated by the leading and trailing edges of the wings. (*NASA Dryden Flight Research Center*)

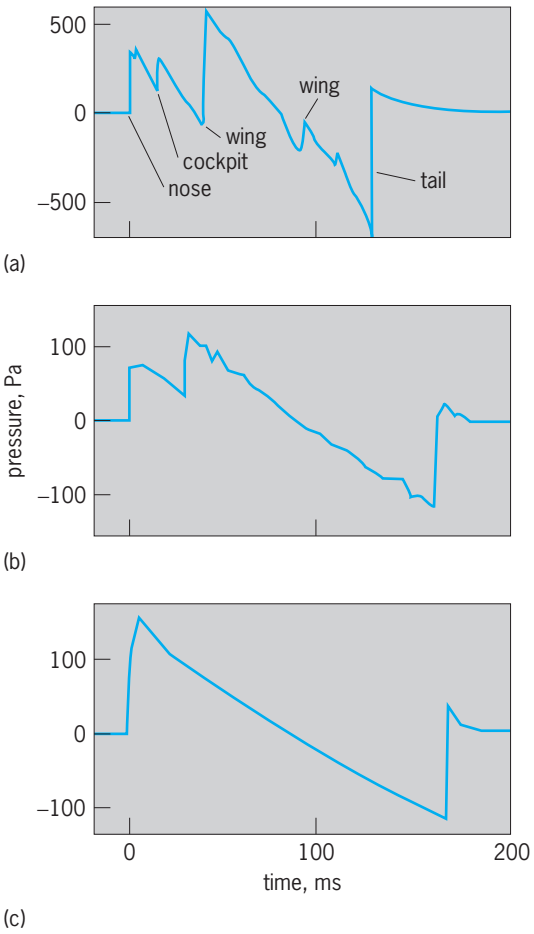


Fig. 2. Measured sonic boom waveforms from an SR-71 flying at Mach 2.4. (*a*) 165 m (540 ft) below the aircraft. Shocks associated with the nose, cockpit, leading edge of the wing, trailing edge of the wing, and the tail can be identified. (*b*) 1575 m (5075 ft) below the aircraft. Nonlinear distortion has resulted in the coalescence of some of the shocks. (*c*) At the ground, 9100 m (30,000 ft) below the aircraft, a classic N-wave signature with two shocks. (*NASA Dryden Flight Research Center*)

Mach cone. The shock waves associated with sonic booms propagate away from the aircraft in a unique fashion. These waves form a cone, called the Mach cone, that is dragged behind the aircraft. **Figure 1** shows the outline of the Mach cone generated by an F-18 fighter aircraft flying at Mach 1.4. The schlieren photographic technique was used to display the sonic boom, which is normally invisible. The half-angle of the cone is determined solely by the Mach number of the aircraft, $\theta = \arctan(1/M)$, 36° for $M = 1.4$. The shock waves travel along rays that are perpendicular to the Mach cone (Fig. 1). As the Mach number increases, θ becomes smaller and the sound travels almost directly downward. *See* SCHLIEREN PHOTOGRAPHY; SUPERSONIC FLIGHT.

Generation. The flow of air is highly complex near the aircraft body. Not only do the nose and tail of the aircraft create shock waves, but each leading edge, for example, the front of the wings, generates a compressive shock wave, and each trailing edge generates a rarefraction shock wave. The shock waves generated by the wings of the F-18 are indicated in Fig. 1.

Figure 2*a* shows a waveform measured close to an SR-71. The compressional shocks generated by the nose, cockpit, and wing can be identified, as well as the rarefraction shocks from the trailing edge of the wing and the tail. The amplitudes of the shock waves are generally larger for faster aircraft and for larger aircraft, although the cross section of the aircraft is also important in determining the shock wave's strength; long, slender planes generate weaker shock waves than short, stubby planes.

Nonlinear distortion. As the sonic boom propagates along the rays away from the aircraft, the wave changes shape because the compression and rarefraction shocks are so strong that they change the speed of sound. The change in sound speed is small, less than 1%, but it is enough that the compression shocks generated by leading edges move faster than the head shock and can catch up to it. Conversely, the rarefraction shocks travel a little slower than the rest of the wave and eventually fall back into the tail shock. This nonlinear distortion is described by the theory of nonlinear acoustics. The nonlinear distortion results in the fine structure of the sonic boom being removed because the shocks coalesce as the boom propagates away from the aircraft (Fig. 2*b*). Eventually the sonic boom consists of a single head shock followed by a linear reduction in pressure to the single tail shock (Fig. 2*c*), a waveform referred to as an N-wave because of its shape. Near the aircraft the time interval between the head shock and the tail shock is $\Delta t = L_{\text{plane}}/c_0$, where L_{plane} is the length of the plane. The presence of nonlinear distortion can result in Δt increasing by as much as 25% by the time that the boom reaches the ground. For longer aircraft, flying at high altitudes (for example, the Concorde), Δt is greater than 0.3 s and the human ear can distinguish the head and tail shocks; that is, the sonic boom sounds like two very loud cracks, or like rifle shots in quick succession. For shorter aircraft (for example, the F-18) flying close to the ground, Δt is less than 0.1 s and the sonic boom sounds like a single bang. *See* NONLINEAR ACOUSTICS.

Atmospheric propagation. It can take up to 60 s for a sonic boom to travel from the plane to the ground (**Fig. 3**). As the sonic boom propagates away from the aircraft, it undergoes geometrical spreading and decreases in amplitude. If the sonic boom follows straight ray paths, the conical nature of the wavefront means that the spreading is cylindrical and the amplitude will decreases as $1/\sqrt{r}$, where r is the distance from the aircraft. In contrast, sound from a subsonic aircraft spreads spherically and the amplitude decreases as $1/r$, a much more rapid reduction. However, for sonic booms, nonlinear distortion adds extra losses, due to the presence of the shocks, and the amplitude reduction becomes $r^{-3/4}$, about half way between cylindrical and spherical spreading.

Refraction. The propagation of real sonic booms is more complex than a cylindrical spreading wave because the acoustic rays do not follow straight lines in the atmosphere. The rays undergo refraction due to the change in sound speed with altitude. Refraction in acoustics is directly analogous to that in optics, and

acoustic waves also obey Snell's law. Refraction acts to bend rays away from regions of high sound speed and toward regions of low sound speed. Therefore, as the sonic boom travels toward the ground, where the sound speed is highest, the rays are bent away. Refraction in the atmosphere is strong enough that some rays never reach the ground; they are refracted back up into the atmosphere (a form of total internal reflection). The sonic boom is therefore heard only in a narrow strip below the aircraft, referred to as the sonic boom carpet. The width of the sonic boom carpet is about 1 km for each 200 m of altitude (1 mi for each 1000 ft). For example, a supersonic aircraft flying at 16,000 m (50,000 ft) will be heard only along a sonic boom carpet about 80 km (50 mi) wide. The loudness of the sonic boom varies across the carpet; it is loudest directly beneath the aircraft and decreases with lateral distance from the flight path. *See* REFRACTION OF WAVES.

Effects of wind and turbulence. Wind also affects the propagation of the sonic boom. In the upper atmosphere the winds are typically steady and at fixed altitude; the jet stream is the strongest high-altitude wind. Wind changes the local sound speed and therefore can both affect the width of the sonic boom carpet and move this strip in the lateral direction. Close to the ground, typically within 1 km (0.6 mi), a turbulent region exists referred to as the atmospheric boundary layer. In this region, the wind and the ground interact to make the air move in a chaotic eddying motion. These turbulent eddies can strongly perturb the sonic boom waveforms. When the turbulence is strong, the N-wave structure can be broken up to produce very ragged looking waveforms. On occasion the turbulence can even focus sonic booms, resulting in intensified sound. *See* ATMOSPHERIC ACOUSTICS; BOUNDARY LAYER FLOW; TURBULENT FLOW.

Super booms. The nature of the sonic boom can be affected by aircraft maneuvers. When a supersonic aircraft accelerates, climbs, or turns, the sonic boom no longer forms a conical wavefront, but generates a curved wavefront. Such a wavefront can lead to focusing, resulting in what are termed super booms, with amplitudes up to twice that of booms generated by steady flight. These super booms can occur outside the sonic boom carpet; for example, when the Concorde decelerates over the Atlantic as it approaches Paris, a focused sonic boom can be heard in Sweden.

Impacts on the ground. The typical peak pressure amplitude (or overpressure) of a sonic boom on the ground is about 50–100 pascals (Fig. 2*c*). A sonic boom with 50 Pa (1 lbf/ft^2 or 0.007 psi) overpressure will produce no damage to buildings. Booms in the range of 75–100 Pa are considered disturbing by some people. Occasionally there is minor damage to buildings from sonic booms in the range of 100–250 Pa; however, buildings in good condition will be undamaged by overpressures up to 550 Pa. Very low flying aircraft (30 m or 100 ft) can produce sonic booms of 1000–7000 Pa. These pressures are still about five times less than that needed to injure the human ear, but can lead to damage to buildings, such as the breaking of glass windows and the cracking of plaster.

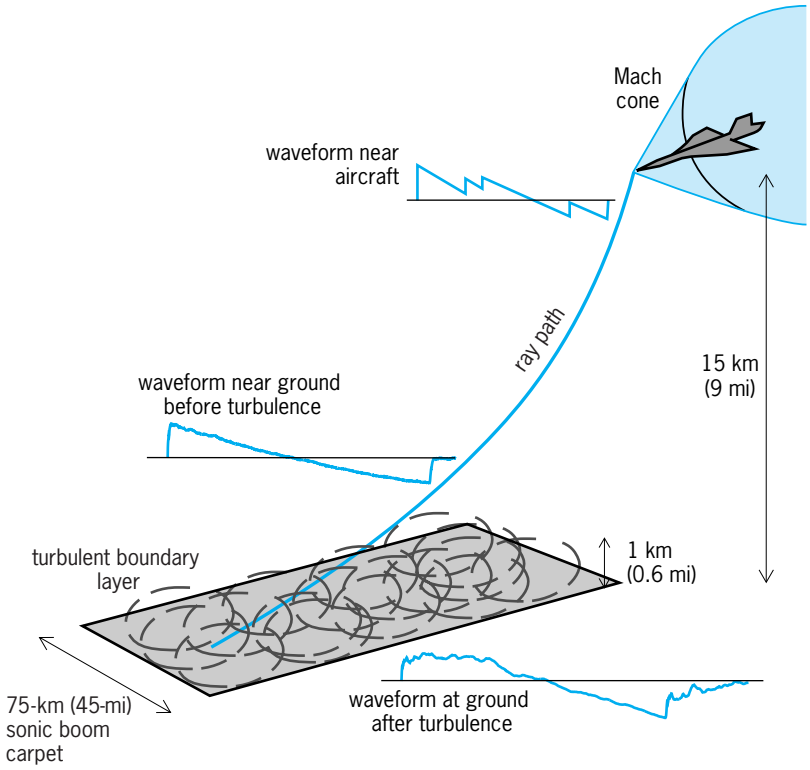


Fig. 3. Propagation of a sonic boom through the atmosphere from an aircraft flying at 15,000 m (50,000 ft).

Although sonic booms are not dangerous, they can evoke a strong startle response in people and animals. A consequence of the nonlinear distortion is that the shocks at the front and back of the sonic boom have a high-frequency content and sound like a gunshot. The high-frequency components are considered very annoying by people who are outdoors. People within houses are normally disturbed by the rattling effect of the sonic boom on the building. Experiments have shown that regular impacts by sonic booms generated by current aircraft are not acceptable to the general population, and so far it has not been possible to design aircraft that can fly supersonically and produce a boom that is acceptable to people underneath the flight path. For this reason, civilian aircraft are not permitted to generate sonic booms overland; therefore, the Concorde's flight path over the Atlantic is chosen to be longer than that for regular subsonic flights, to ensure that the plane is over water for the longest time and therefore able to remain in supersonic flight. Most military aircraft carry out supersonic flights over open water at altitudes greater than 3000 m (10,000 ft) and at least 23 km (15 mi) from shore. Supersonic operations over land are conducted above 10,000 m (30,000 ft) except for some specially designated areas, for example, near Fallon, Nevada, where low-level supersonic flights are permitted. Robin Cleveland

Bibliography. C. M. Darden et al., *Status of Sonic Boom Methodology and Understanding*, NASA CP

3027, 1989; A. D. Pierce and D. J. Maglieri, Effects of atmospheric irregularities on sonic-boom propagation, *J. Acous. Soc. Amer.*, 51:702–721, 1972; K. J. Plotkin, Review of sonic boom theory, *AIAA Pap.*, no 89-1105, April 1989; H. E. von Gierke and C. W. Nixon, Human response to sonic boom in the laboratory and the community, *J. Acous. Soc. Amer.*, 51:766–782, 1972.

Sonobuoy

An expendable device that enables aircraft to detect underwater objects, such as submarines, acoustically. Acoustics is the preferred energy form for use in salt water, because it tends to be the least attenuated by the medium.

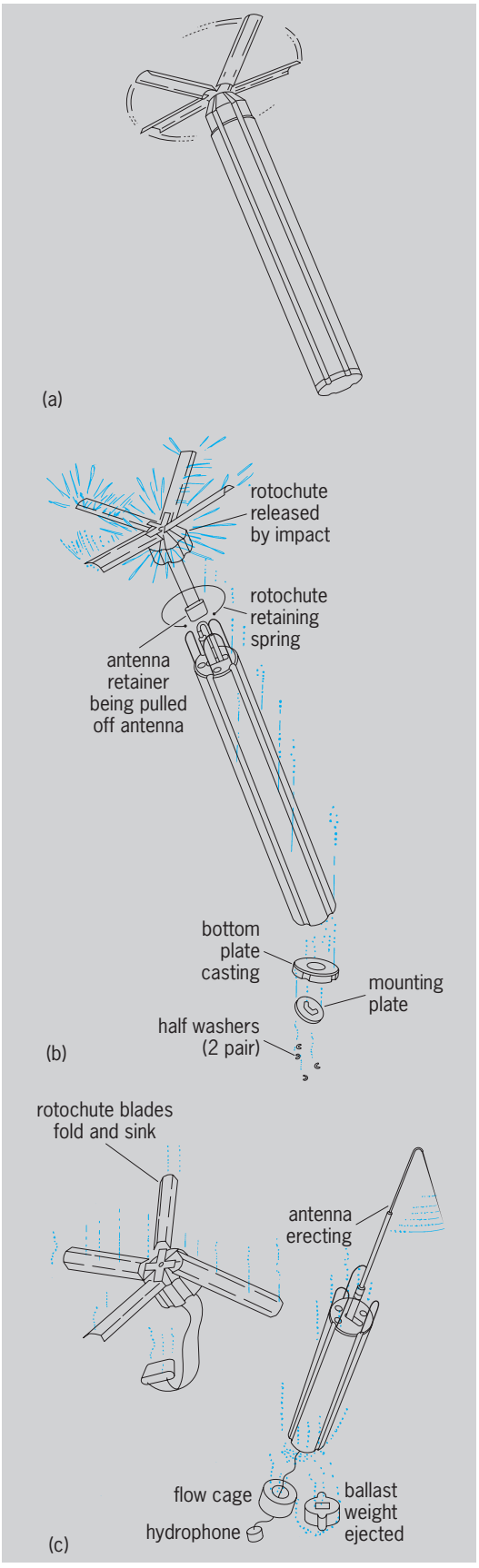
A sonobuoy consists of an electronic radio link and antenna connected to a miniature sonar system. It contains the means for its launch from the aircraft, its entry into the water, separation of a floating antenna from the underwater transducer and sonar, and activation of a seawater battery upon entry, as well as a scuttling means for final sinking upon completion of its intended useful life (see **illus.**). The short life requires that each component be highly reliable and effective at low cost. The package must also be small and lightweight, since large numbers of packages are to be carried on an aircraft. *See* SONAR.

The aircraft is an extremely mobile platform that can rapidly follow up on an initial detection and classification of a submarine. Successive sonobuoy deployments permit tracking the submarine's course and taking or directing offensive action against a submarine target. *See* ANTISUBMARINE WARFARE; UNDERWATER SOUND.

The simplest sonobuoy is a passive sonar that senses sound with its hydrophone, amplifies and converts it to a radio signal, and transmits the signal from its antenna for analysis, evaluation, and storage in the aircraft. The sonobuoy system may be independently activated by an underwater sound source, usually an explosive device dropped from the aircraft. Two or more simple sonobuoys may be deployed to permit processing of directional information, passively or actively.

A sonobuoy that can transmit sound into the water and perform echo ranging on a submarine usually requires one of the larger standard-size configurations and larger launching tubes on the aircraft. Also, a command feature aboard the aircraft and a radio receiver in the sonobuoy are required. Directional transmission may be achieved using elementary superdirective beams such as dipoles and cardioids. Multiple sonobuoys and superdirective beamforming may be combined for sound reception. *See* DIRECTIVITY.

An Englishman, P. M. S. Blackett, proposed the concept of the sonobuoy in 1941. The original development was transferred to the National Defense and Research Committee of the United States, because British research gave higher priority to other tasks. By the end of World War II, substantial numbers of



Launch of a sonobuoy from an antisubmarine warfare (ASW) aircraft. (*a*) In flight. (*b*) Buoy strikes the water in a nearly vertical attitude. (*c*) Antenna and hydrophone are positioned automatically.

sonobuoys had been designed, manufactured, and used in antisubmarine warfare (ASW).

Buoys of various types are used as sensors for oceanographic data such as sound speed, geophysical data such as earthquakes, bioacoustic data such as snapping shrimp, and other signal and noise sources. Some have been designed for long-term deployment on station, with provision for storing data until interrogated from an aircraft or a ship. Sonobuoys providing some of these functions have also been designed and deployed, and new applications can be anticipated. *See* BUOY.

Stanley L. Ehrlich

Bibliography. T. A. Stefanick, *Strategic Antisubmarine Warfare and Naval Strategy*, 1987; R. J. Urick, *Principles of Underwater Sound*, 3d ed., 1983, reprint 1996.

Sonochemistry

The study of the chemical changes that occur in the presence of sound or ultrasound. Industrial applications of ultrasound include many physical and chemical effects, for example, cleaning, soldering, welding, dispersion, emulsification, disinfection, pasteurization, extraction, flotation of minerals, degassing of liquids, defoaming, and production of gas-liquid sols. *See* SOUND; ULTRASONICS.

Surprisingly, when liquids are exposed to intense ultrasound, high-energy chemical reactions occur, often accompanied by the emission of light. There are three classes of such reactions: homogeneous sonochemistry of liquids, heterogeneous sonochemistry of liquid-liquid or liquid-solid systems, and sonocatalysis (which overlaps the first two). In some cases, ultrasonic irradiation can increase reactivity by nearly a millionfold. Especially for liquid-solid reactions, the rate enhancements via ultrasound have proved extremely useful for the synthesis of organic and organometallic compounds. Because cavitation occurs only in liquids, chemical reactions are not generally seen in the ultrasonic irradiation of solids or solid-gas systems.

Mechanisms. Ultrasound spans the frequencies of roughly 20 kHz to 10 MHz (human hearing has an upper limit of less than 18 kHz). Since the velocity of sound in liquids is approximately 5000 ft/s (1500 m/s), ultrasound has acoustic wavelengths of roughly 7.5–0.015 cm. Because the wavelength of ultrasound are so much larger than molecular dimensions, sound does not cause such chemistry in a direct interaction. Instead, the chemical effects of ultrasound derive from several different physical mechanisms, depending on the nature of the system. For both sonochemistry and sonoluminescence, the most important of these mechanisms is acoustic cavitation: the formation, growth, and implosive collapse of bubbles in liquids irradiated with high-intensity sound (**Fig. 1**). During the final stages of cavitation, compression of the gas inside the bubbles produces enormous local heating and high pressures (but with very short lifetimes, in otherwise cold liquids). In clouds of cavitating bubbles, these hot spots have

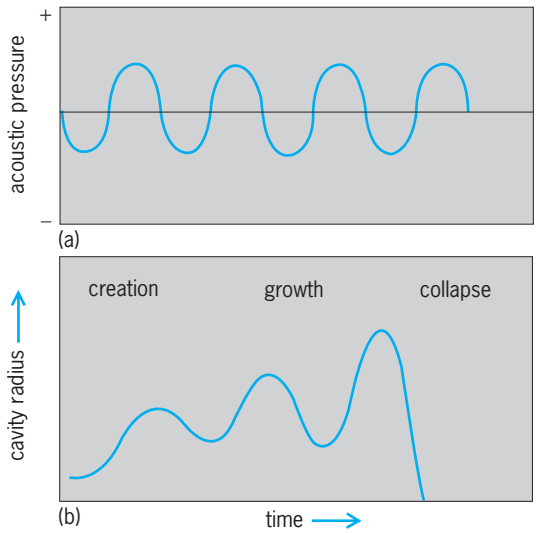


Fig. 1. Idealized representation of (*a*) expansion and compression waves and (*b*) the formation, growth, and collapse of a bubble in the process of acoustic cavitation. (*After K. S. Suslick, in F. G. A. Stone and R. West, eds., Advances in Organometallic Chemistry, vol. 25, 1986*)

equivalent temperatures nearly as high as the surface of the Sun (roughly 5000 K or 8540°F), pressures as large as at the bottom of the ocean (about 1000 atmospheres or 10^8 pascals), and lifetimes much less than a millionth of second. The dynamics of cavity growth and collapse are strikingly dependent on local environment, and cavitation in a homogeneous liquid should be considered separately from cavitation near an interface. The symbol $\xrightarrow{)))}$ will be used in this article to indicate ultrasonic irradiation (sonication or insonation). *See* CAVITATION.

When a liquid-solid interface is subjected to ultrasound, cavitation still occurs, but with major changes in the nature of the bubble collapse. No longer do cavities implode spherically. Instead, a markedly asymmetric collapse occurs, which generates a jet of liquid directed at the surface (**Fig. 2**). The jet velocities are greater than 330 ft/s (100 m/s). The origin of this jet formation is essentially a shaped-charge effect. The impingement of this jet can create a localized erosion (and even melting),

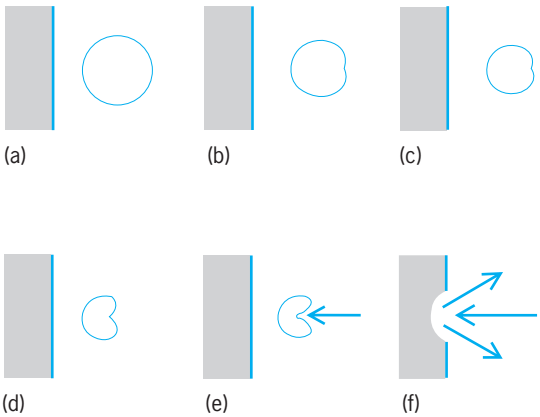


Fig. 2. Sequence *a–f* of a single bubble collapsing during cavitation near a surface.

responsible for surface pitting and ultrasonic cleaning. Enhanced chemical reactivity of solid surfaces is associated with these processes. The cavitational erosion generates unpassivated, highly reactive surfaces; causes short-lived high temperatures and pressures at the surface; produces surface defects and deformations; forms fines and increases the surface area of friable solid supports; and ejects material in unknown form into solution. The local turbulent flow associated with acoustic streaming improves mass transport between the liquid phase and the surface, thus increasing observed reaction rates.

Ultrasonic irradiation of liquid-powder suspensions produces another effect: high-velocity interparticle collisions. Cavitation and the shock waves that it creates in a slurry can accelerate solid particles to high velocities. The resultant collisions are capable of inducing dramatic changes in surface morphology, composition, and reactivity.

Homogeneous systems. A variety of novel reactivity patterns are distinct from either normal thermal or photochemical activation. Most of the reported reactions are stoichiometric in terms of a consumed reagent, but a few examples of true sonocatalysis have also appeared. The predominant reactions of homogeneous sonochemistry are bond breaking and radical formation. The effect of ultrasound on aqueous solutions has been studied for many years. The primary products are H_2 and H_2O_2; there is strong evidence for various high-energy intermediates, including HO_2, $H^\bullet$, $OH^\bullet$, and perhaps $e^-_{(aq)}$. Most recently there has been strong interest shown in the use of ultrasound to remediate low levels of organic contamination of water. The $OH^\bullet$ radicals produced from the sonolysis of water are able to attack essentially all organic compounds (including halocarbons, pesticides, and nitroaromatics) and, through a series of reactions, oxidize them fully. The desirability of sonolysis for such remediation lies in its low maintenance requirements and the low energy efficiency of alternative methods (for example, ozonolysis, ultraviolet photolysis). *See* HOMOGENEOUS CATALYSIS.

Sonochemistry also occurs in nonaqueous solutions. For example, the sonolysis of volalite organometallic complexes, whose thermal and photochemical reactivities were well characterized, gave quite unusual reactivities, including controlled multiple-ligand substitution and clusterification; examples are shown in reactions (1) and (2).

$$Fe(CO)_5 \xrightarrow{)))} Fe_3(CO)_{12} \quad (1)$$

$$Fe(CO)_5 + L \xrightarrow{)))} Fe(CO)_{5-n}L_n \quad (n = 1, 2, 3;\ L = \text{Lewis base}) \quad (2)$$

See ORGANOMETALLIC COMPOUND.

Sonoluminescence. In addition to the initiation or enhancement of chemical reactions, irradiation of liquids with high-intensity ultrasound generates the emission of visible light. The production of such luminescence is a consequence of the localized hot spot created by the implosive collapse of gas- and vapor-filled bubbles during acoustic cavitation. In general, sonoluminescence may be considered a special case of homogeneous sonochemistry, and recent discoveries in this field have heightened interest in the phenomenon. In clouds of collapsing bubbles, emission from the excited states of atoms and molecules (very much as in flames) can be observed. For example, high-resolution sonoluminescence spectra were obtained from cavitation clouds in silicone oil under argon. The observed emission comes from excited-state diatomic carbon (C_2), and was modeled as a function of rotational and vibrational temperatures to find an effective cavitation temperature of 5050 ± 150 K (8630 ± 270°F). Recent work on metal atom sonoluminescence from volatile organometallics, such as iron pentacarbonyl [$Fe(CO)_5$], accurately confirms this temperature. Under conditions where an isolated, single bubble undergoes cavitation, recent studies on the duration of the sonoluminescence flash suggest that a shock wave may be created within the collapsing bubble, with the capacity to generate truly enormous temperatures and pressures within the gas—much higher than those seen with interacting clouds of bubbles. *See* CHEMILUMINESCENCE.

In a sense, chemistry is just the interaction of energy and matter. All chemical reactions require energy in one form or another to proceed. To a large degree, the properties of a specific energy source determine the course of a chemical reaction. Ultrasonic irradiation differs from traditional energy sources (such as heat, light, or ionizing radiation) in duration, pressure, and energy per molecule. The immense local temperatures and pressures and the extraordinary heating and cooling rates generated by cavitation bubble collapse mean that ultrasound provides an unusual mechanism for generating high-energy chemistry.

Heterogeneous systems. A major industrial application of ultrasound is emulsification. The first reported and most studied liquid-liquid heterogeneous systems have involved ultrasonically dispersed mercury. The use of such emulsions for chemical purposes has been delineated by extensive investigations. The effect of the ultrasound in this system appears to be due to the large surface area of mercury generated in the emulsion. *See* EMULSION.

The effects of ultrasound on liquid-solid heterogeneous organometallic reactions have been a matter of intense investigation. Various research groups have dealt with extremely reactive metals, such as lithium (Li), magnesium (Mg), or zinc (Zn), as stoichiometric reagents for a variety of common transformations. Examples are shown in reactions (3)–(6), where R represents an organic functional group.

$$RBr + R'R''CO \xrightarrow[Li]{)))} RR'R''COH \quad (3)$$

$$RBr + (CH_3)_2NCHO \xrightarrow[Li]{)))} RCHO \quad (4)$$

$$R_2SiCl_2 \xrightarrow[Li]{)))} \text{cyclo-}(R_2Si)_3 \quad (5)$$

$$RR'CO + BrCH_2CO_2R'' \xrightarrow[Li]{)))} RR'C(OH)CH_2CO_2R'' \quad (6)$$

Ultrasonic cleaning of the reactive metal surface to remove passivating impurities (for example, water, hydroxide, metal halide, or organolithium) is likely to be important in the origin of these effects.

The activation of less reactive metals continues to attract major efforts in heterogeneous catalysis, metal-vapor chemistry, and synthetic organometallic efforts. Given the extreme conditions generated by acoustic cavitation at surfaces, analogies to autoclave conditions or to metal-vapor reactors may be appropriate. It has been found that the use of ultrasonic irradiation facilitates the reduction of a variety of transition-metal salts to an extremely active form.

The effects of ultrasound on heterogeneous systems are quite general, and ultrasonic rate enhancements for many nonmetallic insoluble reagents also occur. Many organic reactions have been studied by various groups, including reductions with lithium aluminum hydride ($LiAlH_4$) or sodium hydride (NaH), oxidations with potassium permanganate ($KMnO_4$), and thioamidation with phosphorus pentasulfide (P_4S_{10}). Ultrasound has been used to enhance the rates of mass transport near electrode surfaces and thus to enhance rates of electrolysis. This has had some useful synthetic applications for the production of both organic and inorganic chalcogenides.

Applications of ultrasound to electrochemistry have also seen substantial recent progress. Beneficial effects of ultrasound on electroplating and on organic synthetic applications of organic electrochemistry have been known for quite some time. More recent studies have focused on the underlying physical theory of enhanced mass transport near electrode surfaces. Another important application for sonoelectrochemistry has been developed for the electroreductive synthesis of submicrometer powders of transition metals.

Applications to materials chemistry. Of recent interest is the development of sonochemistry as a synthetic tool for the creation of unusual inorganic materials (**Fig. 3**). As one example, the discovery of a simple sonochemical synthesis of amorphous iron helped settle the long-standing controversy over its magnetic properties. More generally, ultrasound has proved extremely useful in the synthesis of a wide range of nanostructured materials, including high-surface-area transition metals, alloys, carbides, oxides, and colloids. Sonochemical decomposition of volatile organometallic precursors in high-boiling solvents produces nanostructured materials (that is, built from nanometer-sized particles) in various forms. Nanometer colloids, nanoporous high-surface-area aggregates, and nanostructured oxide-supported catalysts can be prepared by this general route. The sonochemically prepared materials generally have very large surface areas, with surfaces that are extremely active as heterogeneous catalysts for a variety of important industrial reactions, including hydrogenations and dehydrogenations, hydrodesulfurization, and carbon monoxide reductions. Heterogeneous catalysts often require rare and expensive metals. The use of ultrasound offers some hope of activating less reactive, but also less costly, metals.

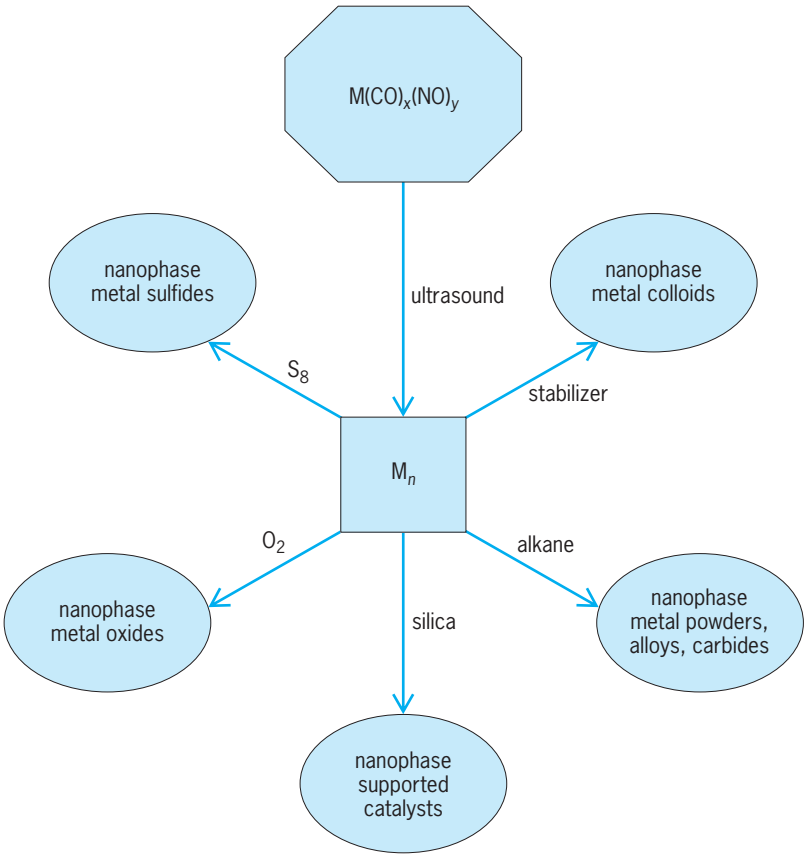


Fig. 3. Sonochemical synthesis of various forms of nanostructured materials. Starting with volatile precursors, such as $M(CO)_x(NO)_y$ where M can be any of several metals, ultrasound can be used to break the precursor apart and liberate metal atoms that form small metal clusters of 100–1000 atoms. These clusters are then used to form various nanostructured materials.

Such effects can occur during the formation of supported catalysts, by activation of preformed catalysts, or through enhancement of catalytic behavior during catalysis. *See* NANOCHEMISTRY; NANOSTRUCTURE.

Polymers. Sonochemistry is also proving to have important applications with polymeric materials. Substantial work has been accomplished in the sonochemical initiation of polymerization and in the

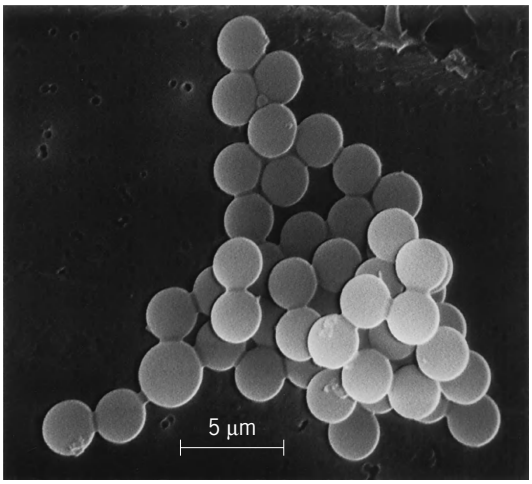


Fig. 4. Scanning electron micrograph of sonochemically synthesized hemoglobin microspheres. (*Courtesy of M. Wong, K. J. Kolbeck, and K. S. Suslick*)

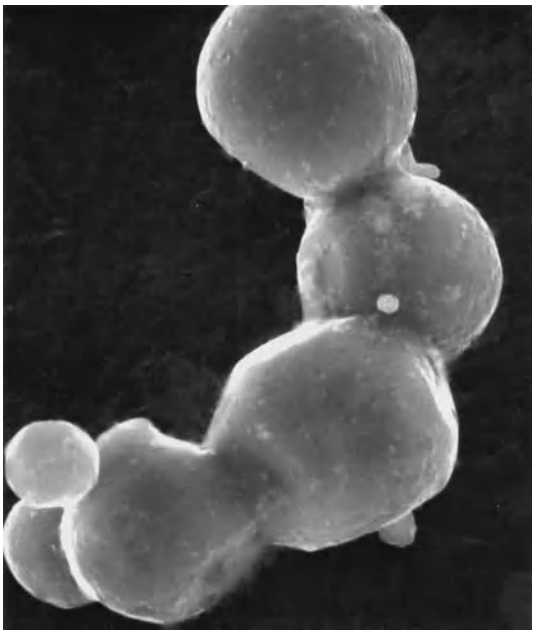


Fig. 5. Scanning electron micrograph showing particle agglomeration after ultrasonic irradiation of slurries of zinc powder (average diameter 5 μm). (*Courtesy of S. J. Doktycz and K. S. Suslick*)

modification of polymers after synthesis. The use of sonolysis to create radicals that function as radical initiators has been well explored. Similarly the use of sonochemically prepared radicals and other reactive species to modify the surface properties of polymers is actively being developed. Other effects of ultrasound on long-chain polymers tend to be mechanical cleavage, which produces relatively uniform size distributions of shorter chain lengths. *See* POLYMER.

Protein microspheres. Sonochemistry has found another recent application in the preparation of unusual biomaterials, notably protein microspheres (**Fig. 4**). Using high-intensity ultrasound and simple protein solutions, a remarkably easy method to make both air-filled microbubbles and nonaqueous liquid-filled microcapsules has been developed. Sonochemically prepared microspheres are stable for months, and being slightly smaller than red blood cells (erythrocytes), they can be intravenously injected and pass unimpeded through the circulatory system. The mechanism responsible for microsphere formation is a combination of two acoustic phenomena: emulsification and cavitation. Ultrasonic emulsification creates the microscopic dispersion of the protein solution necessary to form the proteinaceous microspheres. However, by itself, emulsification is insufficient to produce long-lived microspheres. The long life of these microspheres comes from a sonochemical cross-linking of the protein shell. Protein cysteine residues are oxidized during microsphere formation by sonochemically produced superoxide to create disulfide links between proteins (these same interprotein cross-links are responsible for the mechanical strength of hair, horn, and nails). These protein microspheres have a wide range of biomedical applications, including their use as echo contrast agents for sonography, magnetic resonance imaging contrast enhancement, and drug delivery. *See* PROTEIN.

Applications to heterogeneous catalysis. Ultrasonic irradiation can alter the reactivity observed during the heterogeneous catalysis of a variety of reactions. Sonication has shown such behavior by altering the formation of heterogeneous catalysts, by perturbing the properties of previously formed catalysts, or by affecting the reactivity during catalysis. However, ultrasonic rate enhancements of heterogeneous catalysis have usually been relatively modest (less than tenfold). The effect derived from irradiating operating catalysts is often due simply to improved mass transport. In addition, increased dispersion during the formation of catalysts under ultrasound will enhance reactivity, as will the fracture of friable solids (for example, noble metals on carbon or silica). It has been discovered that hydrogenation of alkenes by nickel powder is dramatically enhanced (by nearly a millionfold) by ultrasonic irradiation. Ultrasonic irradiation of liquid-solid slurries radically alters the surface composition of metal powders. The passivating oxide layer, which normally coats nickel and most metals, is rapidly removed by ultrasound. *See* CATALYSIS; HETEROGENEOUS CATALYSIS.

Scanning electron micrographs of metal powders after ultrasonic irradiation of slurries reveal the importance of interparticle collisions (**Fig. 5**). Such collisions are driven by the turbulent flow and shock waves that are created by high-intensity ultrasound. It appears likely that the observed changes in surface morphology, in the degree of aggregation, and in the surface's elemental composition originate from such interparticle interactions. The chemical effects derive in large part from the surface depassivation that results.

Kenneth S. Suslick

Bibliography. I. Hua, R. H. Hochemer, and M. R. Hoffmann, *Environ. Sci. Technol.*, 29:2790, 1995; T. G. Leighton, *The Acoustic Bubble*, Academic Press, London, 1994; K. S. Suslick, *MRS Bull.*, no. 20, 29, 1995; K. S. Suslick, Sonocatalysis, in G. Ertl et al. (eds.), *Handbook of Heterogeneous Catalysis*, vol. 3, Wiley-VCH, Weinheim, 1997; K. S. Suslick and L. A. Crum, Sonochemistry and sonoluminescence, in M. J. Crocker (ed.), *Handbook of Acoustics*, Wiley-Interscience, New York, 1998.

Sonoluminescence

The transformation of acoustic energy into light through collapsing bubbles. In particular, single-bubble sonoluminescence (SBSL) is the periodic light emission of a single acoustically trapped and periodically driven gas bubble. It occurs when the bubble collapses so violently that the energy focusing during collapse is intense enough for partial ionization of the gas inside the bubble to occur, leading to subsequent light emission. The phenomenon of single-bubble sonoluminescence was discovered in 1989 by Felipe Gaitan, then a graduate student, working with Lawrence A. Crum. A

micrometer-sized bubble is acoustically trapped in a water-filled flask at resonance. The driving pressure is typically $P_a = 1.2$–1.4 bar (1 bar $= 10^5$ Pa $\cong$ 1 atm), the driving frequency is 20–40 kHz, and the air saturation in the water is typically 20–40%. Once per cycle, at bubble collapse, the bubble emits a short pulse of light, that typically lasts 100–300 picoseconds. The origin of the light is thermal bremsstrahlung: During the adiabatic collapse, the gas inside the bubble is heated, presumably to about 15,000 K. Consequently, the gas partly ionizes. The ions and electrons interact, decelerate, and finally recombine, and the deceleration is connected with light emission. *See* BREMSSTRAHLUNG.

Though the energy-focusing power at the bubble collapse is about 12 orders of magnitude and the light emission is rather spectacular, it is negligible from an energy-balance point of view. The majority of the incoming acoustic energy is emitted again as sound (during the violent bubble collapse, and therefore at much higher frequencies), is converted into heat, or is eaten up by chemical reactions. Therefore single-bubble sonoluminescence can be understood as "illuminated cavitating bubble dynamics."

Bubble dynamics. The backbone of the theoretical understanding of single-bubble sonoluminescence is the Rayleigh-Plesset equation, given below. It describes the dynamics of a spherical bubble, with radius $R(t)$, in a liquid. Here σ denotes the surface tension, η the dynamic viscosity, P_0 the ambient pressure (usually 1 atm), $p_g(R(t))$ the gas pressure inside the bubble, and $P(t)$ the time-dependent external driving pressure, which in the case of single-bubble sonoluminescence is simply $P(t) = P_a \sin(2\pi f t)$, with the driving amplitude P_a and the driving frequency f.

$$R\ddot{R} + \frac{3}{2}\dot{R}^2 = \frac{1}{\rho}\left(p_g - P_0 - P(t) - 4\eta\frac{\dot{R}}{R} - \frac{2\sigma}{R}\right)$$

The typical behavior of the bubble radius in the sonoluminescent regime is shown in **Fig. 1**, together with the driving pressure: On pressure reduction, the bubble strongly expands to typically 10 times its equilibrium radius R_0. Once the pressure increases, the bubble violently collapses, followed by a "ringing" of the bubble at its eigenfrequency. As seen from Fig. 1, the Rayleigh-Plesset dynamics nicely describes the experimental data for the bubble radius.

The energy-focusing property of the Rayleigh-Plesset dynamics can be seen by neglecting all terms on the right side of Eq. (1), that is, by considering only the inertial terms, to obtain the simpler equation $R\ddot{R} + \frac{3}{2}\dot{R}^2 = 0$. This equation can be integrated to obtain the solution $R(t) = R_0[(t_* - t)/t_*]^{2/5}$, with the remarkable feature of a divergent bubble-wall velocity as t approaches the time t_* of the void collapse. It is this finite time singularity which leads to the adiabatic heating of the gas inside the bubble. The divergence is eventually cut off by the pressure building up inside the bubble and by the sound emission at bubble collapse.

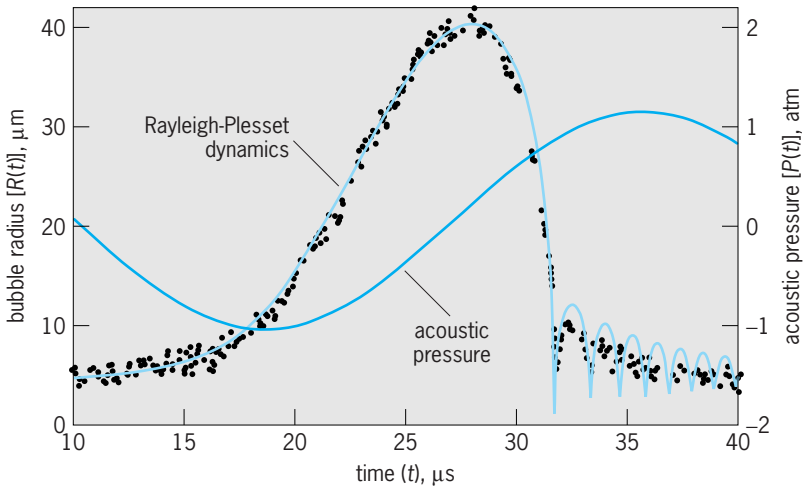


Fig. 1. Radius of a sonoluminescing bubble [$R(t)$, measured on the left axis] and acoustic pressure [$P(t)$, right axis] as functions of time. Rayleigh-Plesset dynamics (solid line) well describes the experimental data (circles) for the bubble radius.

The bubble dynamics also directly or indirectly determines the conditions under which stable single-bubble sonoluminescence can occur. These conditions are (1) that the threshold in the driving pressure for the Rayleigh collapse to occur be exceeded; (2) the shape stability of the bubbles (that is, the bubble must remain spherical and pinch-offs of microbubbles must not occur); (3) its diffusive stability (that is, the average number of gas molecules inside the bubble must not shrink or grow); and (4) the chemical stability of its ingredients. By applying these criteria, the phase diagrams of sonoluminescence can be calculated quantitatively (**Fig. 2**).

Chemical activity. The chemical activity inside the bubble is a consequence of the high temperatures

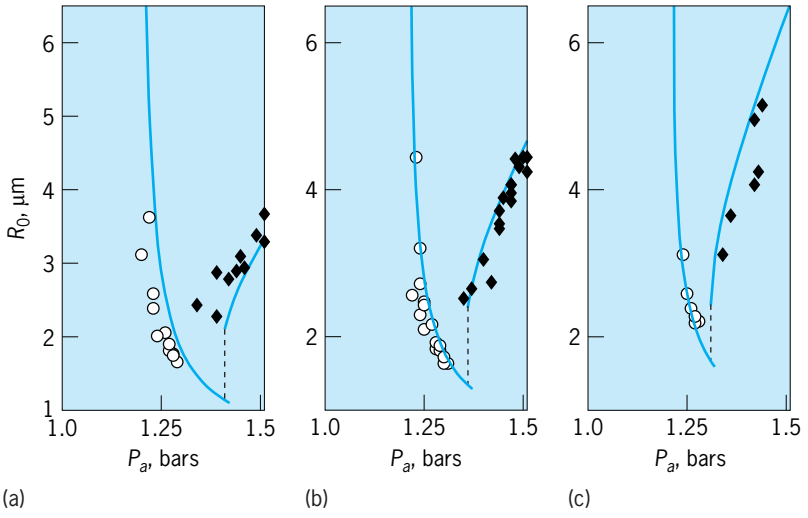


Fig. 2. Comparison of parameter-free model results (solid lines) and experimental data for air bubbles at (*a*) 10%, (*b*) 20%, and (*c*) 40% saturation in the ambient radius (R_0) versus forcing pressure (P_a) parameter space. Closed symbols refer to stable sonoluminescing bubbles, open symbols to stable bubbles that do not emit light. The ambient radius (that is, the radius of the bubble under normal pressure and temperature) is self-adjusting through diffusive processes like growth or shrinkage; there is no free parameter that the experimenter could choose to help the model results and the data agree. The driving frequency and the water temperature are $f = 33.4$ kHz, $T_0 = 293.15$ K $= 20^\circ$C $= 68^\circ$F.

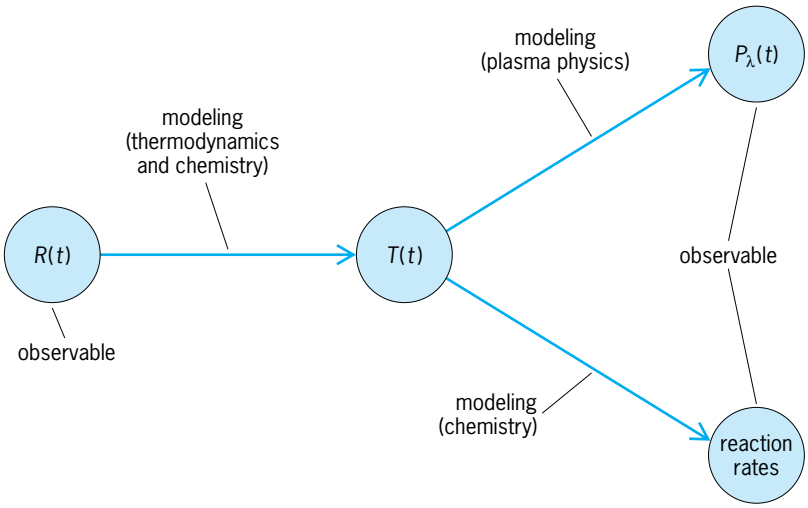


Fig. 3. Modeling steps that are involved in connecting the temperature $T(t)$ in the sonoluminescing bubble to observables. $T(t)$ is obtained from the radius dynamics $R(t)$, and in turn the spectral radiance $P_\lambda(t)$ and chemical reaction rates are deduced from the temperature. In 2005, the temperature of single-bubble sonoluminescence was directly measured and found to be 15,000 K.

achieved. The sonoluminescing bubble can be viewed as a high-temperature, high-pressure microlaboratory or reaction chamber, which can be controlled through the external parameters such as forcing pressure, frequency, or dissolved gas concentration. When the bubble is expanding, gas dissolved in the liquid and liquid vapor enter the bubble. During the adiabatic collapse, these gases are partly trapped inside the hot bubble and react. For example, nitrogen molecules will first dissociate to form nitrogen radicals and later react to form NH, NO, and so forth, which all dissolve readily in water when the bubble cools down and reexpands. Subsequent, reaction cycles, burn off further molecular gases. Eventually, all that remains in the bubble are the inert gases. In the case of air, this means argon. In this sense, stable sonoluminescing bubbles rectify argon.

Bubble temperature. The ultimate question in cavitation physics has always been how hot does the bubble get. That is, what limits the temperature increase, or in other words, what regularizes the singularity? The chief limiting processes are the sound emission from the collapsing bubble and the dissociation and eventually ionization of the gas inside the bubble, all of which consume energy. *See* SONOCHEMISTRY.

Direct measurement of the temperature in the bubble is difficult; measurements of most observable quantities enable only indirect deductions of the temperature, and modeling assumptions are necessary for its calculation. The situation is sketched in **Fig. 3**.

For the case of transient caviation, in which bubbles are destroyed at collapse and reform elsewhere (multibubble sonoluminescence or MBSL), Kenneth S. Suslick and coworkers succeeded in extracting the value of the temperature from observations of spectral lines in the early 1990s. Their result was 5000 K. However, MBSL light emission is rather weak compared to SBSL light emission, and so there was hope that much higher temperatures could be achieved in SBSL, nourished by the absence of any observed lines in SBSL spectra.

However, in 2005 Suslick and coworkers finally reported observing spectral lines in SBSL, enabling a direct temperature determination in that case as well. To achieve particularly strong SBSL, they used xenon and argon bubbles in sulfuric acid. This system combines three advantages: (1) high fluid viscosity, to allow the formation of large shape-stable bubbles; (2) low vapor pressure; and (3) monoatomic gases, so that the focused energy is not eaten up by the internal degrees of freedom of the gas–vapor mixture in the gas bubble. In this way, 1000 times more photons are emitted than with xenon or argon bubbles in water, allowing for good photon statistics and detailed spectral observations. From the atomic emission lines of argon, a temperature of 15,000 K could be deduced. Suslick and coworkers also observed strongly excited emissive states that are inconsistent with any thermal process, that is, light emission must originate from collisions of high-energy ions or electrons. This is the first direct experimental evidence that the light emission process of SBSL is thermal bremsstrahlung by an optically opaque plasma. Both the light-emitting process and the typical SBSL temperatures had been theoretically predicted, giving scientists confidence in the standard model of single-bubble sonoluminescence.

The reason for the much higher temperature in stable SBSL bubbles as compared to transient MBSL bubbles is that argon accumulates over hundreds of cycles in the SBSL case (which evidently is not possible if the bubble dies at collapse). Argon has an adiabatic exponent [equal to $(N + 2)/N$, where N is the number of degrees of freedom of the respective gas molecule] of 5/3, in contrast to nitrogen where it is only 7/5; water vapor has an even smaller adiabatic exponent. So, if the bubble is mainly filled with nitrogen, a large fraction of the energy from the collapse goes into rotational, vibrational, and dissociation energy, rather than leading to additional heating. Therefore, typically a temperature of only 5000 K can be achieved for transient cavitation. It is the bubble's stability in the SBSL case that allows for argon accumulation and therefore for higher temperatures of up to 15,000 K. Eventually, of course, further energy consuming processes take over, such as ionization (in addition to the rotation, vibration, and dissociation of the water vapor molecules, which are unavoidable). *See* ADIABATIC PROCESS; ISENTROPIC PROCESS.

Single-bubble sonoluminescence can be viewed as the hydrogen atom of cavitation physics: Single spherical bubble cavitation is the simplest building block of a sound-driven bubbly fluid, just as the hydrogen atom is the building block for more complicated atoms, molecules, or condensed matter. It is remarkable how many subdisciplines of physics and chemistry have been necessary to understand this conceptally simple building block: a single bubble oscillating in a sound field. These

subdisciplines range from acoustics to fluid dynamics, plasma physics, thermodynamics, atomic physics, spectroscopy, chemistry, dynamical system theory, and applied mathematics in general.

Detlef Lohse

Bibliography. B. P. Barber et al., Defining the unknowns of sonoluminescence, *Phys. Rep.*, 281:65, 1997; M. P. Brenner, S. Hilgenfeldt, and D. Lohse, Single bubble sonoluminescence, *Rev. Mod. Phys.*, 74:425, 2002; L. A. Crum, Sonoluminescence, *Phys. Today*, 47:22, 1994; D. J. Flannigan and K. S. Suslick, Plasma formation and temperature measurement during single-bubble cavitation, *Nature*, 434:52, 2005; S. Hilgenfeldt, S. Grossmann, and D. Lohse, A simple explanation of light emission in sonoluminescence, *Nature*, 398:402, 1999; W. C. Moss, D. B. Clarke, and D. A. Young, Calculated pulse widths and spectra of a single sonoluminescing bubble, *Science*, 276:1398, 1997; R. Toegel and D. Lohse, Phase diagrams for sonoluminescing bubbles: A comparison between experiment and theory, *J. Chem. Phys.*, 118:1863, 2003.