W

Water

F. H. Stillinger

Because of the large amount of water present on our planet, especially in liquid form, this substance has become centrally important for many aspects of science and technology. This importance is connected partly to the peculiar behavior of pure water, partly to its qualifications as a liquid solvent, and partly to its role as a fluid medium for support of life.

The most prominent peculiarities exhibited by pure water are the reduction in volume upon melting at 0 ℃ (by 8.3%), followed by further shrinkage to a density maximum as the liquid is heated to 4 ℃. These attributes are also shared by D_2O (m.p. 3.8 ℃, density max. at 11.2 ℃) and by T_2O (m.p. 4.5 ℃, density max. at 13.4 ℃). Although rare, these observations are not unique with water; the elements Si, Ge, and Bi also shrink upon melting, while In_2Te_3 appears to have a liquid-phase density maximum.

Additional water anomalies are (a) large number of ice polymorphs (including those that form at high pressure); (b) tendency toward reduced viscosity when liquid water below 30 ℃ is compressed; and (c) minimum in isothermal compressibility $[-(\partial \ln V/\partial p)_T]$ for the liquid at 46 ℃.

Observable properties of water in pure form and as a solvent stem from the structure of the individual water molecules and from the way that intermolecular forces between those molecules cause aggregation into liquid and solid.

The isolated H_2O molecule is shaped like a wide-open V, with the oxygen nucleus at the central bend and hydrogen nuclei forming "arms" of length 0.96 Å. The HOH angle is 104.5°. These dimensions can vary slightly as the molecule vibrates and interacts with neighbors in a crystal or the liquid, but the overall shape remains.

The dominant effect in water molecule interactions is the formation of hydrogen bonds. When two water molecules form a hydrogen bond, one (the hydrogen donor) points one of its OH groups toward the back side of the oxygen atom of the second (the hydrogen acceptor). This arrangement is illustrated in Fig. 1. The oxygen–oxygen lengths of these bonds normally lie in the range 2.7–3.0 Å, so that the donated hydrogen resides only about one third of the way between oxygens, and so still "belongs" to the donor.

Encyclopedia of Physics, Third Edition. Edited by George L. Trigg and Rita G. Lerner

ISBN: 3-527-40554-2

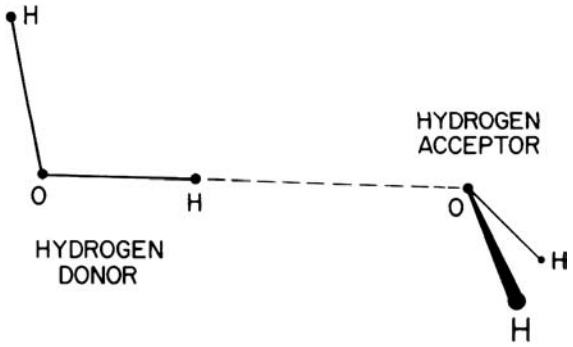

Fig. 1: Hydrogen bond (dashed line) linking two water molecules.

The maximum hydrogen bond strength is achieved when the molecules are arranged as shown in Fig. 1. This strength is about 4.2×10^{-13} ergs (6.0 kcal/mole of bonds), and exceeds thermal energy $k_B T$ by a factor of 10 at room temperature. The existence of these relatively strong hydrogen bonds between water molecules explains the relatively high melting and boiling temperatures for water, compared to other substances of comparable molecular weight (e. g., Ne, CH_4, NH_3, O_2, CO).

In a large aggregate of water molecules, optimum hydrogen bonding is achieved if each water molecule hydrogen-bonds to four others. Toward two of these four it donates its hydrogens, while it accepts hydrogens from the other two. This fourfold bonding is present in ordinary ice, causing formation of hydrogen bond hexagons. Without disturbing hydrogen bond strengths substantially, four-coordinated water networks can also form, which additionally incorporate squares, pentagons, and heptagons of hydrogen bonds. These patterns also exist in high-pressure ice polymorphs and in hydrate crystals.

Evidently the capacity for water molecules to form a diverse collection of three-dimensional networks of hydrogen bonds, while maintaining fourfold bonding at each molecule, has structural relevance for the liquid. Currently available evidence, both experimental and theoretical, indicates that liquid water consists of a structurally random network of hydrogen bonds uniformly filling the volume occupied by the liquid. That random network incorporates strained and broken hydrogen bonds, with greater frequency the higher the temperature. Furthermore, the network is labile, with bonds breaking in one place and reforming nearby, so that normal liquid flow and molecular diffusion are possible.

An isolated water molecule has dipole moment 1.86 debye (D), with hydrogens acting as though each bore one third of a protonic charge, and the oxygen as though it bore minus two thirds. Neighboring molecules in the liquid tend to have their dipoles somewhat aligned, so as to act in concert under the polarizing influence of an electrical field. The net result of this alignment, and of molecular polarizability, is a large static dielectric constant (88.0 at 0 ℃, but declining to 55.3 at 100 ℃).

The ease with which water dissolves many ionic crystals, such as the alkali halides, stems partly from its high static dielectric constant. However, it is also connected with the relatively small size of the water molecules, which permits them to approach ions closely, solvate them strongly, and thus overcome the largely electrostatic binding of the ionic crystals.

The solvating power of water for ions facilitates the dissociation of water molecules themselves into H^+ and OH^- ions. In the liquid at room temperature, roughly one molecule in 55 million will have dissociated. The H^+ and OH^- formed in this way can readily be incorporated into the liquid' s random hydrogen bond network, and they tend to form shortened hydrogen bonds in their vicinity. Both H^+ and OH^- have high apparent mobilities in water, due to the possibility of moving a succession of hydrogens along a chain of hydrogen bonds so as to cause a net transfer of ionic charge along that chain.

Nonionic substances with high solubility in water tend to have molecules with which water can hydrogen-bond. Usually this requires that molecules of those "hydrophilic" substances contain oxygen or nitrogen atoms.

Hydrocarbons (such as methane, hexane, acetylene, benzene) form an important group of "hydrophobic" molecular substances that are sparingly soluble in water. They cannot form hydrogen bonds with water strong enough to compete with those already present in that liquid itself. Consequently the random water network is obliged to restructure around the rare dissolved hydrocarbons so as to form a "cage" of hydrogen bonds of the required size. The corresponding geometric constraints on the water network cause entropies of solution for hydrocarbons in water to be negative.

Biologically important molecules (e. g. lipids, enzymes, RNA, hemes) often contain both hydrophilic and hydrophobic chemical groups. Consequently the biologically active conformations of these molecules, to the extent it is possible, place the hydrophilic groups on the outside to be in contact with water while the hydrophobic groups cluster within to avoid water contact. Since conformation is crucial to operation in most cases, it is obvious that specific solvation properties of water have profound effects in biology and doubtless have exerted a powerful influence on the course of chemical evolution from the first rudimentary "protolife" to present complex biochemistry.

See also: Hydrogen Bond; Ice.

Bibliography

D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*. Oxford, London and New York, 1969. (I, A)

F. Franks, ed., *Water, A Comprehensive Treatise*, Vols. I–V. Plenum, New York, 1972–1975. (I, A)

R. A. Home, ed., *Water and Aqueous Solutions*. Wiley (Interscience), New York, 1972. (I, A)

Waves

D. R. Tilley

Introduction

A simple wave is a single-frequency disturbance traveling at some speed v; the disturbance might be elastic strain, as in an acoustic wave; a combination of electric and magnetic fields, as in light; or some other quantity. All traveling waves transport energy. Single-frequency

waves are of basic importance, but it is sometimes necessary to consider more general disturbances. Because of the pervasiveness and importance of the concepts, general undergraduate texts [1, 2] contain major sections on waves, and a number of texts specifically on the topic have been written [3–6].

This review starts with waves that are both linear and nondispersive. That is, amplitudes can be added (linearity) and velocity is independent of frequency (nondispersive propagation). One section discusses the implications of linearity alone, another treats dispersive propagation in linear media, and the last section introduces nonlinearity.

Nondispersive Media

The displacement u for a wave moving in the x direction in a nondispersive medium satisfies the wave equation

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \; . \tag{1}$$

This applies to electromagnetic waves in vacuum [7, 8], in which case $v = c$ and u is an electric or magnetic field vector in the y–z plane (transverse wave). It also applies to long-wavelength acoustic waves, which may be longitudinal (u along x) or transverse. For acoustic waves, $v^2 = C/\rho$, where ρ is the density and C is an appropriate elastic modulus. The general solution of (1) is

$$u(x,t) = f(x-vt) + g(x+vt) \; , \tag{2}$$

where f and g are arbitrary (twice-differentiable) functions. The expression $f(x-vt)$ represents a disturbance $f(x)$ traveling to the right with speed v; we can see this by sketching $f(x-vt)$ at successive instants. The disturbance $g(x+vt)$ travels to the left with speed v. With single-frequency time dependence, the two terms take the form

$$u(x,t) = u_0 \exp(\pm ikx - i\omega t) \; . \tag{3}$$

Here the convention is that the real part of the right-hand side gives the physical displacement u, and the complex amplitude u_0 may include a constant phase factor. The parameters k and ω are the angular wave number and angular frequency, respectively; also used are wave number $K = k/2\pi$ and frequency $f = \omega/2\pi$. Equation (3) represents a traveling sinusoidal wave. Displacement $u(x,t)$ repeats after distance λ such that $k\lambda = 2\pi$; thus wavelength $\lambda = 2\pi/k = 1/K$. Similarly, periodic time $T = 2\pi/\omega = 1/f$. From (2) and (3),

$$v = \omega/k = f\lambda \; . \tag{4}$$

It was mentioned that with the appropriate v, Eq. (1) describes either a wave in which u is the longitudinal displacement or a wave in which u is the transverse displacement. Not all waves are either simply longitudinal or simply transverse. In a surface wave on water and in the Rayleigh surface wave on an elastic solid, for instance, the displacement is part longitudinal, part transverse [9]. When light propagates through an anisotropic crystal, vector $\mathbf{D}$ is transverse, because $\nabla\mathbf{D} = 0$ implies $\mathbf{k}\cdot\mathbf{D} = 0$, but $\mathbf{E}$, which is not parallel to $\mathbf{D}$, is part

longitudinal, part transverse [7]. In transverse, or partially transverse, waves, the displacement can take any direction in the plane normal to the propagation direction. One must then discuss the polarization of the wave, that is, the orientation of the displacement in that plane.

The time-averaged energy density $\langle E \rangle$ in an acoustic wave is a sum of kinetic and potential (strain) contributions, and the rate of transport of energy, or intensity $I = v\langle E \rangle$, of (3) is found to be [6] $I = \frac{1}{2} Z|u_0|^2\omega^2$ where $Z = \rho v = (\rho C)^{1/2}$ is the impedance. If a traveling wave (3) in a medium of impedance Z_1 meets an abrupt interface with a medium of impedance Z_2, the transmitted and reflected amplitudes are [6] Tu_0 and Ru_0, with $T = 2Z_1/(Z_1 + Z_2)$ and $R = (Z_1 - Z_2)/(Z_1 + Z_2)$. Similar expressions hold for electromagnetic waves. If $Z_1 = Z_2$, there is impedance matching at the interface, and $T = 1$, $R = 0$. Frequently, as in optical systems, it is desirable to maximize transmission and minimize reflection at the interface between two different media. The simplest way to achieve this form of impedance matching is by the deposition of a thin *blooming layer* at the interface. In optics, a blooming layer on a lens of refractive index $(n_1 n_2)^{1/2}$ and thickness $\lambda/4$, where n_1 and n_2 are the refractive indices of the lens and of the medium in contact and λ is the optical wavelength in the layer, gives zero reflection.mm

Superposition

The wave equation (1) is linear in u, so that the sum of two solutions is itself a solution, as in (2); there is a *superposition principle*. Superposition holds in many media besides the nondispersive media of the previous section. We now discuss the consequences of superposition for waves of a single frequency ω, and the discussion includes dispersive media.

The simplest superposition is of two sinusoids of opposite velocities and equal amplitudes:

$$u(x,t) = A\sin(kx - \omega t) + A\sin(kx + \omega t) = 2A\sin kx\cos\omega t \ . \tag{5}$$

This is a *standing wave*, namely, the fixed spatial form $\sin kx$ oscillating with variable amplitude $2A\cos\omega t$. At points $kx_n = n\pi$, the nodes, u is always zero, while at $kx_a = (n + \frac{1}{2})\pi$, the antinodes, u oscillates over a maximum range. The energy density at a point on a standing wave oscillates between kinetic and potential, but there is no transport of energy, as is clear from the nature of the superposition in (5). Standing waves are excited in a medium of fixed length (more generally, volume), and boundary conditions determine where a node or antinode lies relative to the end points. Thus if transverse waves are excited on a string stretched between fixed points L apart, the end points must be nodes. The *modes* of vibration then have successively $1, 2, 3, \ldots, N, \ldots$ antinodes, wavelengths $2L/N$, and discrete frequencies $N\pi v/L$, as is seen from (4) in the form $\omega = 2\pi v/\lambda$. Often, standing waves have low N, but in typical lasers $N \sim 10^5$. This leads to some complications [6].

A second application of superposition is to interference, in which waves traveling by different paths produce interference fringes. Consider Young's slits experiment (Fig. 1), by which the wave picture of light was finally established. The waves from slits S_1 and S_2 are focused by lens L on screen F. If S_1 and S_2 radiate unending sine waves, with the same phase at each slit, the amplitude at F is

$$u = u_0 2^{1/2}(1 + \cos\delta)^{1/2} \sin\omega t \ , \tag{6}$$

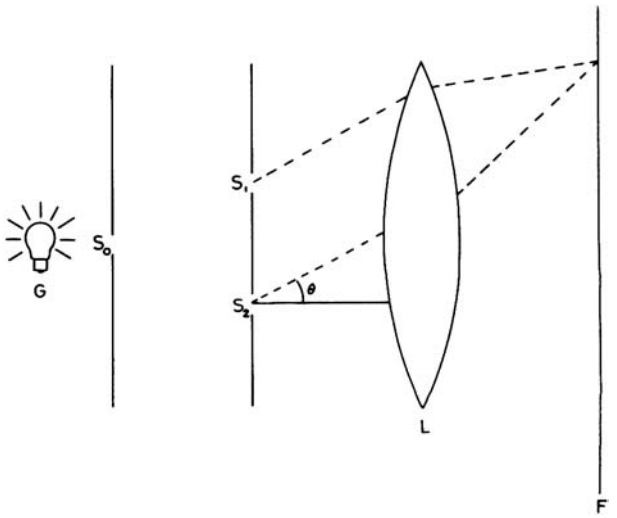

Fig. 1: Young's slits experiment.

where u_0 is the amplitude at S_1 and $\delta = 2\pi d \sin\theta/\lambda$ [6]. Intensity I is proportional to $\langle u^2 \rangle$, the time average, and so

$$I = 2I_0(1+\cos\delta) = 2I_0[1+\cos(2\pi d \sin\theta/\lambda)] \,. \tag{7}$$

What appear on F are fringes equally spaced in $\sin\theta$. In making the foregoing calculation we assume that the only effect of the different path lengths from S_1 and S_2 is to produce a phase difference, and that amplitude differences due to inverse-square-law diminution are negligible. This is called the *Fraunhofer condition*. Some beautiful examples of Young's fringes and other interference effects are found in Ref. [10]. We now return to the assumption that S_1 and S_2 radiate unending sine waves. They do not if G is an ordinary source, which emits short ($\sim 10^{-9}$ s) wave trains of random relative phase. The purpose of S_0 is to produce spatial coherence between S_1 and S_2. Waves may travel through S_1 and S_2 from any part of S_0. If the final phase difference for paths from the top of S_0 is essentially the same as that for paths from the bottom, there is spatial coherence. Thus S_0 must be narrow. The simple calculation above gives the interference for a single wave train at S_0. The observed interference pattern is the sum of many single-train patterns, and is still given by (7).

Spatial coherence, as described, is a correspondence in phase between different points in space. For some experiments, such as the Michelson interferometer [7], the requirement is for *temporal coherence*, that is, a definite phase relation between values of displacement u at different time but at one particular point. Thorough treatments of coherence are given elsewhere [7, 11].

Diffraction, for example by a single slit, a pinhole, or a grating, is handled similarly to Young's slits [6, 7]. Thus if slits S_1 and S_2 in Fig. 1 are replaced by a central slit of width d, a calculation similar to that for Young's slits shows that the Fraunhofer diffraction pattern is $I \propto \sin^2 p/p^2$ with $p = \pi d \sin\theta$. If we define an aperture function for the diffracting obstacle as equal to unity for transmitting regions and to zero for opaque regions, then for the single

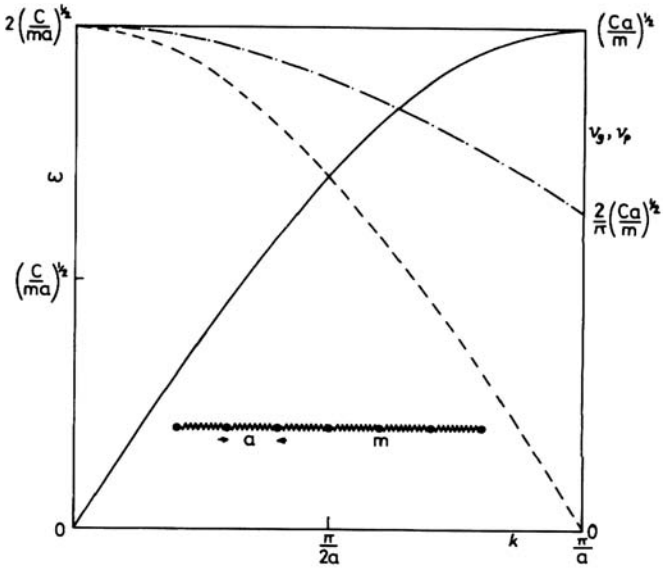

Fig. 2: Dispersion curve (—) of Eq. (13), with group (– –) and phase (–·–) velocities. Inset shows spring-mass system under consideration.

slit (as for other configurations) the intensity is the squared modulus of a Fourier transform of the aperture function [7, 12]. Of particular interest is the *diffraction grating*, which consists of N parallel slits D apart. The diffraction pattern is a series of lines at positions $\sin\theta = m\lambda/D$ ($m =$ integer). Since the spacing depends on λ, the grating can be used to resolve lines of different wavelength, and is therefore a fundamental spectroscopic tool.

Diffraction sets a limit to the resolution of optical instruments. For instance, if a telescope of aperture d is used at wavelength λ, the angular spread of the image to the first minimum of the diffraction pattern is of the order of $\theta_d = \lambda/d$. Consequently, objects with an angular separation much less than θ_d are not resolved.

Dispersive Media: Group Velocity

Group Velocity As mentioned, the wave equation (1) has the special property that all sinusoids travel at the same speed v independent of ω. In general, v does depend on ω, and in this case of dispersive propagation u obeys some equation other than (1). Thus in the mass-spring system of Fig. 2, which is a model for longitudinal acoustic waves in a crystal, the displacement u_n of the nth mass satisfies

$$m\frac{\mathrm{d}^2 u_n}{\mathrm{d}t^2} = C(u_{n-1} - 2u_n + u_{n+1}) \,, \tag{8}$$

where C is the spring constant. Again, Schrödinger's equation for a free particle of mass m is

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} \,, \tag{9}$$

where ψ is the probability amplitude. For light in a refracting medium, it is normally said that (1) holds with a frequency-dependent velocity, but this statement is only true for single-frequency light, in which case (1) may be replaced by the ordinary differential equation

$$-\frac{\omega^2}{v^2}u = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} \tag{10}$$

and (3) is the general solution.

The basic result for dispersive media concerns a *wave packet*

$$u(x,t) = \int f(\omega)\, \mathrm{e}^{\mathrm{i}(kx-\omega t)}\, \mathrm{d}\omega\ . \tag{11}$$

With an appropriate $f(\omega)$, this represents a fairly localized disturbance. Each sinusoid travels at its own phase velocity $v_\mathrm{p} = \omega/k$; consequently the relative phases change with time and the packet changes shape as it travels. For short times the change of shape is negligible, and the peak travels at the group velocity

$$v_\mathrm{g} = \mathrm{d}\omega/\mathrm{d}k\ . \tag{12}$$

The peak corresponds to in-phase superposition, so that the phase is stationary for variations of k and ω, $\mathrm{d}(kx-\omega t) = 0$ or $x\,\mathrm{d}k - t\,\mathrm{d}\omega = 0$, which gives (12). This argument, based on Ref. [13], is due to Professor R. B. Dingle; the elementary proof uses the superposition of two sinusoids [6]. The distinction between group and phase velocity can be seen in an excellent film [14] of the motion of wave packets along a water trough.

A dispersive medium is characterized by the dispersion relation $\omega = \omega(k)$ from which v_p and v_g can be derived. For the mass and spring system of Fig. 2, substitution of (3) in (8) yields

$$\omega = 2\left(\frac{C}{ma}\right)^{1/2} \sin\left(\frac{ka}{2}\right)\ . \tag{13}$$

Graphs of this dispersion relation, and of the dependence of v_g and v_p on k, are shown in Fig. 2. For the Schrödinger equation, (9), substitution of (3) yields $\omega = \hbar k^2/2m$. The group velocity is $\hbar k/m$, equal to the classical particle velocity once $\hbar k$ is identified as the momentum [6]. Other examples of dispersion relations can be found from many branches of physics [3–6].

In isotropic media it is normally stated that v_g is equal to the velocities v_E and v_I for transport of energy and information; the statement is justified by the observation that a localized disturbance moves with velocity v_g. A neat elementary proof that $v_\mathrm{E} = v_\mathrm{g}$ for waves in a bulk medium with small dissipation was given Rayleigh in the Appendix to Vol. 1 of his famous book [15]. For optics, a careful discussion needs to take account of the fact that optical dispersion has its origin in the series of resonant absorptions that characterize the medium, and for frequencies very close to a strong absorption v_E cannot be clearly defined. General proofs that $v_\mathrm{E} = v_\mathrm{g}$ when v_E is defined in a bulk dispersive optical medium have been given [16, 17]. It has been shown analytically [18] that $v_\mathrm{E} = v_\mathrm{g}$ for an optical wave (surface polariton) localized on the surface of a dispersive medium, and numerical investigation [19] gives the same result for waves traveling along an optically dispersive fiber. However, a proof applicable to all propagation geometries has not yet been given.

It is a famous principle of the theory of relativity that information cannot travel faster than the velocity of light, c. With v_I taken equal to v_g, this means that $v_g \leq c$, and it is important to realize that this places no restriction on the value of the phase velocity v_p. What the phase velocity describes is the passage of crests and troughs at the single frequency ω, and in order for information to be conveyed this passage must be modulated in some way; the modulation travels at the group velocity. An example is given by a simple waveguide, in which the dispersion is $\omega^2 = \omega_c^2 + c^2k^2$ with ω_c constant; the same relation holds propagation in an ionized plasma. It follows from the dispersion equation that $v_g v_p = c^2$ with $v_g < c$ and $v_p > c$.

Nonlinear Wave Equations: Solitons

The previous sections dealt with linear systems, and it was seen that in a dispersive linear medium, a general wave form changes shape as it travels. In a nonlinear system, by contrast, solutions can be found that maintain shape. For example, the Korteweg–de Vries (KdV) equation

$$\frac{\partial u}{\partial t} + \alpha u \frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \tag{14}$$

(α is a constant), used to describe shallow-water waves, has the solitary wave solution

$$u(x,t) = 3\alpha^{-1} v \operatorname{sech}^2[v^{1/2}(x - vt)/2] \tag{15}$$

traveling at speed v with amplitude proportional to v. This can be verified by substitution or derived if we seek solutions that depend only on $x - vt$ [20–23]. A striking property of some solitary waves is that if two collide, each emerges from the collision unaltered in shape; solitary waves of this kind, such as (15), are called *solitons*. This property, hardly expected in a nonlinear system, emerges from computer experiments, and can be seen in some explicit "two-soliton" solutions that have been found.

Because of their stability, solitons have been used as models of elementary particles. More generally, they are expected to occur in a wide variety of nonlinear physical systems including optical communication fibers.

See also: Acoustics; Diffraction; Dispersion Theory; Electromagnetic Radiation; Gratings, Diffraction; Interferometers and Interferometry; Maxwell's Equations; Nonlinear Wave Propagation; Optics, Geometrical; Optics, Nonlinear; Polarized Light; Reflection; Schrödinger Equation; Surface Waves on Liquids.

References

[1] H. C. Ohanian, *Physics*. Norton, New York and London, 1985. (E)
[2] D. Halliday and R. Resnick, *Fundamentals of Physics*, 6th ed. Wiley, New York, 2002. (E)
[3] F. S. Crawford, *Waves* (Berkeley Physics Course, Vol. 3). McGraw–Hill, New York, 1965. (E)
[4] A. P. French, *Vibrations and Waves* (M.I.T. Introductory Physics Series). Van Nostrand Reinhold, London, 1982. (E)

[5] I. G. Main, *Vibrations and Waves in Physics*, 3rd ed. Cambridge University Press, London and New York, 1993. (E)
[6] D. R. Tilley, *Waves*. Macmillan, London, 1974. (E)
[7] S. G. Lipson and H. Lipson, *Optical Physics*, 3rd ed. Cambridge University Press, London and New York, 1995. (A)
[8] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. Wiley, New York, 1998.
[9] M. G. Cottam and D. R. Tilley, *Introduction to Surface and Superlattice Excitations*, 2nd ed. Institute of Physics Publishing, London, 2004. (A)
[10] M. Cagnet, M. Francon, and J. C. Thrierr, *Atlas of Optical Phenomena*. Springer-Verlag, Berlin and New York, 1962. M. Cagnet, M. Francon, and S. Mallick, *Supplement to Atlas of Optical Phenomena*. Springer-Verlag, Berlin and New York, 1971.
[11] R. Loudon, *The Quantum Theory of Light*, 3rd ed. Oxford University Press, London and New York, 2000. (A)
[12] D. C. Champeney, *Fourier Transforms and Their Physical Applications*. Academic Press, New York, 1973. (A)
[13] H. and B. S. Jeffreys, *Mathematical Physics*. Cambridge University Press, London and New York, 1972.
[14] E. David and G. Bekow, *Gruppen und Phasengeschwindigkeit*. Film C614 of the Institut fur den Wissenschaftlichen Film, Göttingen.
[15] J. W. S. Rayleigh, *The Theory of Sound*, 2 vols. Dover,'New York, 1945.
[16] L. Brillouin, *Wave Propagation and Group Velocity*. Academic Press, New York, 1960. (A)
[17] R. Loudon, *J. Phys. A* **3**, 233 (1970).
[18] J. Nkoma, R. Loudon, and D. R. Tilley, *J. Phys. C* **7**, 3647 (1974).
[19] H. Khosravi, R. Loudon, and D. R. Tilley, *J. Opt. Soc. Am.* (to be published).
[20] A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, *Proc. IEEE* **61**, 1443 (1973).
[21] R. K. Bullough, "Solitons", in *Interaction of Radiation with Condensed Matter* (Proc. Winter College, Trieste, 1976), vol. 1, p. 382. IAEA, Vienna, 1977.
[22] G. L. Lamb, *Elements of Soliton Theory*. Wiley, New York, 1980.
[23] R. K. Dodd *et al.*, *Solitons and Nonlinear Wave Equations*. Academic, New York, 1984.

Weak Interactions

V. L. Fitch

More than 80 years of painstaking experimental work and theoretical development culminated in the early 1970s with a highly successful theory, called the standard model, which shows that the weak and electromagnetic interactions are separate manifestations of one common electroweak force. The joining of the weak and electromagnetic phenomena into one interaction constitutes the first unification of different forces since Maxwell showed electricity and magnetism to be different manifestations of the same phenomenon.

The weak interactions refer to a class of forces which are 10^{14} times weaker than the strong forces which hold the nucleus together. The neutrino, a spin-$\frac{1}{2}$ particle without electrical charge, best exemplifies the weak interactions because it reacts with other matter only through the weak forces. Whereas a neutron, a strongly interacting particle, will travel on the average

through 10 cm of iron before scattering, the neutrino will travel through 10^{16} cm. Recalling that the earth is largely iron with a diameter of 1.2×10^4 km, we see that the probability of the neutrino scattering in going through the earth is only one part in 10^6. The earth is essentially transparent to neutrinos! Indeed, at low energies the sun is nearly transparent. These are not irrelevant observations since the thermonuclear processes in stars lead to copious production of neutrinos. The mass of the neutrino has never been measured but the existing experimental limits indicate that it is tiny if not zero.

Despite the considerable weakness of the interactions we have been able to learn much about them because of two fortuitous circumstances. First, nuclear reactors and accelerators are such intense sources of neutrinos that the tiny probability of a single neutrino scattering in a detector of reasonable size is compensated by enormous fluxes. Second, the weak interactions manifest themselves in a large variety of rather common decay processes, many of which have been studied exhaustively. Due to these weak effects many radioactive nuclei decay with the emission of an electron and neutrino (the emitted electrons are called beta rays – the process is called nuclear beta decay). During the decay process, a constituent neutron transforms to a proton which remains as a part of the new nucleus, an isobar of the original. And because of the abundance of neutrons in nuclear reactors it has been possible to study the decay of the free neutron to electron, proton, and antineutrino (why antineutrino instead of neutrino will become apparent later)

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu} \,. \tag{1}$$

The mean life for this decay is 15 min. In addition, since a reactor core has the most intense concentration of neutrons, their decay provides a rich source of antineutrinos.

A well-known example of nuclear beta decay is $^{60}\mathrm{Co} \rightarrow {}^{60}\mathrm{Ni} + \mathrm{e}^- + \nu$. The mean life is 144 days. The cobalt is produced as the result of fission of uranium nuclei. Another example is a naturally occurring beta emitter, ^{40}K. Its mean life is 4.5×10^9 years. This isotope of potassium constitutes about 0.01% of the total potassium in the earth and is a source of part of the natural radioactive background. The completely stable isotopes of potassium, ^{39}K and ^{41}K, have relative abundances of 93 and 7%. When the earth was formed, one expects that ^{40}K was formed with a relative abundance between that of its two neighbors. The fact that, because of its decay, it now constitutes only one part in ten thousand of the total potassium on the earth enables us to calculate the age of the earth, a measure of the time that has elapsed since the materials of the earth were formed.

In some nuclei the relative energies of the pertinent nuclear levels permit one of the orbiting electrons to be captured by one of the constituent protons which becomes a neutron and the electron becomes a neutrino:

$$\mathrm{e}^- + \mathrm{p} \rightarrow \nu + \mathrm{n} \,. \tag{2}$$

This process is called K capture because it is the K-shell electrons which are usually involved. One should note the similarity between this reaction and neutron decay. The two processes can be related exactly when the energies of all the particles are taken into account.

Despite the complications associated with the decaying particles being submerged inside nuclei, the study of nuclear beta decay has been extraordinarily fruitful. The energy of the electrons and neutrinos is typically $\frac{1}{2}$ to 10 MeV. Their momenta are sufficiently small that

the two leptons generally do not carry off any orbital *angular* momentum. Such decays are called "allowed" transitions. However, the combined spins of the two spin-$\frac{1}{2}$ leptons may total either 0 or 1 unit of angular momentum. Correspondingly, the nuclear spin may change by 0 or 1 unit in the decay. Those beta-decay processes where the leptons come off with their spins antiparallel, 0 angular momentum, are known as Fermi (F) transitions. When one unit of angular momentum is carried off by the leptons, the spin of the nucleus can change by 0 or 1 unit; the process is called a Gamow–Teller (GT) transition. The decay of ^{35}S to ^{35}Cl is an example of an F transition, while the decay of ^{60}Co (spin 5) to ^{60}Ni (spin 4) is a pure GT transition. Neutron decay is a mixed GT and F transition. In this case, the spin of the nucleon may or may not flip when changing from neutron to proton and the lepton spins may be parallel or antiparallel.

The energy spectrum of the electrons emitted in so-called allowed beta decays is dominated by statistical considerations (phase space with rather obvious corrections for Coulomb effects – the electron being attracted by the positive nucleus from which it is escaping). For those beta decays where the spin of the nucleus changes by 2 units, the leptons must also carry off *orbital* angular momentum and the process is suppressed. These are referred to as first forbidden transitions. The example of ^{40}K mentioned above is a case of a third forbidden transition (spin change is 4). This accounts for the extraordinarily long mean life. In the simplest nuclei the study of beta decay has illuminated the weak interactions. In more complicated nuclei the beta-decay process has been used as a probe of nuclear structure.

A most important characteristic of the beta-decay process is that the leptons are emitted with the polarization equal to $-v/c$, where v/c is their speed relative to that of light. Since neutrinos have a very tiny or zero rest mass their velocity is equal to or nearly equal to c and they are correspondingly highly polarized along their line of flight. The spin of a lepton in relation to its flight direction is as a left-handed screw. Leptons are said to have left helicity. Correspondingly, antileptons are right-handed; they have right or positive helicity. The net polarization of the electron (or neutrino) emitted in beta decay is a manifestation of parity nonconservation in the weak interactions.

From reactions (1) and (2) we see that not only neutrinos but also electrons (or positrons), protons, and neutrons engage in weak interactions. A distinction is made between those weakly interacting particles which also interact strongly and those which do not. The electron and neutrino are the leptons. The neutron, proton, π, and K mesons are examples from a large class of strongly interacting particles called hadrons. In a sense the electron and its neutrino are two charge states of the same object as are the neutron and proton. Other kinds of charged leptons exist: the muon has a mass 207 times that of the electron and the tau, 3490 times, or 1.9 times the mass of the proton. Each variety of charged lepton is associated with a neutrino distinct from all the other neutrinos. The electron and mu neutrinos have been directly detected whereas the existence of the tau neutrino is still only inferred. Modern particle accelerators produce large numbers of π and K mesons both of which decay quickly to muons and neutrinos. Taus are made in electron–positron colliding-beam machines.

All evidence supports the view that the number of leptons minus the number of antileptons is a conserved quantity (conservation of leptons). Arbitrarily, the electron has been called a particle, the positron an antiparticle. The conservation of leptons in reaction (1) forces that neutrino to be an antiparticle.

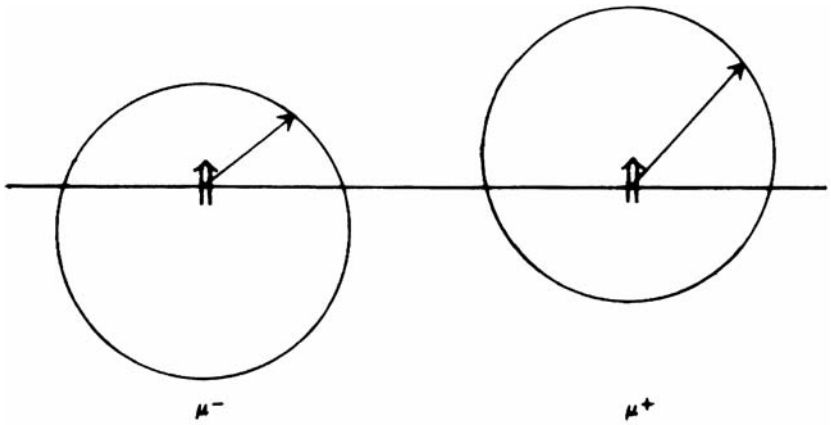

Fig. 1: The relative intensity of electrons (positrons) emitted from a source of polarized μ^- (μ^+) mesons as a function of the angle from the direction of polarization.

Parity nonconservation is most dramatic in the decay of the muon to an electron plus two neutrinos,

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \ . \tag{3}$$

Polarized muons with negative charge decaying at rest emit more electrons in the hemisphere opposite to the direction of the spin than in the reverse direction. The inverse is true for positive muons. In Fig. 1, the length of the vector indicates the number of electrons (positrons) from polarized μ^+ (μ^-) decay as a function of angle. The figure shows that the decay process is not invariant to a reflection about the line, a characteristic of parity nonconservation. It also shows that the process is not invariant under charge conjugation – the operation that transforms world to antiworld, in this case, μ^- to μ^+.

When it was first discovered that parity nonconservation occurred, leading to asymmetric decays, it was thought the process could be used to establish an absolute direction in space in violation of Mach's principle. However, it is noted from Fig. 1 that the combination of parity reversal, *P*, and charge conjugation, *C*, restores the fundamental symmetry. The fact that one cannot, with perfect *CP* symmetry, distinguish a distant galaxy from a distant antigalaxy would prevent one from establishing a preferred direction in space.

An important example of a weak interaction is the decay of the *K* meson (kaon) to two or three π mesons (pions). Indeed, the observation that the kaon, which has zero spin, could decay to three pions in a state of zero relative angular momenta and also decay to two pions was the first evidence that parity was not conserved in the weak interactions (the pion has odd intrinsic parity). Detailed studies show that not only is parity not conserved in this decay but also the combined operations of charge conjugation and parity, *CP*, is not an exactly conserved quantity. The effect is small (the *CP*-nonconserving effect is about 2×10^{-3} of the *CP*-conserving process) but it has been exhaustively studied and the detailed parameters are extremely well known.

Modern field theory depends on the validity of a very general theorem which says that all interactions must be invariant under the combined operations of charge conjugation, parity, *and* time reversal, *CPT*. Accordingly, a violation of *C* and *P* invariance requires a compensating violation of time-reversal invariance. In fact, detailed experimental studies of the neutral kaon decay strongly support the *CPT* theorem. This means that, to a small but finite degree, fundamental interactions are different depending on the direction of time. The absence of complete *CP* symmetry also reopens the question of the complete validity of Mach's principle.

As noted above, nuclear reactors are sources of antineutrinos and, because of their abundance, the reaction

$$\nu_e + p \rightarrow e^+ + n \tag{4}$$

has been extensively studied. Indeed, the production of positrons using reaction (4) provided the first direct evidence for the existence of neutrinos.

It was thought for many years that the exchange of electrical charge in reaction (4) between the leptons was a unique characteristic of weak forces. However, it is now known that the reaction proceeds without charge exchange about 25% of the time. Similar reactions occur for the mu-like neutrinos, viz.,

$$\nu_\mu + n \rightarrow \mu^- + p \tag{5}$$

or

$$\bar{\nu}_\mu + p \rightarrow \mu^+ + n \tag{6}$$

and without charge exchange, e. g.,

$$\nu_\mu + p \rightarrow \nu_\mu + p\ . \tag{7}$$

The theoretical understanding of the weak interactions was guided from the earliest days by the work of Fermi. The Fermi theory was drawn in close analogy with electrodynamics. As a charged particle interacts with the electromagnetic (photon) vector field, so a weakly interacting particle, e. g., a neutron, interacts with a lepton field. The theory enjoyed enormous success. The only change since the original work of Fermi was the addition of axial-vector interactions with a strength nearly equal to the original vector interaction assumed by Fermi. This addition produced the effects of parity nonconservation. The modified Fermi theory was very successful in accounting for low-energy weak interaction phenomena but predicted disturbing behavior at high energies – for example, a neutrino–nucleon cross section growing with energy without limit. To avoid these problems it was proposed that the weak interactions were mediated by massive bosons. These bosons could either be charged and account for the charge changing reactions through "charged currents" or neutral in analogy with the interaction of "neutral" currents in electrodynamics. The standard model evolved out of these observations.

Despite their similarity, these two forces are grossly different with respect to their range. The electromagnetic interaction is long range (the electrical potential depends on distance as $1/r$), whereas the weak interaction is short range, indeed, much shorter even than the strong

interactions. The electrical force between charged particles is due to the exchange of photons which have zero rest mass. A shortened range can be arranged by requiring that the exchanged particle have a nonzero mass. From the simplest point of view, the exchange of a particle of rest mass m leads to an interaction potential energy

$$U(r) = (e^2/r)\exp(-rmc/\hbar)\ , \tag{8}$$

where r is the range, $\hbar$ is Planck's constant, and c is the velocity of light. As $m \to 0$ it reduces to the Coulomb potential. By an increase in the mass of the exchanged particle the range of the interaction can be reduced to any arbitrary value. For example, if m equals the mass of the pion, the range is characteristic of nuclear forces, 1.4×10^{-13} cm. In applying these ideas to the weak interactions, to produce an effect with a given magnitude, one may vary the weak charge e_{w} or the range, i. e., the mass of the exchanged particle. For the same effect the "weak charge" could be very large if the characteristic range of the force is small. Therefore, unlike the electrical charge, it is necessary to specify not only the magnitude of the "weak" charge but also the volume in which it is effective. The weak charge squared is measured to be $e_{\mathrm{w}}^2 = 0.88 \times 10^{-37}$ eV cm (for comparison, the electromagnetic charge squared is $e^2 = 1.44 \times 10^{-7}$ eV cm.) Over a typical nuclear volume this leads to a weak interaction potential of about 1 eV. However, experiments show the weak forces to be confined to very much smaller distances than nuclear dimensions and the question is – just how short is the range? Because the fundamental structure of the weak and electromagnetic interaction is similar, aside from the range of the force, a natural possibility is that the basic strength of the weak and electromagnetic charge is the same. This necessitates that the characteristic range of the weak force be set in the region of 3×10^{-16} cm corresponding to a mass of 70–80 GeV for the exchanged particle.

The standard model leads to the prediction of a charged vector boson, the $W^{\pm}$, which accounts for those weak interactions in which the lepton charge is changing, and the Z_0 which is involved in the neutral interactions. The theory predicts that the masses of these vector bosons are related through the Weinberg mixing angle, θ_{W}, viz.,

$$m_W = 137.3\,\mathrm{GeV}/\sin\theta_{\mathrm{W}} \qquad \text{and} \qquad m_Z = m_W/\cos\theta_{\mathrm{W}}\ .$$

The angle, θ_{W}, has been measured in a wide variety of the neutral current processes and is found to be $\sin\theta_{\mathrm{W}} = 0.230 \pm 0.005$.

With the advent of particle accelerators which have sufficient energy, both of these particles are observed with $m_W = 81 \pm 1.4$ GeV and $m_Z = 91.10 \pm .05$ GeV. These measured values are in striking agreement with those predicted from the standard model.

Since hadrons are composed of quarks, the weak interactions involving hadrons are now described in terms of these more fundamental entities. A neutron is composed of two d quarks and one u quark, whereas the proton is made of two u quarks and one d quark. The beta decay of the neutron then involves, in this picture, a transformation of a d- to a u-type quark, viz.,

$$d \to u + \mathrm{e}^- + \nu\ . \tag{9}$$

Because they are composed of a limited number of primitive quarks, striking symmetries appear among the hadrons and since weak interactions occur between quarks as well as lep-

tons, similar symmetries in the interactions appear. A table of quarks and leptons displays a remarkable correspondence:

Lepton	Charge	Quark	Charge
ν_e	0	u	$\frac{2}{3}$
e	-1	d	$-\frac{1}{3}$
ν_μ	0	c	$\frac{2}{3}$
μ^-	-1	s	$-\frac{1}{3}$
ν_τ	0	t	$\frac{2}{3}$
τ^-	-1	b	$-\frac{1}{3}$

As seen from the table, for each lepton there appears a corresponding quark with its electrical charge displaced by $+\frac{2}{3}$. This striking symmetry between the leptons and quarks suggests that the strong as well as the weak and electromagnetic interactions be included in the unification process. The symmetry also suggests that should new leptons be discovered the chances are high that corresponding quarks exist.

A question of most fundamental importance is whether the table above is complete or whether there exist still heavier charged leptons with their associated neutrinos. Studies of the Z_0 decay reveal that, most likely, there exist *no more than the three generations of leptons* as listed in the table. This conclusion comes from the following considerations. Since it couples to neutral currents, the Z_0 can decay to neutrino–antineutrino pairs. In fact, pure neutrino decay contributes significantly to the decay rate and therefore to the width of the Z_0 resonance which is observed in electron–positron colliders. The conclusion from the data is that there are 3.1 ± 0.2 generations of neutrinos. Such conclusions should hold except in the unlikely event that any new neutrino would be so massive that the Z_0 could not decay into them.

The old Fermi theory, modified with axial-vector contributions, required that the neutrino mass be strictly zero. The standard model imposes no such constraint and there is great interest in the question of neutrino masses. It is now known that the electron neutrino has a mass less than 10 eV, the mu neutrino less than 0.25 eV, and the tau neutrino less than 35 MeV. Pursuing the parallelism between quarks and leptons noted in the table, since the weak interactions allow transitions between the different flavored quarks could it not be that transitions also occur between the different generations of neutrinos especially since they can have finite and different masses? Such transitions would lead to oscillations between the different generations of neutrinos.

There is now good evidence that oscillations do occur. For a long time it has been known that the number of neutrinos from the sun is less than half of what is expected on the basis of solar models, suggestive but not definitive evidence for oscillations. An experiment has been conducted in a deep mine in Sudbury, Ontario, Canada, using deuterium as a solar neutrino target. The rate of deuterium breakup is measured with and without the emission of an electron. Only the electron type neutrinos could lead to the emission of electrons in the breakup process whereas any flavor of neutrino would lead to the breakup without electron emission. From a comparison of the relative occurences of these two breakup reactions the evidence is compelling that more than one flavor of neutrino is passing through the detector.

Recent experiments at the Super-Kamiokande underground detector in Japan have measured the flux of high energy (multi-GeV) neutrinos produced when cosmic rays interact in the atmosphere of the earth. The number of neutrinos, nearly all ν_μ, passing downward through the detector is compared with the number passing upwards. Those passing upwards, of course, have had to pass the extra distance through the diameter of the earth. The numbers are not the same, the flux of those passing upwards is about half that of those passing downward. Auxillary information suggests that the ν_μ neutrinos are oscillating into ν_τ. Detailed analysis of the data shows that the difference in the mass of the ν_μ and ν_τ, squared, lies between $1.3 \times 10^{-3}\,\mathrm{eV}^2$ and $3.0 \times 10^{-3}\,\mathrm{eV}^2$. The mixing is close to maximum.

The weak interaction between the quarks is quantified through a mass mixing matrix which arises because the mass eigenstates are not the same as the weak eigenstates. All of the mixing is expressed through a 3×3 unitary matrix operating on the charge-$\frac{1}{3}$ quark states (d, s, and b), viz.,

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

This matrix is identified with the names of Cabibbo, Kobayashi, and Maskawa, the CKM matrix. When it was first proposed only three flavors of quarks were known, the u, d, and s. In a remarkably prescient paper, Kobayashi and Maskawa pointed out that if, in fact, six quarks existed there was room in the matrix for a phase angle that would characterize *CP* violation. Now that six quarks have been experimentally identified and measured, c, b, and t in addition to the original u, d, and s, *CP* violation has as natural home.

Much experimental effort has been directed toward evaluating the various elements in the matrix. These are determined from the weak decay rates of the various quarks which are, in turn, extracted from the decay of the relevant mesons, and in some instances, from neutrino scattering. The details are beyond the scope of this review but may be found in the references below.

The decay of the neutral B meson is of special significance. It was predicted that it would show a large *CP* violation. Indeed, this has been the case. It has now been observed in the system of neutral B mesons at the electron–positron colliders at Stanford and KEK in Japan.

A full discussion of the this new and exciting development is contained in the last reference below.

See also: Beta Decay; Currents in Particle Theory; Electron; Elementary Particles in Physics; Grand Unified Theories; Leptons; Mesons; Neutrino; Parity; Positron; Quarks; Weak Neutral Currents.

Bibliography

E. D. Commins and P. H. Buchsbaum, *Weak Interactions of Leptons and Quarks*. Cambridge University Press, Cambridge, New York and Melbourne, 1983.

Particle Data Group, "Review of Particle Physics", *Phys. Lett. B*, **592**, issues 1–4, July 2004.

Weak Neutral Currents

A. K. Mann

Introduction

Weak neutral currents were first observed and their strength determined quantitatively in 1973 [1]. The interaction between any two weak neutral currents gives rise to much sought-after and initially elusive reactions characterized in present thinking by the exchange between the currents of a virtual, massive, electrically neutral vector boson, the Z^0. It had long been known that the exchange of a virtual, massless, electrically neutral vector boson, the photon, between electromagnetic currents – which are "neutral" in this terminology – described electromagnetic reactions. However, prior to 1973, only the reactions arising from the interactions of two weak charged currents, mediated by the exchange of virtual, massive, electrically charged vector bosons, the $W^{\pm}$, had been observed. Feynman diagrams schematically illustrating these processes are shown in Fig. 1.

Somewhat earlier, a unified theory of quantum electrodynamics and quantum weak dynamics had been formulated [2] in which an integral constituent was the weak neutral current. The success of this theory, now known as the electroweak theory (EWT), became apparent with the experimental discovery of weak neutral currents (WNC) and is exhibited in the ability of the EWT to describe in precise detail the wide variety of phenomena which have been intensively studied experimentally since 1973. The junction of the EWT and experiment culminated in the direct observation in 1983 of the massive vector bosons, the $W^{\pm}$ and Z^0, at the mass values predicted by the EWT [3], the predictions of which require as input the measured strength of the WNC found in WNC experiments.

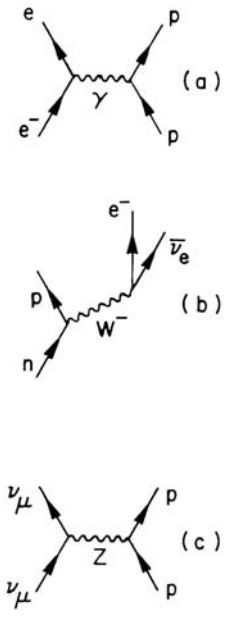

Fig. 1: Feynman diagram for photon exchange in quantum electrodynamics (a). In the unified electroweak theory, the weak interactions are mediated by massive vector bosons: the charged W for β decay (b), and the neutral Z for elastic neutrino scattering (c).

More generally, the EWT provides a unified, renormalizable, gauge-invariant description of weak charged current (WCC) and electromagnetic neutral current (ENC) interactions, as well as WNC interactions [4]. It is, however, WNC phenomena and the $W^{\pm}$ and Z^0 masses which contribute the major quantitative tests of the predictions of the EWT. These extend over the wide range of squared momentum transfer from $10^{-6}\,(\mathrm{GeV}/c)^2$ to $10^{4}\,(\mathrm{GeV}/c)^2$ and encompass such diverse processes as deep inelastic neutrino and longitudinally polarized electron scattering from isoscalar and nonisoscalar targets, ν–p scattering, ν–e scattering, parity violation in atoms via an induced transition electric dipole moment, asymmetries in high center-of-mass energy $\mathrm{e}^+ + \mathrm{e}^- \to \mu^+ + \mu^-$ reactions, and the masses of $W^{\pm}$ and Z^0.

In what follows we record in the next section the EWT expressions for various WNC interactions [5], primarily to illustrate their structure and emphasize the importance of radiative corrections in the theory. We then summarize the comparisons of EWT predictions with experimental data. Finally, a short discussion is given of future exploration of still higher-precision WNC measurements as more stringent tests of the EWT and as sensitive probes of new physics beyond the EWT.

EWT Expressions for WNC Interactions [5]

The effective Lagrangian for WNC interactions between massless neutrinos (ν) and quarks (q) is

$$-L^{\nu\mathrm{H}} = \frac{G_\mathrm{F}}{\sqrt{2}}\bar{\nu}\gamma^\mu(1+\gamma_5)\nu J^\mathrm{H}_\mu \,, \tag{1}$$

where G_F is the Fermi coupling constant ($= 10^{-5}/M_\mathrm{p}^2$),γ^μ and γ_5 are Dirac matrices, and

$$J^\mathrm{H}_\mu = \sum_i \bar{q}_i\gamma_\mu(g^i_V + g^i_A\gamma_5)q_i \,. \tag{2}$$

Here, identical V, A couplings are assumed for all neutrino flavors, and, with $g^i_{V,A} \equiv \varepsilon_L(i) \pm \varepsilon_R(i)$, flavor independence of the quark couplings is also assumed, i. e., $\varepsilon_{L,R}(u) = \varepsilon_{L,R}(c)$, etc., where u and c are "up" and "charm" [5] quarks, respectively.

In the EWT the $\varepsilon_L(i)$ and $\varepsilon_R(i)$ are completely specified in terms of constants, calculable radiative-correction parameters, and the only undetermined parameter in the theory, the relative coupling strength of the WNC, usually written as the function $\sin^2\theta_\mathrm{W}$ of the angle θ_W.

The effective Lagrangian for the WNC interaction between muon-type neutrinos (and autineutrinos), ν_μ (and $\bar{\nu}_\mu$), and electrons is

$$-L^{\nu\mathrm{e}} = \frac{G_\mathrm{F}}{\sqrt{2}}\bar{\nu}_\mu\gamma^\mu(1+\gamma_5)\nu_\mu J^\mathrm{e}_\mu \,, \tag{3}$$

where

$$J^\mathrm{e}_\mu = \bar{\mathrm{e}}\gamma_\mu(g^\mathrm{e}_V + g^\mathrm{e}_A\gamma_5)\mathrm{e} \tag{4}$$

and g^e_V and g^e_A are, again in the EWT, given in terms of $\sin^2\theta_\mathrm{W}$ and radiative-correction parameters.

The parity-violating (but *CP*-conserving) interaction between electrons and quarks is given by

$$-L^{\mathrm{eH}} = \frac{G_{\mathrm{F}}}{\sqrt{2}} \sum_i (C_{1i} \bar{\mathrm{e}} \gamma_\mu \gamma_5 \mathrm{e} \bar{q}_i + C_{2i} \bar{\mathrm{e}} \gamma_\mu \mathrm{e} \bar{q}_i \gamma^\mu \gamma_5 q_i) \ . \tag{5}$$

In the EWT, the constants C_{1u}, C_{1d}, and C_{2u} depend on $\sin^2 \theta_{\mathrm{W}}$ and radiative-correction parameters.

Finally, in the EWT the $W^{\pm}$ and Z^0 masses are predicted to be

$$M_W = \frac{A_0}{\sin \theta_{\mathrm{W}} (1 - \Delta r)^{1/2}} \ , \tag{6}$$

$$M_Z = \frac{M_W}{\cos \theta_{\mathrm{W}}} \ , \tag{7}$$

where

$$A_0 = \left(\frac{\pi \alpha}{\sqrt{2} G_{\mathrm{F}}} \right)^{1/2} = 37.281 \, \mathrm{GeV}/c^2 \ , \tag{8}$$

and Δr is a calculable radiative-correction parameter.

It is customary to define the so-called renormalized weak angle in the EWT as

$$\sin^2 \theta_{\mathrm{W}} \equiv 1 - M_W^2 / M_Z^2 \ . \tag{9}$$

The effective Lagrangians in Eqns. (1) through (8) and the definition of $\sin^2 \theta_{\mathrm{W}}$ in Eq. (9) are sufficient to describe quantitatively all the WNC reactions mentioned in the Introduction. In principle, once a value of $\sin^2 \theta_{\mathrm{W}}$ is extracted from a given class of experiments and an appropriate radiative correction applied to the empirical value, all other classes of experiments then yield (when radiatively corrected) values of $\sin^2 \theta_{\mathrm{W}}$ which serve to test in detail the correctness and self-consistency of the EWT over a multiplicity of phenomena and a wide range of energies and momentum transfers. This is the strategy employed to compare the results of one class of experiments with another class and to facilitate and make clear comparison of experiment with the EWT.

WNC Data and the EWT

In this section we briefly summarize WNC experimental results and compare them with the predictions of the EWT.

In Fig. 2 are shown values of $\sin^2 \theta_{\mathrm{W}}$ obtained from various classes of WNC experiments and measurements of the W and Z masses plotted against Q^2, the square of the four- momentum transfer typical of each class of experiment [7J. Each point in the plot is labeled by the experiment type that produced it. Note the wide range of Q^2 and the convergence of the different determinations of $\sin^2 \theta_{\mathrm{W}}$ on the single, universal value shown by the horizontal line in Fig. 2 at $\sin^2 \theta_{\mathrm{W}} = 0.230 \pm 0.0048$.

The data in Fig. 2 are also summarized in Table 1 [7] which gives, in addition to the precise numerical values of $\sin^2 \theta_{\mathrm{W}}$ from the different classes of experiments, values of $\sin^2 \theta^0$, i. e.,

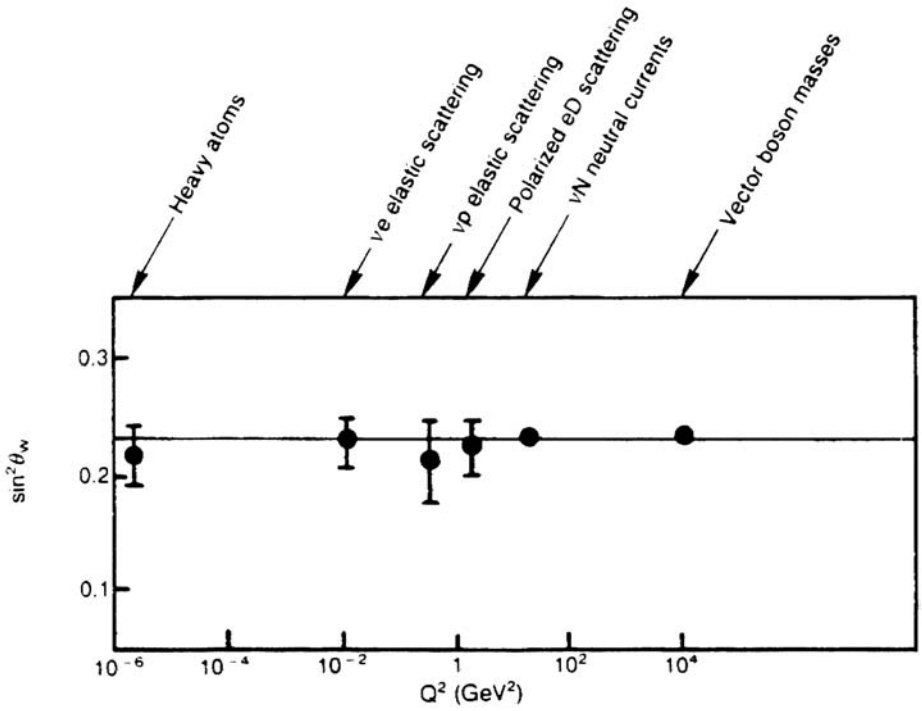

Fig. 2: Various determinations of $\sin^2\theta_W$, the fundamental unspecified parameter of the electroweak theory, plotted against Q^2, the square of the four-momentum transfer typical of each class of experiments. Each determination is labeled by experiment type (see Table 1). The horizontal line is at $\sin^2\theta_W = 0.230 \pm 0.0048$. Reprinted from Ref. [7].

the values obtained directly from experiment without the radiative corrections provided by the EWT. One sees the importance of the radiative corrections which are approximately 4% in the determination of $\sin^2\theta_W$ from the parity violation in heavy atoms, approximately 7% in extracting $\sin^2\theta_W$ from measurement of W and Z masses, and less than 1% in $\nu_\mu e$ scattering.

In addition to the wide range of Q^2 shown in Fig. 2, the wide diversity of the phenomena and the experimental techniques represented by the different classes of measurement in that figure should be emphasized. For example, parity violation in heavy atoms, at the extreme low end of momentum transfer, originates through a modified transition probability for the scattering of circularly polarized photons by a gas of heavy atoms. The modification is due to an induced transition electric dipole moment generated by the parity-violating WNC through exchange of a virtual Z^0 boson between electron and nucleus in addition to virtual photon exchange. The experiment consists in measuring very small rotations of the axis of polarization of the incident circularly polarized light following its passage through the atomic gas. Compare this with production of the massive Z^0 boson ($M_Z \simeq 91\,\mathrm{GeV}/c^2$) in $\bar{p}p$ collisions at center-of-mass energy of 560 GeV, and detection of the multi-GeV e^+ and e^- from the subsequent decay, $Z^0 \to e^+e^-$.

Table 1: Determination of $\sin^2\theta_W$ from Various Reactions [7].

Reaction	$\sin^2\theta_W$ [a]	$\sin^2\theta_W$ [b]
Deep inelastic $\nu_\mu N$ scattering	$0.233 \pm 0.003 \pm [0.005]$	0.242
$\nu_\mu p \to \nu_\mu p$	0.210 ± 0.033	0.208
$\nu_\mu e \to \nu_\mu e$	$0.223 \pm 0.018 \pm [0.002]$	0.221
W and Z masses	$0.229 \pm 0.007 \pm [0.002]$	0.214
Parity violation in heavy atoms	$0.220 \pm 0.007 \pm [0.018]$	0.212
Polarized eD scattering	$0.221 \pm 0.015 \pm [0.013]$	0.226
All data	0.230 ± 0.0048	

Data from Ref. [5]. See also Fig. 2.

[a] Where two uncertainties are shown, the first is experimental, and the second, in squared brackets, is theoretical. The latter includes the effect of letting the unknown masses of the top quark and Higgs boson range widely. The central values assume $M_1 = 45$ and $M_{11} = 100\,\text{GeV}$.

[b] Values that would be obtained from the data without radiative corrections.

Table 2: Values of the Model-Independent Neutral-Current Parameters Compared With the Prediction for $\sin^2\theta_W = 0.230$. After Ref. [7].

Quantity	Experimental Value	Prediction
$\varepsilon_L(u)$	0.339 ± 0.017	0.345
$\varepsilon_L(d)$	-0.429 ± 0.014	-0.427
$\varepsilon_R(u)$	-0.172 ± 0.014	-0.152
$\varepsilon_R(d)$	$0.011 \pm {}^{0.081}_{0.057}$	0.076
g_A^e	-0.498 ± 0.027	-0.503
g_V^e	-0.044 ± 0.036	-0.045
C_{1u}	-0.249 ± 0.071	-0.191
C_{1d}	0.381 ± 0.064	0.340
$C_{2u} - \frac{1}{2}C_{2d}$	0.19 ± 0.37	-0.039

To compare experiment and theory in still greater detail, we show in Table 2 the experimental values of the quantities appearing in the effective Lagrangians and the values predicted by the EWT [7] when the value of $\sin^2\theta_W$ is taken to be 0.230.

The content of Fig. 2 and Tables 1 and 2 may be summarized as demonstrating the "symbiotic" relationship between WNC phenomena and the EWT. The experimental tests provided by these phenomena and passed successfully by the EWT help to establish the theory as the quantitatively correct description of weak and electromagnetic interactions at the present level of experimental error.

Future Progress

There exists the prospect that the precision of several of the classes of experiments discussed in the last section might be substantially improved in the coming decade. The precision of $\sin^2\theta_W$ determined from a single class of experiments is of the order of 5–7% at present,

but improvements to the level of 1% or better are likely to be made in the near future. This increased precision will in turn stimulate calculations of higher-order radiative corrections. Taken together, the result will be that WNC processes will serve not only as tests of the validity of the EWT, but also as probes of new physics beyond the EWT if any significant discrepancy between experiment and theory is found.

See also: Currents in Particle Theory; Elementary Particles in Physics; Grand Unified Theories; Weak Interactions.

References

[1] F. J. Hasert *et al.* (Gargamelle Collaboration). *Phys. Lett.* **46B**, 121, 138 (1973); A. Benvenuti *et al.* (HPW Collaboration), *Phys. Rev. Lett.* **32**, 800, 1454, 1457 (1974).

[2] S. L. Glashow, *Nucl. Phys.* **22**, 579 (1961). S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967). A. Salam, in *Elementary Particle Theory* (N. Svartholm, ed.), p. 367. Almqvist and Wiksells, Stockholm, 1969.

[3] G. Arnison *et al.* (UA1 Collaboration), *Phys. Lett.* **166B**, 484 (1986); C. Albajar *et al.*, *Z. Phys.* **C44**, 15 (1989); R. Ansari *et al.* (UA2 Collaboration), *Phys. Lett.* **186B**, 440 (1987).

[4] P. W. Higgs, *Phys. Rev. Lett.* **12**, 132 (1964); **13**, 321 (1964); *Phys. Rev.* **145**, 1156 (1966); G. 't Hooft and M. Veltman, *Nucl. Phys.* **B50**, 318 (1972), and references therein.

[5] U. Amaldi *et al.*, *Phys. Rev. D* **36**, 1385 (1987); G. Costa *et al.*, *Nucl. Phys.* **B297**, 244 (1988); P. Langacker, *Phys. Rev. Lett.* **63**, 1920 (1989).

[6] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev. D* **2**, 1285 (1970); J. J. Aubert *et al.*, *Phys. Rev. Lett.* **33**, 1404 (1974); J. E. Augustin *et al.*, *Phys. Rev. Lett.* **33**, 1406 (1974); A. Benvenuti *et al.*, *Phys. Rev. Lett.* **34**, 419 (1975); E. J. Cazzoli *et al.*, *Phys. Rev. Lett.* **34**, 1125(1975).

[7] P. Langacker and A. K. Mann, *Phys. Today* **42**, 22 (1989).

Whiskers

R. V. Coleman

The term *whisker* describes any fibrous growth of a solid, and such forms were studied extensively in relation to the development of microscopic theories of crystal growth, particularly those involving screw dislocations. If crystal growth proceeds by the motion of a dislocation step and if conditions are such that the rate of step generation dominates over the motion of steps away from the dislocation source, then a whisker profile results. This is a simple and elegant explanation for the growth of single-crystal whiskers, but many modifications and alternative theories involving oxidation, solid-state diffusion, stress recrystallization, or some combination of these have been developed. Impurities may also play a significant role.

Whisker crystals can be produced by vapor deposition, chemical reaction, electrolytic deposition, or oxidation of surfaces, and at vapor–liquid–solid interfaces. A famous form of whisker growth is that of Sn whiskers growing from tin-plated metal where applied stress can enhance the growth rate, and this has led to the term *squeeze whisker*. Low-melting-point met-

als such as Zn, Cd, Mg, Hg, and K have been grown from vapor, while higher-melting-point metals and semiconductors such as Fe, Co, Ni, Cu, Au, Ag, Pt, Si, and Ge are often grown by the hydrogen reduction of metallic salts. Whiskers of oxides, carbides, nitrides, metallic salts, graphite, polymers, organic materials, and metallic alloys have been reported.

Whisker-like crystals also result from the growth of very anisotropic solids such as the quasi-one-dimensional solids $NbSe_3$ and TaS_3 which contain linear chains of metal atoms. This extreme structural anisotropy is also reflected in the electronic structure which undergoes a phase transition to a charge-density-wave state at lower temperature characterized by a superlattice of charge modulation. Quasi-one-dimensional organic compounds such as $(TMTSF)_2PF_6$ also grow as whiskers and exhibit a phase transition to a spin-density-wave state characterized by a superlattice modulation of spin. Quasi-two-dimensional compounds such as the high-T_c superconductor $Bi_2Sr_2Ca_1Cu_2O_x$ can also be grown in single-crystal whisker form.

The high surface and volume perfection of whiskers originally made them very useful for the study of mechanical properties. Stress–strain curves were obtained with elastic strain regions extending to 4 or 5%. The crystal perfection of whiskers has also made them excellent samples for the study of a wide range of electric and magnetic properties in solids. Many of these experiments are discussed in the review articles listed in the Bibliography, which also contain extensive reference to the original work.

See also: Crystal Defects; Crystal Growth.

Bibliography

R. H. Doremus, B. W. Roberts, and D. Turnbull (eds.), *Growth and Perfection of Crystals*. Wiley, New York, 1958, and Chapman and Hall, London, 1958. (A)

J. J. Gilman (ed.), *The Art and Science of Growing Crystals*. Wiley, New York and London, 1963. (A)

E. M. Nadgornyi, "The Properties of Whiskers", *Sov. Phys.–Usp.* **5**, 462 (1962). (I)

R. V. Coleman, "The Growth and Properties of Whiskers", *Metall. Rev.* **9**, 261 (1964). (I)

C. C. Evans, *Whiskers*. Mills and Boon, London, 1972. (I)

Work Function

T. M. Donovan and A. D. Baer

The work function W is the minimum energy required to remove an electron at the Fermi energy from a solid. Figure 1 shows a simplified model of the surface of metal, with electronic states in the metal occupied up to the Fermi level. The work function is

$$W = E_v - E_f \,,$$

where E_v is the energy of an electron at rest outside the solid, and E_f is the energy of an electron at the Fermi energy inside the solid.

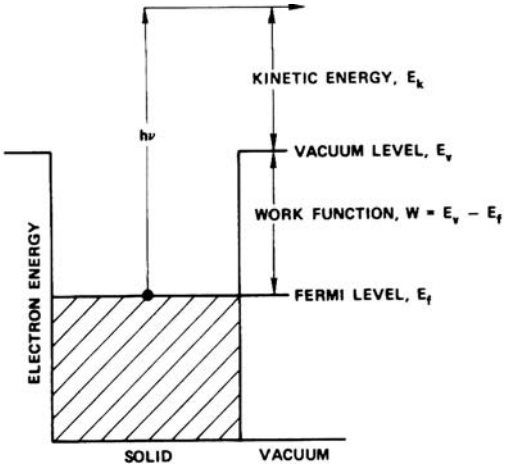

Fig. 1: Simplified model of electrons in a metal showing electron emission.

The work function is determined experimentally by measuring the minimum amount of excitation energy required to emit an electron from the solid. There is a high density of filled electronic states near the Fermi level of a metal, and clear-cut theoretical relationships exist which relate photoemission and thermionic emission of electrons from these states to the work function [1–3]. For example, in Fig. 1, if an electron is excited from the Fermi level by a photon of energy $h\nu$, and the electron is emitted from the solid into vacuum without energy loss, then

$$W = h\nu - E_{\mathrm{k}} \ ,$$

where E_{k} is the kinetic energy of the electron outside the solid.

Since the Fermi level of a semiconductor or insulator generally lies in the band gap, the work function of a nonmetal sample is usually measured by first determining the work function of a metal surface and then measuring the difference between the work functions of the metal and the sample. The difference is equal to the contact potential difference which can be measured directly in a Kelvin probe [2, 3] or retarding-potential experiment [4, 5].

See also: Electron Energy States in Solids and Liquids; Photoelectron Spectroscopy; Thermionic Emission.

References

[1] R. H. Fowler, *Phys. Rev.* **38**, 45 (1931). (I)
[2] A. J. Decker, *Solid State Physics*, pp. 211, 230. Prentice-Hall, Englewood Cliffs, 1957. (E)
[3] J. C. Riviere, *J. Solid State Surf. Sci.* **1**, 179 (1969). (A)
[4] L. Apker, E. Taft, and J. Dickey, *Phys. Rev.* **74**, 1462–1474 (1948). (I)
[5] C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1030, (1964); **136**, A1044(1964).