

CP (Charge Conjugation Parity) Violation

John F. Donoghue
Barry R. Holstein
University of Massachusetts

I. CP Symmetry
II. Observation of CP Violation
III. Mechanisms of CP Violation
IV. Recent Progress
V. Future Areas of Research

GLOSSARY

B meson Elementary particle containing one heavy quark called a b quark as well as a lighter antiquark, or a $\bar{b}$ antiquark, and a light quark. The B mesons are about five times more massive than a proton.

Charge conjugation invariance Invariance of physical laws under the process of interchanging particle and antiparticle.

Decay Elementary particles can transform into other combinations of particles as long as energy, momentum, charge, etc., are conserved. The process of an isolated particle transforming into several lighter particles is called *decay*.

Electric dipole moment Classically, an electric dipole moment is a separation of charges so that, although the whole system is electrically neutral, the distribution of charge has a region of positive charge and a region of negative charge separated along some axis.

Kaon Elementary particle with a mass of about one-half of that of the proton. The kaon is the lightest of the particles, with a quantum number called "strangeness" and containing a "strange" quark or antiquark.

Parity invariance Invariance of physical laws under the process of reversing spatial coordinates. If accompanied by a 180° rotation, this is equivalent to a reflection in a mirror.

CHARGE CONJUGATION PARITY (CP) violation is said to occur when two processes, which differ by the combined action of charge conjugation and parity reversal, do not occur at the same rate. This phenomenon is rare, but it has been observed in the decay of the neutral K meson system. The origin of this slightly broken symmetry is not currently understood, and it may tell us more about the structure of the fundamental interactions.

I. CP SYMMETRY

From ancient times, the concept of symmetry has commanded a powerful influence upon our view of the universe. However, many symmetries are only approximate, and the way in which they are broken can reveal much about the underlying dynamics of physical law. Perhaps the earliest example of a broken symmetry was the required modification of the presumed perfect circular orbits of the outer planets by epicycles—circles on circles—in order to explain the observation that occasionally the trajectories of theses planets through the sky double back on themselves. This breaking of perfect symmetry, although small, forced scientists to search more deeply into the basic forces responsible for celestial orbits, leading ultimately to Newton's law of universal gravitation.

More recently, in the mid-1950s, the concept of parity invariance—left-right symmetry—was found to be violated by the weak interaction, that is, the force responsible for such processes as nuclear beta decay. The concept of right or left in such a process is realized by particles whose direction of spin is respectively parallel or antiparallel to the particle momentum. Wrapping one's hand around the momentum vector with fingers pointing in the direction of rotation, as in Fig. 1, the particle is said to be right-left-handed if the right/left thumb points in the direction of the momentum vector. Parity invariance would require the absence of handedness, that is, the emission of equal numbers of both right- and left-handed particles in the decay process. Beta decay processes occur with both electron (e^-) and positron (e^+) emission:

$$A \rightarrow B + e^- + \bar{\upsilon}_e$$

$$B' \rightarrow A' + e^+ + \bar{\upsilon}_e$$

where υ_e ($\bar{\upsilon}_e$) is the accompanying neutrino (antineutrino). In such decays, the neutrinos or antineutrinos are found to be completely left- or right-handed, indicating a maximal violation of parity. We now understand this as being due to the left-handed character of the particles that mediate the weak interactions, the W and Z bosons.

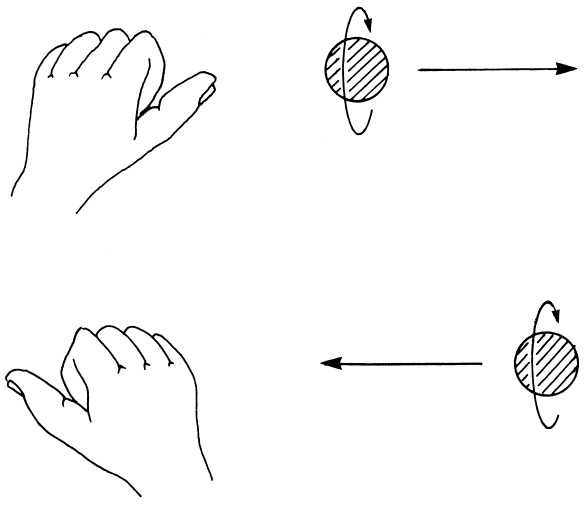


FIGURE 1 A spinning particle is described as either left-handed or right-handed, depending on which hand, when wrapped around the direction of motion with fingers pointing in the direction of the spin, has the thumb pointing in the direction of travel. Thus, the top figure indicates left-handed motion and the bottom figure shows right-handed motion.

Even after the overthrow of parity in 1957, it was believed that a modified remnant of the symmetry remained, that of CP. Here P designates the parity operation, while C signifies charge conjugation, which interchanges particle and antiparticle. Thus, CP invariance requires equality of the rates for:

$$A \rightarrow B + e^- + \bar{\upsilon}_e(right)$$

and

$$\bar{A} \rightarrow \bar{B} + e^+ + \bar{\upsilon}_e(left)$$

Here $\bar{A}$, $\bar{B}$ are the antiparticles of A, B. Such CP invariance occurs naturally in the theories that have been developed to explain beta decay. It would then also be expected to extend to other weak interaction processes, such as the decays of other elementary particles.

There exists also a related symmetry called *time reversal*, T. In this case, the symmetry corresponds to the replacement of the time t by $-t$ in all physical laws (plus the technical addition of using the complex conjugate of the transition amplitude). Pictorially, this consists of taking a film, say, of a scattering amplitude $A + B \rightarrow C + D$ and then running the film backwards to obtain $C + D \rightarrow A + B$ (see Fig. 2). Time reversal invariance requires that the two processes occur with equal probability. In addition, there is a very powerful and fundamental theorem, called the *CPT theorem*, that asserts that in all of the currently known class of field theories the combined action of CP and T transformations must be a symmetry. Of course, this CPT invariance is also being subjected to experimental scrutiny and may in fact be violated in a new class of theories called *string theory* in which the fundamental units are not particles but elementary strings. In most reactions, both CP and T appear to be true symmetries and it is only in exotic reactions that any violation is possibly manifest.

It should be noted that there is an extremely important way in which the world is not CP invariant. This concerns the observed contents of the universe. When we look throughout the visible world, we see mostly electrons but very few positrons and mostly protons but very

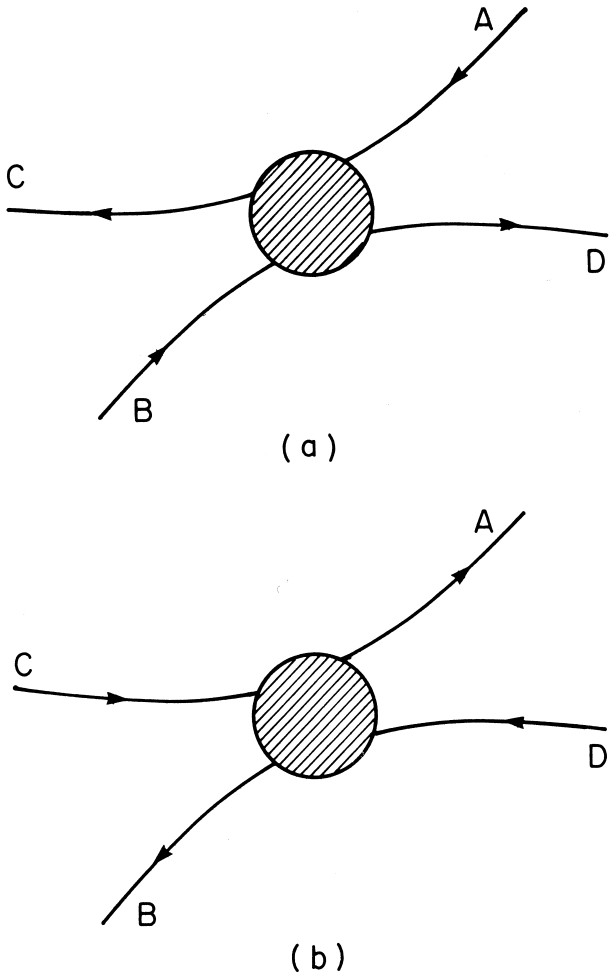


FIGURE 2 According to time-reversal symmetry, the reaction $A + B \rightarrow C + D$ indicated in (a) must be identical to the reaction $C + D \rightarrow A + B$ shown in (b).

few antiprotons. Thus, the matter that exists around us is not CP invariant! This may very well be an indication of CP violation in the early history of the universe. It is natural to assume that at the start of the "big bang" equal numbers of particles and antiparticles were produced. In this case, one requires a mechanism whereby the interactions of nature created a slight preference of particles over antiparticles, such that an excess of particles can remain at the present time. It can be shown that this scenario requires CP-violating interactions. Thus, it can be said that our existence is very likely due to the phenomenon of CP violation.

II. OBSERVATION OF CP VIOLATION

In 1964, a small breaking of CP symmetry was found in a particular weak interaction. In order to understand how this phenomenon was observed, we need to know that under a left–right transformation a particle can have an intrinsic eigenvalue, either $+1$ or -1, since under two successive transformations one must return to the original state. Pi mesons, for example, are spinless particles with a mass of around 140 MeV/c^2. Via study of pion reactions, they are found to transform with a negative sign under parity operations. Pions exist in three different charge states—positive, negative, and neutral—and under charge conjugation, the three pions transform into one another:

$$C|\pi^+\rangle = |\pi^-\rangle$$
$$C|\pi^0\rangle = |\pi^0\rangle$$
$$C|\pi^-\rangle = |\pi^+\rangle$$

(Note that the π^0 is its own antiparticle.) Under CP, then, a neutral pion is negative:

$$CP|\pi^0\rangle = -|\pi^0\rangle$$

so that a state consisting of two or three neutral pions is, respectively, even or odd under CP. The argument is somewhat more subtle for charged pions, but it is found that spinless states $|\pi^+\pi^-\rangle$ and a symmetric combination of $|\pi^+\pi^-\pi^0\rangle$ are also even and odd, respectively, under the CP operation.

Before 1964, it was believed that the world was CP invariant. This had interesting implications for the system of K mesons (spinless particles with a mass of about 498 MeV/c^2) which decay into 2π and 3π by means of the weak interaction. There are two neutral K species, K^0 and $\bar{K}^0$, particle and antiparticle, with identical masses and opposite strangeness quantum numbers. We now understand these particles in the quark model, where K^0 is a bound state of a down quark and a strange antiquark, while $\bar{K}^0$ contains a down antiquark and a strange quark. Under CP, we have:

$$CP|K^0\rangle = |\bar{K}^0\rangle$$
$$CP|\bar{K}^0\rangle = |K^0\rangle$$

Although neither K^0 nor $\bar{K}^0$ is a CP eigenstate, one can form linear combinations:

$$|K_1\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle + |\bar{K}^0\rangle) \qquad CP = +1$$
$$|K_2\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle - |\bar{K}^0\rangle) \qquad CP = -1$$

If the world were CP invariant, then the particle that decays into a two-pion final state must itself be an eigenstate of CP with CP $= +1$, while that which decays into a three-pion final state must have CP $= -1$. Therefore, we expect that the decay scheme is

$$K_1 \to \pi^+\pi^-, \pi^0\pi^0$$
$$K_2 \to \pi^+\pi^-\pi^0, \pi^0\pi^0\pi^0$$

In fact, precisely this phenomenon is observed. The neutral kaons that decay weakly into these pionic channels are different particles (K_L, K_S) with different lifetimes, labeled L, S for long and short:

$$\tau_S \approx 10^{-10}\,\text{sec}$$
$$\tau_L \approx 10^{-8}\,\text{sec}$$

In the limit of CP conservation, $K_S = K_1$ and $K_L = K_2$. One can even observe strangeness oscillations in the time development of the neutral kaon system, which is a fascinating verification of the quantum mechanical superposition principle at work, but it is not our purpose to study this phenomenon here.

Rather, we return to 1964 when an experiment at Brookhaven National Laboratory by Christenson, Cronin, Fitch, and Turlay observed that the same particle—the longer-lived kaon—could decay into both 2π and 3π channels. The effect was not large—for every 300 or so $K_L \to 3\pi$ decays, a single $K_L \to 2\pi$ was detected—but it was definitely present. Since these channels possess opposite CP eigenstates, it was clear that a violation of CP symmetry had been observed.

III. MECHANISMS OF CP VIOLATION

The phenomenon of CP violation is of interest because we do not yet understand its origin. It is possible, but not yet proven, that it could be a manifestation of the Standard Model, which is the current theory of the fundamental interactions that appears to describe most of what we see in particle physics. However, because the breaking of CP symmetry is so small, it is also possible that it is a manifestation of some new type of interaction that is not part of our current Standard Model. In this case, the phenomenon is our initial indication of a deeper theory that will tell us yet more about the workings of nature. The goal of present and future research in this field is to identify the origin of the CP-violating interaction.

The mechanism that allows CP violation within the Standard Model was first articulated by Kobayashi and Maskawa (KM). It makes use of the fact that the interaction of the quarks with the charged bosons that mediate the weak force $W^{\pm}$ have different strengths. Generalizing from work by Cabibbo in the 1960s, the strength of the couplings can be described by angles, and therefore obey a "unitarity" constraint which is a generalization of the relation $\cos^2\theta + \sin^2\theta = 1$. Kobayashi and Maskawa noted that this can be generalized to *complex* angles (i.e., including complex phases) as long as a unitarity condition is satisfied. This has the form:

$$\sum_j V_{ij}^* V_{ji} = 1$$

where V_{ij} are the elements of a 3×3 matrix (the KM matrix); i, j refer to the different types of quarks, and $*$ denotes complex conjugation. It is the addition of complex phases to these couplings that allows for the existence of CP violation. Normally the phase of an amplitude $A = |A|e^{i\phi}$ is not observable, since the decay probability is given by the square of the absolute value of the amplitude $|A|^2$. However, relative phases can sometimes be observed, and kaon mixing can involve the two different amplitudes, $K^0 \to \pi\pi$ and the mixing-induced amplitude $K^0 \to \bar{K}^0 \to \pi\pi$, which can have different relative phases. For example, if the mass eigenstates were K_1 and K_2 described previously, then these two relative phases would be observable:

$$A(K^0 \to \pi\pi) = |A|e^{i\phi}$$
$$A(\bar{K}^0 \to \pi\pi) = |A|e^{-i\phi}$$
$$A(K_1 \to \pi\pi) = \sqrt{2}|A|\cos\phi$$
$$A(K_2 \to \pi\pi) = \sqrt{2}|A|\sin\varphi$$

In the KM scheme, this phase resides in the coupling of light quarks (u, d, and s quarks) to heavy quarks (c, b, and t quarks) and this makes the effect naturally small. The origin of this phase in the heavy quark couplings is not well understood and its magnitude is not predicted, but its existence is compatible with the theory.

There are other theories that have been proposed to explain the phenomenon of CP violation. Indeed, one of these, the superweak model of Wolfenstein, predates the Standard Model. This theory proposes a new, very weak force which can mix K_1 and K_2. In the modern framework of gauge theory, this would involve the exchange of a very heavy particle. Because it is so weak, however, it is very unlikely to be seen in any other effect besides the mixing of the neutral kaons. There are also other mechanisms. For example, in the theory of supersymmetry, which postulates a symmetry between fermions and bosons, there are many complex phases in addition to the one of the KM model, and these can lead to a rich variety of CP-violating processes. Likewise, if there exist extra Higgs bosons, which are spinless particles often postulated in new theories, their couplings will almost always involve CP-violating phases.

IV. RECENT PROGRESS

The most important recent progress involves the observation of an effect that is clearly *not* simply the mixing of

K_1 and K_2. This is then a new effect, often referred to as *direct* CP violation. This emerges from the study of the two different charged states that can emerge from kaon decay. Conventionally, we describe the ratio of two rates by two parameters, ε and ε', defined:

$$\frac{A(K_L \to \pi^+\pi^-)}{A(K_S \to \pi^+\pi^-)} = \varepsilon + \varepsilon'$$

$$\frac{A(K_L \to \pi^0\pi^0)}{A(K_S \to \pi^0\pi^0)} = \varepsilon - 2\varepsilon'$$

If the mixing of the neutral kaons were the only phenomenon contributing to this process, then both these ratios would be identical and $\varepsilon' = 0$. The superweak model leads to this prediction. On the other hand, the KM theory has a mechanism such that the two decays can differ by a small amount. The prediction of this difference is very difficult because of the need to calculate decay amplitudes within the theory of the strong interactions. In fact, the range of predictions in the literature encompasses an order of magnitude $\varepsilon'/\varepsilon = 0.0003 \to 0.003$.

Very beautiful and precise experiments have been carried out over the last decade at CERN (the European Laboratory for Particle Physics) in Geneva and Fermilab (near Chicago) which now agree on a value $\varepsilon'/\varepsilon = 0.0022 \pm 0.0003$. This result is a major advance for the field. It offers convincing proof that direct CP violation exists. Since not all effects are in the mixing mechanism, it rules out the superweak theory. The result also appears compatible with the Standard Model within the present range of theoretical uncertainty. However, further theoretical work is required in order to refine the prediction if this is to become a firm test.

V. FUTURE AREAS OF RESEARCH

Despite a long history of investigation, CP violation has only been detected in the neutral kaon system. The recent observations of ε' have been extremely important but have not decisively identified the mechanism responsible for this phenomenon. Clearly, in order to understand the origin of CP non-conservation, additional experimental observations are required. This is recognized as an important problem in the field and work is underway around the world that may help to clarify this situation.

The most focused effort at present is in the area of B meson decay. B mesons are particles that carry a heavy (b) quark, as well as a lighter (u, d, or s) antiquark. These heavy particles are produced only in high-energy reactions and decay with a lifetime of about 10^{-12} seconds. The neutral particles in this family—called B_d^0 and B_s^0, where the subscript labels the antiquark flavor—undergo mixing with their corresponding antiparticles in a fashion similar to the neutral kaon. This mixing, together with possible direct CP violation in the decay amplitude, leads to a possibility of CP non-conservation in the decay of B mesons. However, because they are heavier there exist far more channels open for B meson decay than are possible for kaons, so the experimental exploration is both richer and more difficult.

There are a few decay channels for which the Standard Model yields precise predictions. The most accessible of these is the reaction, $B_d^0 \to \Psi K_S^0$, where the symbol Ψ denotes the bound state of a charmed quark and a charmed antiquark. The signal being looked for is the difference between the decay to this state, as a function of time, of B_d^0 and its antiparticle $\bar{B}_d^0$. In the ratio of these decay rates, the magnitude of the decay amplitude cancels out, leaving only a well-defined combination of the KM angles, as described above. In addition, this decay mode is experimentally accessible. Both the Ψ and the K_S^0 are readily identified by the particle detectors, and indeed the decay rate relevant for this process has already been measured. Experimentally, the most stringent requirement is the observation of the time dependence of the decay, for which the asymmetric B factories are needed (see below). At the time of this writing, there exist preliminary indications for a CP-violating asymmetry, although the present precision is not sufficient to know if it agrees with the Standard Model prediction. This asymmetry is a valuable test of the Standard Model mechanism and by itself could signal the need for new nonstandard interactions. Moreover, there are many other decay modes that may exhibit CP violation. The overall pattern of such decays will allow a thorough study of the mechanism of CP violation.

The experiments on these heavy particle decays are being carried out at all of the present high-energy accelerator facilities, but most especially at dedicated B factories. These are specialized accelerators that are designed to provide the maximum number of B mesons in an environment that gives experimenters the clearest access to the relevant decay channels. There are three B factories operating in the world as of this writing: at Cornell University, Stanford Linear Accelerator Center, and the KEK laboratory in Japan. The latter two are "asymmetric" machines, where the energies of the two colliding beams are not equal. This design requirement was specifically chosen in order to facilitate the observation of the time dependence of the decay asymmetries. It is expected that these B factories will soon provide preliminary results on CP asymmetries for some of the more accessible modes, to be followed up by a multiyear precision exploration of all facets of heavy quark physics.

A second important area of current and future research is that of measurement of electric dipole moments of

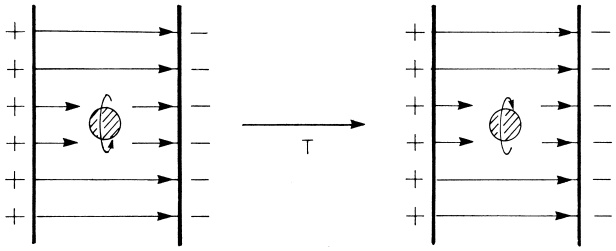


FIGURE 3 Under time reversal, a spinning particle placed between capacitor plates, as shown, will reverse its direction of spin, but the direction of the electric field stays the same. Thus, an interaction of the form $\vec{S} \cdot \vec{E}$ violates time-reversal invariance.

particles. This refers to the interaction between a particle and an applied electric field of the form:

$$H \propto \vec{S} \cdot \vec{E}$$

where $\vec{S}$ is the spin of the particle and $\vec{E}$ is the electric field. Imagine a spinning particle placed in an electric field set up by two oppositely charged capacitor plates (Fig. 3). It is easy to see that if time is reversed, the spin reverses but not the electric field, so that an interaction of this form violates time reversal and, hence, by the CPT theorem also violates CP. There is a long history of experiments that have looked for the possible electric dipole moment of the neutron. The use of a neutral particle is a necessity since a charged particle would accelerate out of the experimental region under the influence of the electric field. At the present time, the experimental upper limit of a possible neutron electric dipole moment is at the level of several times 10^{-26} e-cm. This is an incredible sensitivity. If one imagines a neutron expanded to the size of the earth, the above limit corresponds to a charge separation of only one micron! Similar searches for a nonzero electric dipole moment are being performed with atoms. While no electric dipole moment has yet been found, the interpretation of any such result in terms of a limit is made uncertain by the shielding of the nucleus from the full effect of the electric field because of the shifting of the electron cloud. The Standard Model mechanism for CP violation predicts a dipole moment which is many orders of magnitude too small to be seen by present experiments. However, many other models predict electric dipole moments in the range under investigation, and this may prove to be a powerful indication of new physics.

SEE ALSO THE FOLLOWING ARTICLES

ACCELERATOR PHYSICS AND ENGINEERING • PARTICLE PHYSICS, ELEMENTARY • RADIATION PHYSICS • RADIOACTIVITY

BIBLIOGRAPHY

Bigi, I., and Sanda, A. (2000). "CP Violation," Cambridge University Press, Cambridge, U.K.

Donoghue, J. F., Golowich, E., and Holstein, B. R. (1994). "Dynamics of the Standard Model," Cambridge University Press, Cambridge, U.K.

Georgi, H. (1984). "Weak Interactions and Modern Particle Theory," Benjamin-Cummings, Redwood City, CA.

Gottfried, K., and Weisskopf, V. F. (1984). "Concepts of Particle Physics," Oxford University Press, London.

Halzen, F., and Martin, A. D. (1984). "Quarks and Leptons: An Introductory Course in Modern Particle Physics," Wiley, New York.

Perkins, D. H. (1982). "Introduction to High Energy Physics," 2nd ed., Addison–Wesley, Reading, MA.

Quinn, H., and Witherall, M. (1998). "The asymmetry between matter and antimatter," *Sci. Am.* **Oct.**, 76.

Winstein, B., and Wolfenstein, L. (1993). "The search for direct CP violation," *Rev. Mod. Phys.* **65,** 1113.



Dense Matter Physics

George Y. C. Leung
Southeastern Massachusetts University

I. Background and Scope
II. Basic Theoretical Method
III. Composition of Dense Matter
IV. Equation of State of Dense Matter
V. Transport Properties of Dense Matter
VI. Neutrino Emissivity and Opacity

GLOSSARY

Adiabat Equation of state of matter that relates the pressure to the density of the system under a constant entropy.

Baryons Elementary particles belonging to a type of fermions that includes the nucleons, hyperons, delta particles, and others. Each baryon is associated with a baryon number of one, which is a quantity conserved in all nuclear reactions.

Bosons Elementary particles are divided into two classes called *bosons* and *fermions.* The bosons include the photons, phonons, and mesons. At thermal equilibrium, the energy distribution of identical bosons follows the Bose–Einstein distribution.

Degenerate electrons System of electrons that occupy the lowest allowable momentum states of the system, thus constituting the absolute ground state of such a system.

Fermions Class of elementary particles that includes the electrons, neutrions, nucleons, and other baryons. Identical fermions obey Pauli's exclusion principle and follow the Fermi–Dirac distribution at thermal equilibrium.

Isotherm Equation of state of matter that relates the pressure to the density of the system at constant temperature.

Neutrinos Neutral, massless fermions that interact with matter through the weak interaction. Neutrinos are produced, for example, in the decay of the neutrons.

Neutronization Form of nuclear reaction in which the neutron content of the reaction product is always higher than that of the reaction ingredient. It occurs in dense matter as its density increases from 10^7 to 10^{12} g/cm^3.

Nuclear matter Matter substance forming the interior of a nucleus. Its density is approximately 2.8×10^{12} g/cm^3, which is relatively independent of the nuclear species. It is composed of nearly half neutrons and half protons.

Phonons Lattice vibrations of a solid may be decomposed into a set of vibrational modes of definite frequencies. Each frequency mode is composed of an integral number of quanta of definite energy and momentum.

These quanta are called *phonons.* They are classified as bosons.

Photons Particle–wave duality is an important concept of quantum theory. In quantum theory, electromagnetic radiation may be treated as a system of photons endowed with particle properties such as energy and momentum. A photon is a massless boson of unit spin.

Quarks Subparticle units that form the elementary particles. There are several species of quarks, each of which possesses, in addition to mass and electric charge, other fundamental attributes such as c-charge (color) and f-charge (flavor).

Superconductivity Electrical resistance of a superconductor disappears completely when it is cooled below the critical temperature. The phenomenon is explained by the fact that due to the presence of an energy gap in the charge carriers' (electrons or protons) energy spectrum, the carriers cannot be scattered very easily, and the absence of scattering leads to superconductivity.

Superfluidity Superfluidity is the complete absence of viscosity. The conditions leading to superconductivity also lead to superfluidity in the proton or electron components of the substance. In the case of neutron matter, the neutron component may turn superfluid due to the absence of scattering. The critical temperatures for the proton and neutron components in neutron matter need not be the same.

DENSE MATTER PHYSICS is the study of the physical properties of material substance compressed to high density. The density range begins with hundreds of grams per cubic centimeter and extends to values 10 to 15 orders of magnitudes higher. Although such dense matter does not occur terrestrially, it exists inside stellar objects such as the white dwarf stars, neutron stars, and black holes and possibly existed during the early phase of the universe. Dense matter physics therefore provides the scientific basis for the investigation of these objects.

I. BACKGROUND AND SCOPE

Matter is the substance of which all physical objects are composed. The density of matter is the ratio of its mass to volume and is a measure of the composition of matter and the compactness of the constituent entities in it. In units of grams per cubic centimeter (g/cm^3) the density of water is 1.0 g/cm^3, and the densities of all macroscopic objects on earth do not exceed roughly 20 g/cm^3. However, some stellar objects are believed to be formed of matter with much higher densities. In the 1920s, a star called Sirius B, a binary companion of the star Sirius, was found to be a highly compact object having the mass of our sun but the size of a planet, and thus must be composed of matter of very high density, estimated to reach millions of g/cm^3. Sirius B is now known to belong to a class of stellar objects called the *white dwarf stars.* Dense matter physics began as an effort to understand the structure of the white dwarf stars. It matured into a branch of science devoted to the study of the physical properties of dense matter of all types that may be of interest to astrophysical and cosmological investigations.

Since the type of matter under study cannot be found terrestrially, it is impossible to subject it to direct laboratory examination. Hence, the study of dense matter physics is mainly theoretical in nature. In the 1920s the emergence of quantum mechanics was making a strong impact on physics, and a theory of dense matter based on the quantum mechanical behavior of electrons at high density was constructed. It marked the dawn of dense matter physics, and this theory remains valid today for the study of white dwarf stars. The subsequent identification of other compact stellar objects such as the neutron stars and black holes greatly intensified the study of dense matter physics. We survey here what can be expected theoretically from dense matter and the implications of current theories on the structure of these compact stellar objects.

On the experimental side, the study is benefited by the fact that if the concept of matter density is extended to include microscopic bodies such as the atomic nuclei, then a substance called *nuclear matter*, which possesses extremely high density, may be identified. Through nuclear physics study, it is then possible to subject matter with such high density to laboratory examinations. Such experimental information provides an invaluable guide to the study of matter forming the neutron stars.

Compact stellar objects are mainly the remains of stars whose nuclear fuels have been exhausted and are drained of the nuclear energy needed to resist the pull of the gravitational force. As the gravitational force contracts the stellar body, it also grows in strength. This unstable situation is described as gravitational collapse, which continues until a new source of reaction strong enough to oppose the gravitational force becomes available. The search for the physical properties of dense matter responsible for resistances to gravitational collapse is an important aspect in dense matter physics since the results have important astrophysical implications.

The structure and stability of a compact stellar object depend on its composition and the equation of state of the form of matter that it is composed of. The equation of state expresses the pressure generated by the matter substance as a function of its density and temperature. The determination of the composition and the equation of state of dense matter is a prime objective in dense matter

studies. These topics are discussed in Sections III and IV after a brief introduction of the basic theoretical method involved is presented in Section II.

Compact stellar objects perform rotations, pulsations, and emissions, and to understand these processes we would need to know, in addition to the equation of state, the properties of dense matter under nonequilibrium conditions. These are called the *transport properties*, which include the electrical and thermal conductivity and viscosity. These intrinsic properties of dense matter are discussed in Section V. The effects of a strong magnetic field on the transport properties, however, are not included in this writing. Properties related to radiative transfer, such as emissivity and opacity, are discussed in Section VI. However, radiative transfer by photons in dense matter is completely superceded by conductive transfer, and since it does not play an important role, the photon emissivity and opacity in dense matter will not be discussed. Instead, Section VI concentrates on the much more interesting topic of neutrino emissivity and opacity in dense matter.

The properties of dense matter will be discussed in several separate density domains, each of which is characterized by typical physical properties. In the first density domain, from 10^2 to 10^7 gm/cm^3, the physical properties are determined to a large extent by the electrons among the constituent atoms. The electrons obey an important quantum mechanical principle called *Pauli's exclusion principle* which forbids two electrons to occupy the same quantum state in a system. All electrons must take up quantum states that differ in energy, and as the electron density is increased, more and more of the electrons are forced to take on new high-energy quantum states. Consequently, the total energy of the electrons represents by far the largest share of energy in the matter system. It is also responsible for the generation of an internal pressure in the system. All white dwarf stars are believed to be composed of matter with densities falling in this domain, which is known to sustain stable stellar configurations. The physical mechanism mentioned here for electrons is central to establishing the physical properties of dense matter at all density domains, and for this reason it is first introduced in Section II.

The second density domain ranges from 10^7 to 10^{12} g/cm^3, where nuclear physics plays a key role. Above 10^7 g/cm^3 the constituent atomic nuclei of the dense matter experiences nuclear transmutations. In general, an increase in density above this point leads to the appearance of nuclei that are richer in neutron content than those occurring before. This process, called *neutronization*, continues throughout the entire density domain. The process also suppresses the increase in electron number with increase in matter density and thus deprives the matter system of its major source of energy. Matter with densities belonging to this density domain experiences a gradual reduction in compressibility with increasing density and is no longer able to sustain stable stellar configurations after its density exceeds 10^8 g/cm^3.

As matter density approaches 10^{12} g/cm^3, some nuclei become so rich in neutrons that they cease to bind the excess neutrons; nuclei now appear to be immersed in a sea of neutrons. The onset of such a phenomenon is called *neutron drip*, a term suggesting that neutrons are dripping out of the nuclei. This leads to the third density domain ranging from 10^{12} to 10^{15} g/cm^3. A rapid increase in neutron density accompanying an increasing matter density leads to the production of energetic neutrons, since neutrons (like electrons) obey Pauli's exclusion principle. Hence, the same quantum mechanical mechanism characterizing the first density domain becomes operative here. As soon as neutrons were discovered experimentally in the 1930s, this mechanism was invoked to suggest the possible existence of stable neutron stars, long before neutron stars were actually identified in astronomical observations. Unlike the electrons, however, neutrons interact among themselves with nuclear forces that are comparatively strong and must be handled with great care. The average density of atomic nuclei, or nuclear matter density, is of the order of 10^{14} g/cm^3. Much of the needed physics in understanding matter with density in this density domain must come from nuclear physics.

Our understanding of matter with densities above 10^{15} g/cm^3 is very tentative; for this reason we shall, for discussion, assign matter with densities above 10^{15} g/cm^3 into the fourth and last density domain. In this area we shall discuss the physical basis for topics such as hyperonic matter, pion condensation, and quark matter.

Since the study of dense matter is highly theoretical in nature, we begin our discussion with an introduction to the basic theoretical method needed for such an investigation in establishing the composition of dense matter and its equation of state.

II. BASIC THEORETICAL METHOD

Matter may first be considered as a homogeneous system of atoms without any particular structure. Such a system may be a finite portion in an infinite body of such substance, so that boundary effects on the system are minimized. The system in the chosen volume possesses a fixed number of atoms. At densities of interest all atoms in it are crushed, and the substance in the system is best described as a plasma of atomic nuclei and electrons. We begin by studying pure substances, each formed by nuclei belonging to a single nuclear species, or nuclide. The admixture of other nuclides in a pure substance can be accounted

for and will be considered after the general method of investigation is introduced.

The physical properties of the system do not depend on the size of the volume, since it is chosen arbitrarily, but depend on parameters such as density, which is obtained by dividing the total mass of the system by its volume. The concept of density will be enlarged by introducing a host of densitylike parameters, all of which are obtained by dividing the total quantities of these items in the system by its volume. The term "density" will be qualified by calling it *mass density* whenever necessary. In addition, parameters such as the electron number density and the nuclei number density will be introduced.

Each nuclide is specified by its atomic number Z and mass number A. Z is also the number of protons in the nucleus and A the total number of protons and neutrons there. Protons and neutrons have very nearly the same mass and also very similar nuclear properties; they are often referred to collectively as *nucleons*. Each nuclide will be designated by placing its mass number as a left-hand superscript to its chemical symbol, such as ^{4}He and ^{56}Fe.

Let the system consist of N atoms in a volume V. The nuclei number density is

$$n_{\mathrm{A}} = N/V \tag{1}$$

and the electron number density is

$$n_{\mathrm{e}} = NZ/V \tag{2}$$

which is the same as the proton number density, because the system is electrically neutral. The mass of a nuclide is given in nuclear mass tables to high accuracy. To determine the mass density of a matter system we usually do not need to know the nuclear mass to such accuracy and may simply approximate it by the quantity Am_{p}, where $m_{\mathrm{p}} = 1.67 \times 10^{-24}$ g is the proton mass. The actual mass of a nucleus should be slightly less than that because some of the nuclear mass, less than 1%, is converted into the binding energy of the nucleus. The mass density ρ of the matter system is given simply by:

$$\rho = NAm_{\mathrm{p}}/V \tag{3}$$

For example, for a system composed of electrons and helium nuclei, for which $Z = 2$ and $A = 4$, a mass density of 100 g/cm^3 corresponds to a nuclei number density of $n_{\mathrm{A}} = 4 \times 10^{24}$ cm^{-3} and an electron number density $n_{\mathrm{e}} = 2n_{\mathrm{A}}$.

Like all material substances, dense matter possesses an internal pressure that resists compression, and there is a definite relation between the density of a substance and the pressure it generates. The functional relationship between pressure and density is the equation of state of the substance and is a very important physical property for astrophysical study of stellar objects. The structure of a stellar object is determined mainly by the equation of state of the stellar substance. In this section, the method for the derivation of the equation of state will be illustrated.

Unlike a body of low-density gas, whose pressure is due to the thermal motions of its constituent atoms and is therefore directly related to the temperature of the gas, dense matter derives its pressure from the electrons in the system. When matter density is high and its electron number density is correspondingly high, an important physical phenomenon is brought into play which determines many physical properties of the system. We shall illustrate the application of this phenomenon to the study of dense matter with densities lying in the first density domain, 10^2–10^7 g/cm^3.

A. Pauli's Exclusion Principle

It is a fundamental physical principle that all elementary particles, such as electrons, protons, neutrons, photons, and mesons, can be classified into two major categories called *fermions* and *bosons*. The fermions include electrons, protons, and neutrons, while the bosons include photons and mesons. Atomic nuclei are regarded as composite systems and need not fall into these classes. In this study, one fermionic property, Pauli's exclusion principle, plays a particularly important role. The principle states that *no two identical fermions should occupy the same quantum state in a system*. Let us first explain the meaning of identical fermions. Two fermions are identical if they are of the same type and have the same spin orientation. The first requirement is clear, but the second deserves further elaboration. A fermion possesses intrinsic angular momentum called *spin* with magnitude equal to $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, ... basic units of the angular momentum (each basic unit equals Planck's constant divided by 2π). The electron spin is equal to $\frac{1}{2}$ of the basic unit, and electrons are also referred to as *spin-half particles*. Spin is a vector quantity and is associated with a direction. In the case of electron spin, its direction may be specified by declaring whether it is oriented parallel to or antiparallel to a chosen direction. The fact that there are no other orientations besides these two is a result of nature's quantal manifestation, the recognition of which laid the foundation of modern atomic physics. These two spin orientations are simply referred to as *spin-up* and *spin-down*. All spin-up electrons in a system are identical to each other, as are all spin-down electrons. The term identical electrons is thus defined. Normally, these two types of electrons are evenly distributed, since the system does not prefer one type of orientation over the other.

We come now to the meaning of a quantum state. Each electron in a dynamical system is assigned a quantum state. The quantum state occupied by an electron may be specified by the electron momentum, denoted by **p**, a vector

that is expressable in Cartesian coordinates as p_x, p_y, p_z. When an electron's momentum is changed to $\mathbf{p}'$ in a dynamical process, it moves to a new quantum state. The momentum associated with the new quantum state $\mathbf{p}'$ must, however, differ from $\mathbf{p}$ by a finite amount. For a system of uniformly distributed electrons in a volume V, this amount may be specified in the following way. If each of the three momentum components is compared, such as p_x and p'_x, they must differ by an amount equal to h/L, where h is Planck's constant and $L = V^{1/3}$. In other words, in the space of all possible momenta for the electrons, each quantum state occupies a size of h^3/V, which is the product of the momentum differences in all three components.

B. Fermi Gas Method

According to Pauli's exclusion principle, all electrons of the same spin orientation in a system must possess momenta that differ from each other by the amount specified above. The momenta of all electrons are completely fixed if it is further required that the system exist in its ground state, which is the lowest possible energy state for the system because there is just one way that this can be accomplished, which is for the electrons to occupy all the low-momentum quantum states before they occupy any of the high-momentum quantum states. In other words, there exists a fixed momentum magnitude p_{F}, and, in the ground-state configuration, all states with momentum magnitudes below p_{F} are occupied while the others with momentum magnitudes above p_{F} are unoccupied. Graphically, the momenta of the occupied states plotted in a three-dimensional momentum space appear like a solid sphere centered about the point of zero momentum. The summation of all occupied quantum states is equivalent to finding the volume of the sphere, since all states are equally spaced from each other inside the sphere. Let the radius of the sphere be given by p_{F}, called the *Fermi momentum*, then the integral for the volume of the sphere with the momentum variable expressed in the spherical polar coordinates is given by:

$$\int_0^{p_{\mathrm{F}}} 4\pi p^2\, dp = \left(\frac{4\pi}{3}\right) p_{\mathrm{F}}^3 \tag{4}$$

where p is the magnitude of $\mathbf{p}$, $p = |\mathbf{p}|$. The number of electrons accommodated in this situation is obtained by dividing the above momentum volume by h^3/V and then multiplying by 2, which is to account for electrons with two different spin orientations. The result is proportional to the volume V which appears because the total number of electrons in the system is sought. The volume V is divided out if the electron number density is evaluated, which is given by:

$$n_{\mathrm{e}} = (8\pi/3h^3) p_{\mathrm{F}}^3 \tag{5}$$

Equation (5) may be viewed as a relation connecting the electron number density n_{e} to the Fermi momentum p_{F}. Henceforth, p_{F} will be employed as an independent variable in establishing the physical properties of the system.

The total kinetic energy density (energy per unit volume) of the electrons can be found by summing the kinetic energies of all occupied states and then dividing by the volume. Each state of momentum $\mathbf{p}$ possesses a kinetic energy of $p^2/2m_{\mathrm{e}}$, where m_{e} is the electron mass, and the expression for the electron kinetic energy density is

$$\varepsilon_{\mathrm{e}} = \frac{2}{h^3}\int_0^{p_{\mathrm{F}}} 4\pi p^2\, dp \frac{p^2}{2m_{\mathrm{e}}} = \frac{2\pi}{5h^3 m_{\mathrm{e}}} p_{\mathrm{F}}^5 \tag{6}$$

The average kinetic energy per electron is obtained by dividing ε_e by n_e:

$$\varepsilon_{\mathrm{e}}/n_{\mathrm{e}} = 0.6\left(p_{\mathrm{F}}^2/2m_{\mathrm{e}}\right) \tag{7}$$

where $p_{\mathrm{F}}^2/2m_{\mathrm{e}}$ is the kinetic energy associated with the most energetic electrons in the system, and here we see that the average kinetic energy of all electrons in the system is just six-tenths of it. This is an important result to bear in mind. It tells us that p_{F} is a representative parameter for the system and may be used for order of magnitude estimates of many of the energy-related quantities. The method described here for finding the energy of the system of electrons may be applied to other fermions such as protons and neutrons and is referred to generally as the *Fermi gas method.*

It is now instructive to see what would be the average electron kinetic energy in a matter system that is composed of ^{4}He nuclei at a density of 100 g/cm^3. From the electron number density that we have computed before, $n_e = 8 \times 10^{24}$ cm^{-3}, we find that

$$\frac{\varepsilon_{\mathrm{e}}}{n_{\mathrm{e}}} = \frac{0.3h^2c^2}{m_{\mathrm{e}}c^2}\left(\frac{3}{4\pi}n_{\mathrm{e}}\right)^{2/3} \approx 300\ \mathrm{eV} \tag{8}$$

where eV stands for electron volts. In evaluating expressions such as Eq. (8), we shall follow a scheme that reduces all units needed in the problem to two by inserting factors of h or c at appropriate places. These two units are picked to be electron volts (eV) for energy and centimeters (cm) for length. Thus, instead of expressing the electron mass in grams it is converted into energy units by multiplying by a factor of c^2, and $m_{\mathrm{e}}c^2 = 0.511 \times 10^6$ eV. The units for h can also be simplified if they are combined with a factor of c, since $hc = 1.24 \times 10^{-4}$ eV cm. This scheme is employed in the evaluation of Eq. (8) as the insertions of the c factors are explicitly displayed.

The average kinetic energy per electron given by Eq. (8) is already quite formidable, and it increases at a rate proportional to the two-thirds power of the matter density. For

comparison, we estimate the average thermal energy per particle in a system, which may be approximated by the expression $k_B T$, where k_B is Boltzmann's constant and T the temperature of the system in degrees Kelvin (K). The average thermal energy per particle at a temperature of 10^6 K is about 100 eV. Consequently, in the study of compact stellar objects whose core densities may reach millions of g/cm^3 while the temperature is only 10^7 K, the thermal energy of the particles may justifiably be ignored.

Electrons belonging to the ground state of a system are called *degenerate electrons*, a term that signifies the fact that all electron states whose momenta have magnitudes below the Fermi momentum p_F are totally occupied. Thus, whenever electrons are excited dynamically or thermally to states with momenta above p_F, leaving states with momenta below p_F empty, such a distribution of occupied states is called *partially degenerate*. Partially degenerate electrons will be discussed later in connection with systems at high temperatures.

C. Pressure

The internal pressure P of a system of particles is given by the thermodynamic expression,

$$P = -\left.\frac{dE}{dV}\right|_N \tag{9}$$

where E is the total energy of the system and N its particle number in the volume V. For example, in the case of the electron gas, $E = \varepsilon_e V$ and $N = n_e V$. The evaluation of the derivative in Eq. (9) is best carried out by using p_F as an independent variable and converting the derivative into a ratio of the partial derivatives of E and V with respect to p_F, while keeping N fixed. For a noninteracting degenerate electron system, the pressure can also be derived from simple kinematical considerations, and the result is expressed as:

$$P_e = \frac{2}{h^3}\int_0^{p_F} 4\pi p^2\, dp \frac{pv}{3} \tag{10}$$

where v is the velocity of the electron, $v = p/m_e$. By either method, the electron pressure is evaluated to be

$$P_e = \frac{8\pi}{15 m_e h^3} p_F^5 \tag{11}$$

Pressure due to a degenerate Fermi system is called degenerate pressure.

Pressure generated by the nuclei may be added directly to the electron pressure, treating both as partial pressures. They contribute additively to the total pressure of the system. Since nuclei are not fermions (with the sole exception of the hydrogen nuclei, which are protons), they do not possess degenerate pressure but only thermal pressure, which is nonexistent at zero temperature. In the case of the hydrogen nuclei, the degenerate pressure they generate may also be neglected when compared with the electron pressure, since the degenerate pressure is inversely proportional to the particle mass; the proton mass being 2000 times larger than the electron mass makes the proton pressure 2000 times smaller than the elctron pressure (bearing in mind also that the proton number density is the same as the electron number density). Thus, the pressure from the system P is due entirely to the electron pressure:

$$P = P_e \tag{12}$$

D. Relativistic Electrons

Because the electrons have very small mass, they turn relativistic at fairly low kinetic energies. Relativistic kinematics must be employed for the electrons when the kinetic energy approaches the electron rest energy $m_e c^2$. For a noninteracting degenerate electron system, its most energetic electrons should reach this energy threshold when the electron number density is $n_e = 2 \times 10^{30}$ cm^{-3}, which translates into a helium matter density of $\rho = 3 \times 10^7$ g/cm^3.

In the relativistic formalism, the evaluation of the electron number density of the system remains unchanged since it depends only on p_F. The evaluation of the electron energy density must, however, be modified by replacing the individual electron kinetic energy from the expression $p^2/2m_e$ to the relativistic expression,

$$e_k = \left(p^2c^3 + m_e^2 c^4\right)^{1/2} - m_e c^2 \tag{13}$$

so that

$$\varepsilon_e = \frac{2}{h^3}\int_0^{p_F} 4\pi p^2\, dp\ e_k \tag{14}$$

The electron pressure may also be found from Eq. (9) without modification, but if Eq. (10) is used, the electron velocity must be modified to the relativistic form, which is

$$v = \frac{p}{\left(p^2c^2 + m_e^2 c^4\right)^{1/2}} \tag{15}$$

In summary, the physical quantities of a noninteracting degenerate electron system are

$$n_e = \frac{8\pi}{3\lambda_e^3} t^3 \tag{16a}$$

$$\varepsilon_e = \frac{\pi m_e c^2}{\lambda_e^3}\{t(2t^2+1)(t^2+1)^{1/2} - \ln[t + (t^2+1)^{1/2}]\} \tag{16b}$$

$$P_e = \frac{\pi m_e c^2}{\lambda_e^3}\left\{\frac{t}{3}(2t^2-3)(t^2+1)^{1/2} + \ln[t + (t^2+1)^{1/2}]\right\} \tag{16c}$$

where $\lambda_e = h/m_e c$ is the electron Compton wavelength having the dimension of a length and $t = p_F/m_e c$, which is dimensionless.

E. Equation of State

The mass density of dense matter given by Eq. (3) may be expressed in terms of the electron number density as:

$$\rho = m_p n_e / Y_e \tag{17}$$

where Y_e, called the *electron fraction*, is the ratio of the number of electrons to nucleons in the system. For a pure substance, it is just $Y_e = Z/A$, but for a mixed substance, where a variety of nuclides are present, it is given by:

$$Y_e = \sum_i x_i \left(\frac{Z_i}{A_i}\right) \tag{18}$$

where the subscript i designates the nuclide type and x_i denotes the relative abundance of that nuclide in the system. For example, $Y_e = 1.0$ for a pure hydrogen system, $Y_e = 0.5$ for a pure ^{4}He system, and $Y_e = \frac{26}{56} = 0.464$ for a pure ^{56}Fe system. All pure substances that are composed of nuclides lying between He and Fe in the periodic table have electron fractions bounded by the values 0.464 and 0.5. Since pressure depends on n_e of a system while its mass density depends on n_e/Y_e, it is clear that matter systems with the same Y_e would have the same equation of state.

In Fig. 1, typical equations of state of dense matter at zero temperature are plotted for the cases of $Y_e = 1.0$ and $Y_c = 0.5$. The pressure is due entirely to that of a non-interacting degenerate electron system and is computed from Eq. (16c). Pressure is expressed in units of dynes per square centimeter (dyn/cm^2). From the approximate linearity of these curves on a log–log plot, it is apparent that the pressure and density are related very nearly by a power law. It is therefore both convenient and useful to express the equation of state in the form:

$$P \propto \rho^{\Gamma} \tag{19}$$

where

$$\Gamma = \frac{d(\ln P)}{d(\ln \rho)} \tag{20}$$

is called the *adiabatic index* of the equation of state. Γ should be approximately constant in the density range belonging to the first density domain. We find $\Gamma = \frac{5}{3}$ at the low-density end of the density domain, and it decreases slowly to values approaching $\Gamma = \frac{4}{3}$ at the high-density end of the regime. The magnitude of Γ is a measure of the "stiffness" of the equation of state, in the language of astrophysics. A high Γ means the equation of state is stiff (the substance is hard to compress). Such information

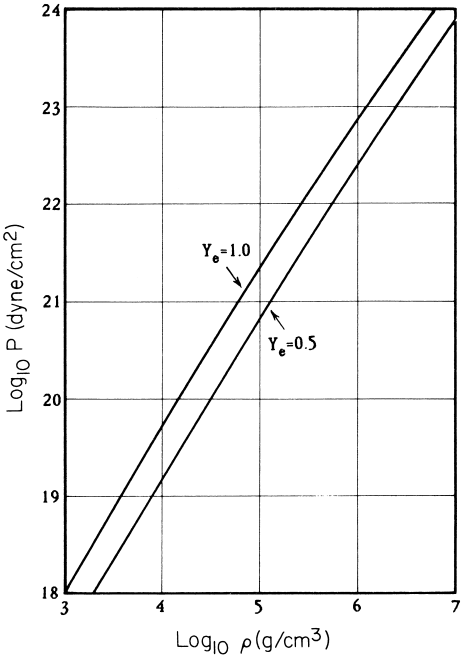


FIGURE 1 Two equations of state at zero temperature, one for matter composed of helium nuclei ($Y_e = 0.5$) and the other for matter composed of hydrogen nuclei ($Y_e = 1.0$), computed by the Fermi gas method.

is of course of utmost importance to the study of stellar structure.

The numerical values of the equation of state of dense matter composed of helium nuclei at zero temperature are listed in Table I. In spite of the fact that this equation of state is obtained by neglecting electrostatic interaction, it should still be applicable to the study of white dwarf stars which are expected to be composed predominantly

TABLE I Equation of State of Matter Composed of Helium Nuclei at Zero Temperature

t	ρ (g/cm^3)	P (dyn/cm^2)	$\text{Log}_{10}\,\rho$	$\text{Log}_{10}\,P$	Γ
0.037	1.00×10^2	6.70×10^{15}	2.0	15.826	$\frac{5}{3}$
0.054	3.16×10^2	4.58×10^{16}	2.5	16.661	$\frac{5}{3}$
0.080	1.00×10^3	3.12×10^{17}	3.0	17.494	$\frac{5}{3}$
0.117	3.16×10^3	2.11×10^{18}	3.5	18.324	1.66
0.172	1.00×10^4	1.43×10^{19}	4.0	19.156	1.66
0.252	3.16×10^4	9.63×10^{19}	4.5	19.984	1.65
0.371	1.00×10^5	6.41×10^{20}	5.0	20.807	1.64
0.544	3.16×10^5	4.16×10^{21}	5.5	21.619	1.61
0.798	1.00×10^6	2.59×10^{22}	6.0	22.414	1.57
1.172	3.16×10^6	1.52×10^{23}	6.5	23.182	1.51
1.720	1.00×10^7	8.36×10^{23}	7.0	23.922	1.45
2.524	3.16×10^7	4.31×10^{24}	7.5	24.634	1.40
3.706	1.00×10^8	2.13×10^{25}	8.0	25.328	1.37

of dense helium matter. As we shall see below, the electrostatic interaction is not going to be important at high densities. The exact composition of the star is also not important in establishing a representative equation of state for the substance forming the white dwarf star, since all nuclides from He to Fe yield very similar Y_e factors; therefore, any mixture of these nuclides would give very nearly the same equation of state. The maximum core density of the largest white dwarf star should not exceed 10^7 g/cm^3.

F. Electrostatic Interaction

In a plasma of electrons and nuclei, the most important form of interaction is the electrostatic interaction among the charged particles, which could modify the total energy of the system. Several estimates of the electrostatic interaction energy will be given here beginning with the crudest.

Each electron in the plasma experiences attraction from the surrounding nuclei and repulsion by the other electrons. Its electrostatic interaction does not depend on the overall size of the system chosen for the investigation, since it is electrically neutral and should depend only on the local distribution of the electric charges. The positive charges carried by the protons are packed together in units of Z in the nucleus, while the negative charges carried by the electrons are distributed in the surrounding space. To investigate the electrostatic energy due to such a charge distribution, let us imagine that the system can be isolated into units, each occupied by a single nucleus of charge Z and its accompanying Z electrons. The volume of the unit is given by Z/n_e, since $1/n_e$ is the volume occupied by each electron and there are Z electrons in the unit. The shape of the unit may be approximated by a spherical cell with the nucleus residing at its center. In each spherical cell, the electrons are distributed in a spherically symmetric way about the center. In other words, if a set of spherical polar coordinates were introduced to describe this cell, with the coordinate origin at the nucleus, the electron distribution can only be a function of the radial variable r and not of the polar angle variables. With such a charge distribution, the cell appears electrically neutral to charges outside the cell, and all electrostatic energy of such a unit must be due to interactions within the cell. The radius of the cell r_s is given by the relation $(4\pi/3)r_s^3 = Z/n_e$. A more accurate treatment would call for the replacement of the spherical cell by polygonal cells of definite shapes by assuming that the nuclei in the system are organized into a lattice structure, but the energy corrections due to such considerations belong to higher orders and may be ignored for the time being.

The determination of the radial distribution of the electrons within the cell is complicated by the fact that it cannot be established by classical electrostatic methods, because in this case with cell size in atomic dimensions, Pauli's exclusion principle plays a major role. As a first estimate we may assume the electrons are distributed uniformly within the cell. The interaction energy due to electron–nucleus interaction is

$$E_{N-e} = -3/2(Ze)^2/r_s \tag{21}$$

and that due to electron–electron interaction is

$$E_{e-e} = 3/5(Ze)^2/r_s \tag{22}$$

giving a total electrostatic energy per cell of

$$E_s = E_{N-e} + E_{e-e} = -9/10(Ze)^2/r_s \tag{23}$$

In these expressions, the electric charge is expressed in the Gaussian (c.g.s.) units, so the atomic fine-structure constant is given by $\alpha = 2\pi e^2/hc = (137.04)^{-1}$. Dividing Eq. (23) by Z gives us the average electrostatic energy per electron, which, if expressed as a function of the electron number density of the system, is

$$\varepsilon_s = E_s/Z = -(9e^2/10)(4\pi Z^2 n_e/3)^{1/3} \tag{24}$$

Note that the interaction energy is negative and thus tends to reduce the electron pressure in the system. Quantitatively, for an electron number density of $n_e = 8 \times 10^{24}$ cm^{-3} and a corresponding helium matter density of 100 g/cm^3, we find $\varepsilon_s = 65$ eV, which is about 20% of the average electron kinetic energy. At higher densities, the relative importance of the electrostatic interaction energy to the total energy of the system is actually reduced.

G. Thomas–Fermi Method

The assumption of uniform electron distribution in a cell, though expedient, is certainly unjustified, since the electrons are attracted to the nucleus, thus the electron distribution in the neighborhood of the nucleus should be higher. The exact determination of the electron distribution in a cell is a quantum mechanical problem of high degree of difficulty. The best-known results on this problem are derived from the Thomas–Fermi method, which is an approximate method for the solution of the underlying quantum mechanical problem. The method assumes the following:

1. It is meaningful to introduce a radially dependent Fermi momentum $p_F(r)$ which determines the electron number density, as in Eq. (5), so that

$$n_e(r) = 8\pi/3h^3[p_F(r)]^3 \tag{25}$$

2. The maximum electron kinetic energy, given by $p_F^2/2m_e$, now varies radially from point to point, but, for

a system in equilibrium, electrons at all points must reach the same maximum total energy, kinetic and potential, thus

$$[p_F(r)]^2/2m_e - e\,\Phi\,(r) = \mu \tag{26}$$

where $\Phi(r)$ is the electrostatic potential energy, and μ the chemical potential, a constant independent of r (μ replaces p_F in establishing the electron number density in the cell).

3. The electrostatic potential obeys Gauss' law:

$$\nabla^2\Phi = -4\pi[Ze\,\delta(r) - en_e(r)] \tag{27}$$

where $\delta(r)$ is the mathematical delta function. From these three relations the electron distribution may be solved. Since the first relation is based on Eq. (5), which is derived from a system of uniformly distributed electron gas, its accuracy depends on the actual degree of variation in electron distribution in the cell. The second relation can be improved by adding to the electrostatic potential a term called the *exchange energy*, which is a quantum mechanical result. The improved method which includes the exchange energy is called the Thomas–Fermi–Dirac method.

The solution of these three relations for $n_e(r)$ is quite laborious, but it has been done for most of the stable nuclides. After it is found, the electron pressure in the system can be evaluated by employing Eq. (16c), setting p_F in the expression equal to $p_F(r_s)$, which is the Fermi momentum of the electrons at the cell boundary. There is no net force acting at the cell boundary, and the pressure there is due entirely to the kinetic energy of the electrons. Thus, the electron pressure can be evaluated by Eq. (16c). Note that this condition would not be fulfilled by an interior point in the cell. Denoting the degenerate electron pressure evaluated for noninteracting electrons by p_e, the pressure P of the matter system corrected for electrostatic interaction by means of the Thomas–Fermi–Dirac method may be expressed as:

$$P = F(\xi)P_e \tag{28}$$

where ξ is a dimensionless parameter related to r_s and Z:

$$\xi = (0.62)Z^{-1/3}(a_0/r_s) \tag{29}$$

where $a_0 = (h/2\pi)^2/e^2m_e = 5.292 \times 10^{-9}$ cm is the Bohr radius and $F(\xi)$ a rather complicated function that is displayed graphically in Fig. 2. The dashed lines represent the correction factor derived from the Thomas–Fermi method, and it holds for all nuclides. The correction factor derived from the Thomas–Fermi–Dirac method depends on Z, and curves for two extreme cases are shown. Curves for nuclides of intermediate Z should fall between these two curves. The correction curve based on the assumption of uniform electron distribution is shown by the dotted curve for comparison. It is computed for the case of Fe matter. All curves tend to converge for large values of ξ. The relation between matter density and ξ is given by:

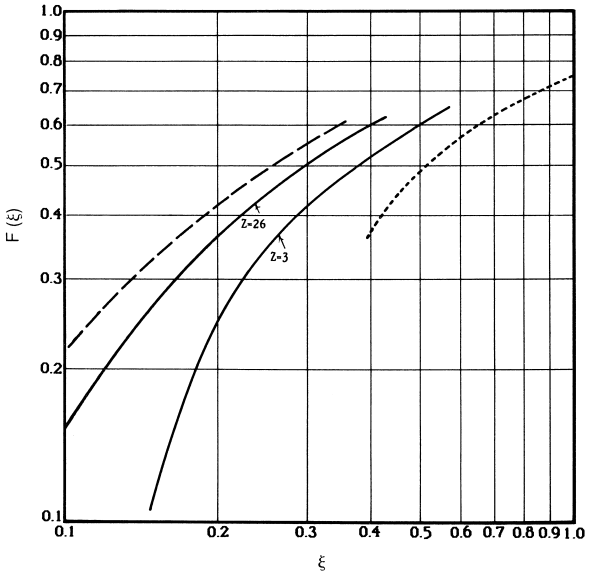


FIGURE 2 Pressure corrections factor $F(\xi)$ due to electrostatic interaction determined from the Thomas–Fermi method (dashed line), the Thomas–Fermi–Dirac method (solid lines), and by the assumption of uniform electron distribution (dotted line).

$$\rho \approx 12AZ\xi^3 \text{ g/cm}^3 \tag{30}$$

The values of $F(\xi)$ for all the curves shown in Fig. 2 are less than unity, and the corrected pressure is always less than that given by the noninteracting degenerate electron system. As ξ increases, $F(\xi)$ tends toward unity, implying that the electrostatic correction on the pressure becomes less and less significant with increasing density. At $\xi = 1$, the pressure of the system is about 20% less than that given by the electron degenerate pressure. For that value of ξ, the density of He matter is about 10^2 g/cm^3, while the density of Fe matter is already 10^4 g/cm^3. Thus, the proper treatment of electrostatic interaction is more relevant to the high-Z nuclei than to the low-Z nuclei. None of the curves shown in Fig. 2 is applicable to dense He matter, whose ξ values would be larger than unity. For He matter, electrostatic correction based on uniform electron distribution should be applicable. At a He matter density of 10^4 g/cm^3, the electrostatic correction accounts for a 5% reduction in pressure. Electrostatic correction based on uniform electron distribution may be applied to all matter system whenever the ξ values exceed unity.

Even though the Thomas–Fermi–Dirac method works quite satisfactorily for dense matter systems, the reader should, however, be reminded that it is not suitable for studying common metals whose densities are rather low. In these cases the cells are so large that the electrons fall into

orbital motions around the nucleus and must be handled differently.

H. High Temperatures

At high temperatures, of the order of millions of degrees Kelvin, the thermal energy of the particles in the system becomes comparable to the average kinetic energy of the electrons, in which case the zero-temperature equation of state must be corrected for thermal effects. We use the term "finite temperature" to denote a situation where temperature shows a perceptible effect on the energy state of the system. The proper treatment of finite temperature will first be discussed by ignoring the electrostatic interaction, which will be discussed later.

Without electrostatic interaction, the nuclei behave like a classical ideal gas, and the pressure P_A they contribute to the system is given by the classical ideal gas law:

$$P_A = \rho / A m_p k_B T \tag{31}$$

The electrons, on the other hand, must be handled by the Fermi gas method. At finite temperatures electrons no longer occupy all the low-energy states; instead, some electrons are thermally excited to high-energy states. Through quantum statistics it is found that the probability for occupation of a quantum state of momentum $\mathbf{p}$ is given by the function,

$$f(\mathbf{p}) = \{1 + \exp[(\varepsilon(\mathbf{p}) - \mu)/k_B T]\}^{-1} \tag{32}$$

where $\varepsilon(\mathbf{p}) = p^2/2m_e$ and μ is the chemical potential, which in the present noninteracting case is given by the Fermi energy ε_F:

$$\mu = \varepsilon_F = p_F^2/2m_e \tag{33}$$

It accounts for the energy needed to introduce an additional particle into the system. The Fermi–Dirac distribution $f(\mathbf{p})$ is the distribution of occupied quantum states in an identical fermion system in thermal equilibrium. Since the distribution that we shall employ depends on only the magnitude of the momentum and not its direction, it shall henceforth be written as $f(p)$. The finite temperature parameters of a noninteracting electron system are to be evaluated from:

$$n_e = \frac{2}{h^3} \int 4\pi p^2 \, dp \;\; f(p) \tag{34a}$$

$$\varepsilon_e = \frac{2}{h^3} \int 4\pi p^2 \, dp \; f(p) \, \varepsilon(p) \tag{34b}$$

$$P_e = \frac{2}{h^3} \int 4\pi p^2 \, dp \; f(p) \frac{pv}{3} \tag{34c}$$

where the range of the p integration extends from zero to infinity. Note that, for relativistic electrons, $\varepsilon(p)$ in Eq. (34b) should be replaced by e_k of Eq. (13) and v in Eq. (34c) should be changed from $v = p/m_e$ to that given in Eq. (15). The integrals in these equations cannot be evaluated analytically, and tabulated results of these integrals are available.

It is also useful to evaluate an entropy density (entropy per unit volume) for the system. The entropy density due to the electrons may be expressed in terms of quantities already evaluated as

$$s_e = (P_e + \varepsilon_e - n_e \mu)/T \tag{35}$$

The entropy density due to the nuclei is

$$s_A = (P_A + \varepsilon_A)/T = (5/3) P_A / T \tag{36}$$

where ε_A is the energy density of the nuclei. The entropy density has the same dimensions as Boltznman's constant k_B.

The equation of state relates the pressure of the system to its density and temperature. Since the pressure is determined by two independent parameters, it is hard to display the result. Special cases are obtained by holding either the temperature constant or the entropy of the system constant as the density is varied. An equation of state that describes a thermodynamic process in which the temperature remains constant is called an *isotherm*, and an equation of state that describes an adiabatic process for which the entropy of the system remains constant is called an *adiabat*. In the case where the total number of particles in the system remains unchanged as its volume is varied, constant entropy means the entropy per particle remains a constant, which is obtained by dividing the entropy density given by either Eq. (35) or Eq. (36) by the particle number density. In Fig. 3, typical isotherms and adiabats for a system of (noninteracting) neutrons at high densities are shown. Such a system of neutrons is called *neutron matter*, which is an important form of dense matter and will be discussed in detail later. Neutrons are fermions, and the method of Fermi gas employed for the study of an electron system may be applied in the same manner to the neutron system. In Fig. 3 it is evident that the adiabats have much steeper rises than the isotherms, and this feature holds for all matter systems in general.

The inclusion of electrostatic interaction to a finite-temperature system can be dealt with by extending the Thomas–Fermi–Dirac method described before. At a finite temperature the electron distribution in the system's quantum states must be modified from that of degenerate electrons to that of partially degenerate electrons as determined by the Fermi–Dirac distribution of Eq. (32). Results of this type have been obtained, but they are too elaborate to be summarized here. Interested readers are referred to

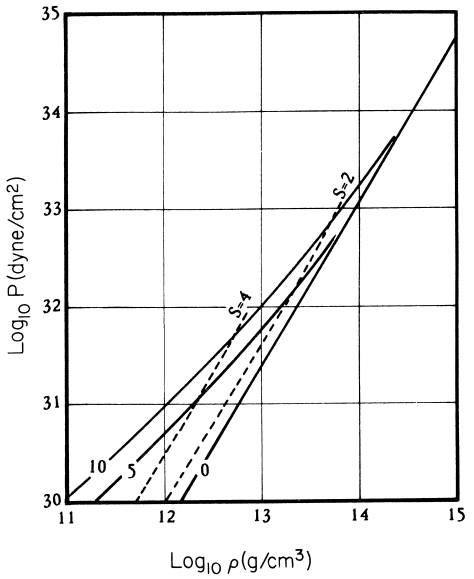


FIGURE 3 Isotherms (solid lines) and adiabats (dashed lines) for a dense system of neutrons computed by the Fermi gas method with complete neglect of nuclear interaction. Isotherms shown are computed for temperatures equivalent to $k_B T = 0, 5$, and 10 MeV, and the adiabats for entropy per neutron equal to $S/k_B = 2$ and 4.

references listed in the bibliography, where reference to the original articles on this topic can be found.

This section concentrates mainly on simple matter systems with densities lying in the first density domain, but the concepts discussed are applicable to the study of dense matter at all densities. At higher densities, the electrostatic interaction is replaced by nuclear interaction as the dominant form of interaction. The discussion of nuclear interaction will be deferred to the next section.

III. COMPOSITION OF DENSE MATTER

The composition of dense matter will be discussed separately for the four density domains delineated in Section I. In the first density domain where density is 10^2–10^7 g/cm^3, dense matter can in principle exist in any composition, especially at low temperatures. The constituent nuclei of the matter system may belong to any nuclide or combination of nuclides listed in the nuclear table. A sufficient number of electrons must be present to render the system electrically neutral. Stated in such general terms, not much more needs to be added. However, since dense matter is normally found in the interior of stellar objects, its composition naturally depends on nucleosynthesis related to stellar evolution. A review of the stellar evolution process will suggest the prevalent compositions of dense matter in this density domain and the conditions under which they exist.

A star derives its energy through the fusion of hydrogen nuclei to form ^{4}He. It is called *hydrogen burning*, and its ignition temperature is about 10^7 K. The resulting helium nuclei are collected at the core of the star. As the helium core becomes hot enough and dense enough (estimated to be 1.5×10^8 K and 5×10^4 g/cm^3), nuclear reactions based on ^{4}He occur, initiating the helium flash. Helium burning generates the following sequence of relatively stable products: $^{12}\text{C} \rightarrow {}^{16}\text{O} \rightarrow {}^{20}\text{Ne} \rightarrow {}^{24}\text{Mg}$. These ^{4}He-related processes account for the fact that after ^{1}H and ^{4}He, which are respectively the most and second most abundant nuclear species in steller systems, ^{16}O is the third most abundant, ^{12}C the fourth, and ^{20}Ne the fifth. Both hydrogen burning and helium burning proceed under high temperatures and high densities. The reason is nuclei are electrically charged and do not normally come close enough for reactions to take place unless their thermal energy is large enough to overcome the repulsive electrical force. The repulsive force increases with the electric charge Z of the nuclei, which is why reactions of this type begin with low-Z nuclei and advance to high-Z nuclei as core temperature rises and density increases. These are exothermic reactions in which energy is generated.

As helium burning proceeds, the helium-exhausted core contracts sufficiently to initiate $^{12}\text{C} + {}^{12}\text{C}$ and $^{16}\text{O} + {}^{16}\text{O}$ reactions, called *carbon burning* and *oxygen burning*, respectively. The final products of these processes are ^{24}Mg, ^{28}Si, and ^{32}S. Upon conclusion of oxygen burning, the remanent nuclei may not wait for the further contraction of the core to bring about reactions such as $^{24}\text{Mg} + {}^{24}\text{Mg}$, since the electrical repulsion between them is very large and the temperature and density needed to initiate such reactions are correspondingly high. Instead, these nuclei react with photons to transform themselves in successive steps to the most tightly bound nuclide in the nuclear table, which is ^{56}Fe, in a process called *photodisintegration rearrangement*. Therefore, the most likely composition of matter with a density reaching the high end of the first density domain is of electrons and ^{56}Fe nuclei.

A. Neutronization

As matter density advances into the second density domain, 10^7–10^{12} g/cm^3, the ground-state composition of the dense matter system changes with density. For example, if a dense matter system composed of ^{56}Fe nuclei at 10^7 g/cm^3 is compressed, then as density increases the constituent nuclei transmute in a sequence such as Fe $\rightarrow$ Ni $\rightarrow$ Se $\rightarrow$ Ge and so on. Transmutations occur because the availability of high-energy electrons in a dense system is making such reactions possible, and transmutations result if the transmuted nuclei lower the total energy of the matter system. This process is called *neutronization*, because

the resulting nuclide from such a reaction is always richer in neutron content than that entering into the reaction.

The physical principle involved can be seen from the simplest of such reactions, in which an electron, e, is captured by a proton, p, to produce a neutron, n, and a neutrino, ν:

$$\mathrm{e} + \mathrm{p} \rightarrow \mathrm{n} + \nu \tag{37}$$

This is an endothermic reaction. For it to proceed, the electron must carry with it a substantial amount of kinetic energy, the reason being that there is a mass difference between the neutron and proton that, expressed in energy units, is $(m_\mathrm{n} - m_\mathrm{p})c^2 = 1.294$ MeV, where $1\,\mathrm{MeV} = 10^6$ eV. However, when matter density exceeds 10^7 g/cm^3, electrons with kinetic energies comparable to the mass difference become plentiful, and it is indeed energetically favorable for the system to have some of the high-energy electrons captured by protons to form neutrons. The neutrinos from the reactions in general escape from the system since they interact very weakly with matter at such densities. Reactions based on the same principle as Eq. (37) are mainly responsible for neutronization of the nuclei in the matter system.

Let us denote the mass of a nucleus belonging to a certain nuclide of atomic number Z and mass number A by $M(Z, A)$ and compare the energy contents of different pure substances, each of which is composed of nuclei of a single species. Each system of such a pure substance is characterized by a nuclei number density n_A. Since n_A is different for different systems, it is best to introduce instead a nucleon number density n_B. All systems with the same density have the same n_B, where the subscript B refers to baryons, a generic term including nucleons and other nucleonlike particles. Each baryon is assigned a unit baryon number, which is a quantity conserved in all particle reactions. Thus, n_B will not be changed by the reaction shown in Eq. (37) or the neutronization process in general, which makes it a useful parameter. It is related to the nuclei number density by $n_\mathrm{B} = An_\mathrm{A}$ and to the electron number density by $n_\mathrm{e} = n_\mathrm{B}Z/A$.

At zero temperature, the energy density of such a system is given by:

$$\varepsilon = (n_\mathrm{B}/A)M(Z, A)c^2 + \varepsilon_\mathrm{s} + \varepsilon_\mathrm{e} \tag{38}$$

where ε_s is the electrostatic interaction energy density and ε_e the degenerate electron energy density, which is to be calculated from Eq. (16b) for relativistic electrons. The masses of most stable nuclides and their isotopes have been determined quite accurately and are listed in the nuclear table. One may therefore compute ε according to Eq. (38) for all candidate nuclides using listed masses from the nuclear table. As a rule, only even–even nuclides, which are nuclides possessing even numbers of protons and even numbers of neutrons, are needed for consideration. These nuclides are particularly stable and are capable of bringing the dense matter system to a low energy state.

The electrostatic interaction energy density ε_s may be estimated from methods discussed in Section II. Even though electrostatic interaction energy plays a minor role in determining the pressure of the system once the density is high, it has an effect on the composition of the system, which depends on a relatively small amount of energy difference. The ε_s may be estimated by evaluating the lattice energy due to the electrostatic interaction energy of a system of nuclei that form into a lattice, while the electrons are assumed to be distributed uniformly in space. The lattice giving the maximum binding is found to be the body-centered-cubic lattice, and the corresponding lattice energy is

$$\varepsilon_\mathrm{L} = -0.89593(Ze)^2/r_\mathrm{s} \tag{39}$$

where r_s is the cell radius defined for Eq. (21) and is related to n_B by:

$$r_\mathrm{s} = (4\pi n_\mathrm{B}/3A)^{-1/3}$$

For each nuclide, the energy density of the system evaluated from Eq. (38) is a function of n_B. Next, plot the quantity ε/n_B versus $1/n_\mathrm{B}$ as shown in Fig. 4, where the curves corresponding to nuclides ^{62}Fe, ^{62}Ni, and ^{64}Ni are drawn. From the curves indicated by ^{62}Fe and ^{62}Ni, it is obvious that a matter system composed of ^{62}Ni nuclei would have an energy lower than that composed of ^{62}Fe nuclei at the densities shown. The ground-state composition of

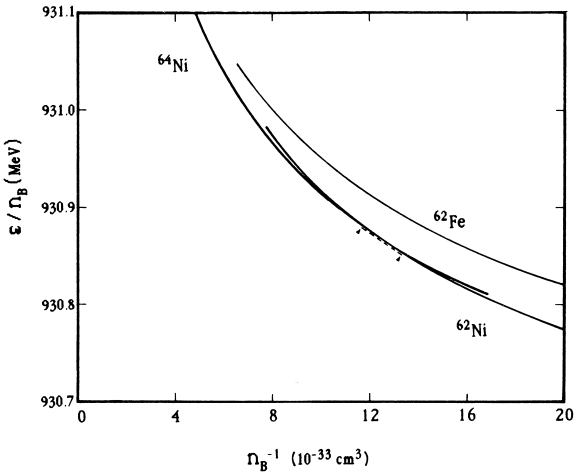


FIGURE 4 Schematic energy curves for matter composed of nuclides ^{62}Fe, ^{62}Ni, and ^{64}Ni illustrating the tangent construction method for establishing the domain of a first-order phase transition in matter. The short dashed line is a line tangent to the ^{62}Ni and ^{64}Ni curves. The triangular pointers indicate the points of tangency. [After Leung, Y. C. (1984). "Physics of Dense Matter," Science Press, Beijing, China. By permission of Science Press.]

dense matter is determined by nuclides whose curves form the envelope to the left of all the curves, as illustrated, for example, by the curves for ^{62}Ni and ^{64}Ni in Fig. 4. Two of these curves on the envelope cross at a certain density. This means that a change in composition occurs in the neighborhood of that density. Such a change, called a *first-order phase transition*, begins at a density below the density where the curves cross, when some of the ^{62}Ni nuclei are being rearranged into ^{64}Ni, and ends at a density above that intersection when all of the ^{62}Ni nuclei are changed into ^{64}Ni. Throughout a first-order phase transition, the pressure of the system remains constant, which is a requirement of thermal equilibrium. The exact density at which a phase transition begins and the density at which it ends can be found from the tangent construction method shown in Fig. 4. By this method one draws a straight-line tangent to the convex curves on the envelope as shown by the dotted lines. The values of n_B at the points of tangency correspond to the onset and termination of phase transition. All points on the tangent exhibit the same pressure since the expression for pressure, Eq. (9), may be converted to read:

$$P = -\frac{\partial(\varepsilon/n_B)}{\partial(1/n_B)} \tag{40}$$

Within the range of phase transition, the tangent now replaces the envelope in representing the energy of the matter system.

The sequence of nuclides constituting the zero-temperature ground state of dense matter in the density range 10^7–10^{11} g/cm^3 obtained by this method with ε_s replaced by ε_L is shown in Table II. As indicated, neutronization begins at a density of 8.1×10^6 g/cm^3, which is quite close to 10^7 g/cm^3, and that is the reason why we choose 10^7 g/cm^3 to mark the begining of the second density domain. Note that the sequence of nuclides in the table shows a relative increase in neutron content with density, which is indicated by the diminishing Z/A ratios.

The lattice structure can be destroyed by thermal agitation. At each density, there is a corresponding melting temperature T_m that can be estimated by comparing the lattice energy with its thermal energy. It is usually taken to be

$$k_B T_m = -c_m^{-1}(Ze)^2/r_s \tag{41}$$

where $c_m \approx 50$–100. Thus, for Fe matter at $\rho = 10^8$ g/cm^3, the melting temperature is about $T_m \approx 2 \times 10^8$ K.

B. Nuclear Semiempirical Mass Formula

As matter density exceeds 4×10^{11} g/cm^3 or so, the ground-state composition of the matter system may prefer nuclides that are so rich in neutrons that these nuclides do not exist under normal laboratory conditions, and their masses would not be listed in nuclear tables. To understand these nuclides, theoretical models of the nucleus must be constructed to deduce their masses and other properties. This is a difficult task since nuclear forces are complicated, and the problem is involved. A preliminary investigation of this problem should rely on as many empirical facts about the nucleus as possible. For the present purpose, the nuclear semiempirical mass formula, which interpolates all known nuclear masses into a single expression, becomes a useful tool for suggesting the possible masses of these nuclides.

TABLE II Ground-State Composition of Dense Matter in the Second Density Domain[a]

Nuclide	Z	Z/A	ρ_{max} (g/cm^3)
^{56}Fe	26	0.4643	8.1×10^6
^{62}Ni	28	0.4516	2.7×10^8
^{64}Ni	28	0.4375	1.2×10^9
^{84}Se	34	0.4048	8.2×10^9
^{82}Ge	32	0.3902	2.2×10^{10}
^{80}Zn	84	0.3750	4.8×10^{10}
^{78}Ni	28	0.3590	1.6×10^{11}
^{76}Fe	26	0.3421	1.8×10^{11}
^{124}Mo	42	0.3387	1.9×10^{11}
^{122}Zr	40	0.3279	2.7×10^{11}
^{120}Sr	38	0.3166	3.7×10^{11}
^{118}Kr	36	0.3051	4.3×10^{11}

[a] ρ_{max} is the maximum density at which the nuclide is present. [From Baym, G., Pethick, C., and Sutherland, P. (1971). *Astrophysical J.* **170,** 299. Reprinted with permission of *The Astrophysical Journal*, published by the University of Chicago Press; © 1971 The American Astronomical Society.]

The nuclear mass is given by the expression:

$$M(Z, A) = (m_p + m_e)Z + m_n(A - Z) - B(Z, A) \tag{42}$$

where m_p, m_e, and m_n denote the proton, electron, and neutron mass, espectively, and $B(Z, A)$ the binding energy of the nucleus. The nuclear semiempirical mass formula for even–even nuclides expresses the binding energy in the following form:

$$B(Z, A) = a_V A - a_S A^{2/3} - a_C Z^2 A^{-1/3} - \frac{a_A(A - 2Z)^2}{A} + a_P A^{-3/4} \tag{43}$$

where the coefficients have been determined to be $a_V = 15.75$ MeV, $a_S = 17.8$ MeV, $a_C = 0.710$ MeV, $a_A = 23.7$ MeV, and $a_P = 34$ MeV. This mass formula is not just a best-fit formula since it has incorporated many

theoretical elements in its construction. This feature, hopefully, will make it suitable for extension to cover unusual nuclides.

Applying Eqs. (42) and (43) to the computation of the matter energy density given by Eq. (38), one finds matter composition with density approaching 4×10^{11} g/cm^3. The general result is that as matter density increases, the constituent nuclei of the system become more and more massive and at the same time become more and more neutron rich. In general, nuclei become massive so as to minimize the surface energy which is given by the term proportional to a_S in Eq. (43), and they turn neutron rich so as to minimize the electrostatic interaction energy within the nuclei, given by the term proportional to a_C in Eq. (43). In general, the nuclear semiempirical mass formula is not believed to be sufficiently accurate for application to nuclides whose Z/A ratios are much below $Z/A \approx 0.3$. For these nuclides, we must turn to more elaborate theoretical models of the nucleus.

C. Neutron Drip

When the nuclei are getting very large and very rich in neutron content, some of the neutrons become very loosely bound to the nuclei; as density is further increased, unbound neutrons begin to escape from the nuclei. The nuclei appear to be immersed in a sea of neutrons. This situation is called *neutron drip*, a term suggesting that neutrons are dripping out of the giant nuclei to form a surrounding neutron sea. At zero temperature, this occurs at a density of about 4×10^{11} g/cm^3. When it occurs, matter becomes a two-phase system, with one phase consisting of the nuclei and the other consisting of neutrons (with possibly a relatively small admixture of protons). The nuclei may be visualized as liquid drops suspended in a gas consisting of neutrons. These two phases coexist in phase equilibrium. The energy densities of these two phases are to be investigated separately. If the surface effects around the nuclei may be neglected, then these two phases are assumed to be uniform systems, each of which exhibits a pressure and whose neutron and proton components reach certain Fermi energies [cf. Eq. (26)]. At phase equilibrium, the pressures of these two phases must be identical, and the Fermi energies of the respective neutron and proton components of these two phases must also be identical. The composition of matter in the density domain of neutron drip is established by the system that is in phase equilibrium and at the same time reaching the lowest possible energy state.

The evaluation of the Fermi energies for the neutron and proton components depends on the nature of the nuclear interaction. Since the nuclear interaction is not as well known as the electrostatic interaction, the results derived for post-neutron drip densities are not as well established as those found for the first density domain. At the present time, several forms of effective nuclear interactions constructed specifically to explain nuclear phenomenology are quite helpful for this purpose. Effective nuclear interaction is to be distinguished from realistic nuclear interaction, which is derived from nuclear scattering data and is regarded as a more fundamental form of nuclear interaction.

D. Liquid Drop Model

A nucleus in many respects resembles a liquid drop. It possesses a relatively constant density over its entire volume except near the surface, and the average interior density is the same for nuclei of all sizes. A model of the nucleus that takes advantage of these features is the liquid drop model. It considers the total energy of the nucleus to be the additive sum of its bulk energy, surface energy, electrostatic energy, and translational energy. The bulk energy is given by multiplying the volume of the nucleus by the energy density of a uniform nuclear matter system with nuclear interaction included. The surface energy is usually taken to be a semiempirical quantity based on calculations for nuclei of finite sizes and on experimental results on laboratory nuclei. In a two-phase situation the surface energy must be corrected for nucleon concentration outside the nucleus. The electrostatic energy involves the interaction among all charged particles inside and outside the nucleus. The position of the nuclei may again be assumed to form a body-centered-cubic lattice. Both electrostatic lattice energy and electrostatic exchange energy contribute to it. The translational energy is due to the motion of the nucleus.

At high temperatures, the bulk energy must be computed according to a system of partially degenerate nucleons. This is usually done by employing effective nuclear interaction, in which case the nucleons are assumed to be uncorrelated, and the Fermi–Dirac distribution of Eq. (32) may be directly applied to the nucleons. The problem is much more complicated if the realistic nuclear interaction is employed, in which case particle correlations must be included, and for this reason the computation is much more elaborate. The surface energy shows a reduction with temperature. This result can be extracted from finite nuclei calculations. The lattice energy also shows temperature modification, since nuclei at the lattice points agitate with thermal motions. The same thermal motion also contributes to the translational energy of the nuclei, which may be assumed to possess thermal velocities given by the Boltzmann distribution.

The results of a study based on the liquid drop model of the nucleus is depicted in Fig. 5, where the variation of the size of a nucleus (its A number) with matter

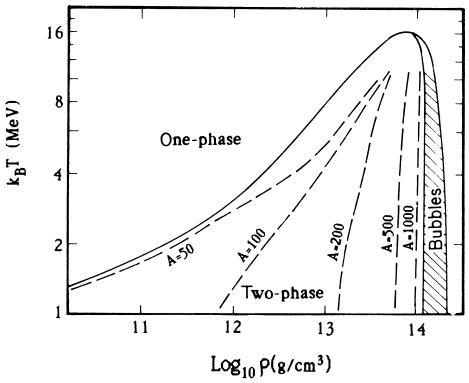


FIGURE 5 Variations in composition of hot dense matter with $Y_e = 0.25$. A two-phase system exists in the region under the uppermost solid line and a one-phase system above it. The dashed lines indicate the types of equilibrium nuclides participating in the two-phase system. [After Lamb, D. Q., Lattimer, J. M., Pethick, C. M., and Ravendall, D. G. (1978). *Phys. Rev.* **41,** 1623. By permission of the authors.]

density and temperature is shown. The study is done with an effective nuclear interaction called the *Skyrme interaction* and for a matter system having an overall electron fraction $Y_e = 0.25$. A wide range of temperatures, expressed in terms of k_BT in units of MeV, is included. Note that the temperature corresponding to $k_BT = 1$ Mev is $T = 1.16 \times 10^{10}$ K. The dashed lines indicate the A numbers of the nuclei at various densities and temperatures. The solid line forms the boundary separating a one-phase system from a two-phase system. The conditions for a two-phase system are included in the plot under the solid line. The two-phase condition disappears completely when matter density exceeds 2×10^{14} g/cm^3. Just before that density, there exists a density range where the nuclei in the system merge and trap the surrounding neutrons into a form of bubbles. This range is indicated by the cross-hatched area in the plot and is labeled "Bubbles."

E. Neutron Matter

In general, when matter density exceeds 2×10^{14} g/cm^3, all nuclei are dissolved into a homogeneous system, and for a range of densities above that, matter is composed almost entirely of neutrons. The admixture of protons is negligibly small (about 1%), because all protons must be accompanied by an equal number of electrons which, being very light, contribute a large amount of kinetic energy to the system. Hence, all ground-state systems tend to avoid the presence of electrons. Such a nearly pure neutron system is called *neutron matter*.

The average density of the atomic nucleus, quite independent of its species, is approximately 2.8×10^{14} g/cm^3; therefore, many of the methods employed for the study of the nucleus may be applied to the study of neutron matter. As we have mentioned before, the distribution of mass density of a nucleus has been found to be quite uniform over its entire volume, and that uniform mass density is very nearly the same from one nuclide to the other. For a heavy nucleus, its volume is extended so as to maintain a mass density common to nuclei of all sizes. This remarkable fact is described as nuclear saturation. Nuclear substance possessing this mass density is called *nuclear matter*, and its mass density, called *nuclear density*, is found to be 2.8×10^{14} g/cm^3.

Neutron matter, however, is not nuclear matter. Neutron matter consists nearly entirely of neutrons, whereas nuclear matter consists of neutrons and protons in roughly equal fractions. While nuclear matter forms bound units or nuclei, neutron matter is an unbound system. The ground-state composition of an extended system is not given by that of nuclear matter but neutron matter. Neutron matter does not exist terrestrially; its existence is only inferred theoretically. Neutron matter may be studied in analogy to nuclear matter. To accomplish this there are methods based on realistic nuclear interaction expressed in the form of nuclear potentials, the best known of which are the Brueckner–Bethe–Goldstone method and the constrained variational method. There are also methods based on the use of effective nuclear interaction. These methods combine the Hartree–Fock method with some form of effective nuclear potential. Though they are considered less fundamental than the two methods mentioned before, they usually yield more direct and accurate results at the nuclear density where they are designed to perform. Since neutron matter exists for a range of densities, the choice of an applicable method depends also on how well these methods may be extended to cover such a wide range of densities. In this respect, a phenomenological model called the *relativistic mean field model* seems very attractive. We shall return to these methods in the next section when the equation of state of neutron matter is discussed.

It has been seriously proposed that known nuclear interaction forces would make neutron matter superfluid and the proton component in it superconductive when its temperature is below a critical temperature T_c which is estimated to be as high as 10^9–10^{10} K. These phenomena would have profound effects on the transport properties of neutron matter and will be discussed in more detail in Sections V and VI.

Since neutrons, like electrons, are fermions and obey Pauli's exclusion principle, many features of a dense degenerate electron system are exhibited by neutron matter. In spite of having a strong nuclear interaction, these features still play dominant roles in the neutron system. Thus, results based on a dense degenerate neutron system serve as useful guides to judge the properties of neutron

matter over a wide range of densities. Such guidance is particularly needed when hyperons and other massive particles begin to make their appearance in dense matter, because the precise nature of their interactions is still poorly known.

F. Baryonic Matter

As matter density advances beyond 10^{15} g/cm^3 and into the fourth and last density domain, many new and unexpected possibilities may arise. It is fairly sure that light hyperons appear first. Hyperons are like nucleons in every aspect except that they are slightly more massive and carry a nonzero attribute called *hypercharge* or *strangeness*. They are $\Lambda(1115)$ and $\Sigma(1190)$, where the numbers in the parentheses give their approximate masses, expressed as mc^2 in units of MeV. Λ is electrically neutral, but there are three species of Σ with charges equal to $+1$, 0, and -1 of the proton charge. More exact values of their masses may be found in a table of elementary particles. These particles are produced in reactions such as:

$$\mathrm{p} + \mathrm{e} \rightarrow \Lambda + \nu \tag{44a}$$

$$\mathrm{p} + \mathrm{e} \rightarrow \Sigma^0 + \nu \tag{44b}$$

$$\mathrm{n} + \mathrm{e} \rightarrow \Sigma^- + \nu \tag{44c}$$

The hyperons appear as soon as the kinetic energies of the neutrons exceed the mass difference between neutron and these particles, in a manner similar to the neutronization process.

In order of ascending mass, the next group of nucleonlike particles to appear are the $\Delta(1232)$, which come in four species with electric charges equal to $+2$, $+1$, 0, and -1 of the proton charge. Δ are produced as excited states of the nucleons. The nucleons, hyperons, and Δ fall into the general classification of baryons. All baryons are assigned the baryon number $+1$ and antibaryons the number of -1. The baryon number is conserved additively in all particle reactions. Matter systems composed of nucleons, hyperons, Δ, and possibly other baryons are called *baryonic matter*.

G. Pion Condensation

Pions, $\pi(140)$, fall into a class of elementary particles called *mesons* which are quite different from the baryons. They participate in nuclear reactions with the baryons, but they are bosons and not restricted by Pauli's exclusion principle. They are produced in a large variety of particle interactions. They do not normally appear in a matter system, because energy is needed to create them, and their presence would mean an increase in the total energy of the matter system. They do appear, however, if their interaction with the nucleons creates an effective mass for the nucleon that is lower than its actual mass by an amount comparable to the mass of the pion. This occurs when the baryonic matter reaches some critical density, estimated to be of the order of 10^{15} g/cm^3. This result is not definite since the elementary particle interaction is still far from being understood. Different estimates yield rather different results. When the pions do appear, they may possess very similar momenta as required by the interaction, which is momentum dependent. Such a state of a boson system is called a *condensate*. The appearance of a pion condensate in dense matter is called *pion condensation.*

H. Quark Matter

With the advent of the quark theory some of the traditional notions about elementary particles must be revised. The baryons and mesons, which have been traditionally called *elementary particles*, must be viewed as composite states of quarks. Current quark models have achieved such success that it is hardly in doubt that quark theory must play a key role in the understanding of the subparticle world. Quarks are spin-half fermions possessing fractional electric charges such as $\frac{2}{3}$, $\frac{1}{3}$ and $-\frac{1}{3}$ of the proton charge. Besides mass, spin, and electric charge, they possess additional attributes such as c-charge (also called *color*) and f-charge (also called *flavor*). The electric charge will be called *q-charge* in the present terminology. In a quark model, the baryons mentioned before are the bound states of three quarks, and the mesons are the bound states of quark and antiquark paris. Multiquark states involving more than three quarks are also possible. The interactions of the quarks are governed by quantum chromodynamics (QCD) through their c-charges, quantum flavodynamics (QFD) through their f-charges, and quantum electrodynamics (QED) through their q-charges. The bound-state configurations of the quarks are described by QCD, which is a highly nonlinear theory, and for this reason the transition from one form of manifestation to another can occur abruptly.

Such an abrupt transition is believed to be responsible for the distinct boundary of a nucleon, for example. Inside the nucleon boundary, where three quarks are in close proximity to each other, the effective quark interaction is very weak, but once any one of them reaches the boundary, the interaction turns strong so rapidly that the quark does not have enough energy to penetrate the boundary, thus, in effect, the quarks are confined inside the boundary. Each nucleon, or baryon, therefore occupies a volume.

What if the baryon density inside a matter system is so high that the volumes occupied by them are crushed? Naturally, the boundaries will be gone, and all quarks merge to form a uniform system. Such a state of dense matter is called *quark matter*. The transition to quark

matter has been estimated to occur at densities as low as 2 to 4×10^{15} g/cm^3 at zero temperature and at even lower densities at high temperatures. Current estimates of the average quark-confining energy is about 200 MeV per quark; therefore, if the temperature is above 2×10^{12} K or a corresponding thermal energy of $k_B T = 200$ MeV, the nucleons are vaporized and the quarks set free, forming a sort of quark gas. These results, we hasten to add, are only tentative.

There is very little confidence in postulating anything beyond the quark matter, and for this reason we shall end our discussion at this point.

IV. EQUATION OF STATE OF DENSE MATTER

The equation of state of a substance is a functional relation between the pressure generated by the substance and its density and temperature. It reflects the composition and internal structure of the substance. The equation of state of dense matter is of prime importance to astrophysical study of compact stellar objects such as white dwarf stars and neutron stars. It will be discussed in detail for the first three density domains. Our knowledge about dense matter in the fourth density domain is insufficient to provide accurate quantitative evaluation of its properties at this time.

A. The First Density Domain

For the first density domain (10^2–10^7 g/cm^3), we shall concentrate our discussion on a matter system composed of Fe^{56} nuclei, which represents the most stable form of matter system in that domain and shall be referred to as Fe matter. The equation of state of Fe matter at zero temperature in this density range is determined basically by the degenerate electrons in the system. Electrostatic interaction among the charged particles plays a relatively minor role, which has been demonstrated in Section II. Nevertheless, the degenerate electron pressure must be corrected for electrostatic effects. For Fe matter with densities below 10^4 g/cm^3, electrostatic interaction must be computed on the basis of the Thomas–Fermi–Dirac method which is discussed in Section II; for Fe matter with densities above 10^4 g/cm^3, the electrostatic interaction may be computed on the assumption that electrons are distributed uniformly around the nuclei.

Figure 6 shows the ground-state equation of the state of dense matter at zero temperature in a log–log plot over the first three density domains. The portion below 10^7 g/cm^3 belongs to the equation of state for Fe matter in the first density domain. The numerical values of this equation of state are given in Table III under the heading BPS. The adiabatic indices are listed under Γ. The equation of state in this density domain should be quite accurate since its composition is well established.

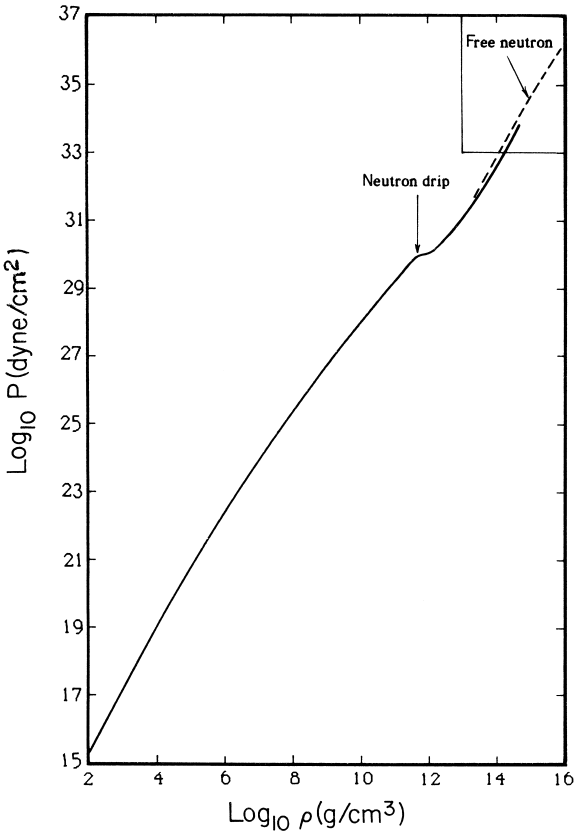


FIGURE 6 The ground-state equation of state of dense matter. The numerical values of the solid line are listed in Table III under the heading BPS. The onset of neutron drip is indicated. The different versions of the equation of state in the high-density region (outlined by rectangle) are given in Fig. 7. For correlation purposes, the equation of state of a noninteracting neutron system, indicated by "free neutron," is drawn here.

At high temperatures, degenerate electrons become partically degenerate in a way described by Eq. (32). Finite temperature equations of state will not be shown here. General behaviors of the isotherms and adiabats should be similar to those shown in Fig. 3. At densities below 10^4 g/cm^3, where the energy correction due to electrostatic interaction is important, the Thomas–Fermi–Dirac method must be extended to include thermal effects. Interested readers are referred to references listed in the Bibliography for further details.

B. The Second Density Domain

The ground-state composition of matter in the second density domain (10^7–10^{12} g/cm^3) is described in Section III. The composition varies with density due to neutronization

TABLE III Ground-State Equation of State of Dense Matter

BPS				
ρ (g/cm^3)	P (dyn/cm^2)	Z	A	Γ
2.12×10^2	5.82×10^{15}	26	56	—
1.044×10^4	9.744×10^{18}	26	56	1.796
2.622×10^4	4.968×10^{19}	26	56	1.744
1.654×10^5	1.151×10^{21}	26	56	1.670
4.156×10^5	5.266×10^{21}	26	56	1.631
1.044×10^6	2.318×10^{22}	26	56	1.586
2.622×10^6	9.755×10^{22}	26	56	1.534
1.655×10^7	1.435×10^{24}	28	62	1.437
3.302×10^7	3.833×10^{24}	28	62	1.408
1.315×10^8	2.604×10^{25}	28	62	1.369
3.304×10^8	8.738×10^{25}	28	64	1.355
1.045×10^9	4.129×10^{26}	28	64	1.344
2.626×10^9	1.272×10^{27}	34	84	1.340
1.046×10^{10}	7.702×10^{27}	32	82	1.336
2.631×10^{10}	2.503×10^{28}	30	80	1.335
6.617×10^{10}	8.089×10^{28}	28	78	1.334
1.049×10^{11}	1.495×10^{29}	28	78	1.334
2.096×10^{11}	3.290×10^{29}	40	122	1.334
4.188×10^{11}	7.538×10^{29}	36	118	1.334
4.460×10^{11}	7.890×10^{29}	40	126	0.40
6.610×10^{11}	9.098×10^{29}	40	130	0.40
1.196×10^{12}	1.218×10^{30}	42	137	0.63
2.202×10^{12}	1.950×10^{30}	43	146	0.93
6.248×10^{12}	6.481×10^{30}	48	170	1.31
1.246×10^{13}	1.695×10^{31}	52	200	1.43
2.210×10^{13}	3.931×10^{31}	58	241	1.47
6.193×10^{13}	1.882×10^{32}	79	435	1.54
1.262×10^{14}	5.861×10^{32}	117	947	1.65
2.761×10^{14}	2.242×10^{33}	—	—	1.82
5.094×10^{14}	7.391×10^{33}	—	—	2.05

FP[b]		SKM	
ρ (g/cm^3)	P (dyn/cm^2)	ρ (g/cm^3)	P (dyn/cm^2)
4.11×10^{13}	1.01×10^{32}	4.11×10^{13}	8.00×10^{31}
5.64×10^{13}	1.52×10^{32}	5.64×10^{13}	1.42×10^{32}
9.75×10^{13}	3.36×10^{32}	9.75×10^{13}	4.34×10^{32}
1.55×10^{14}	8.60×10^{32}	1.55×10^{14}	1.23×10^{33}
2.31×10^{14}	2.30×10^{33}	2.31×10^{14}	3.21×10^{33}
3.21×10^{14}	5.71×10^{33}	3.29×10^{14}	7.64×10^{33}
4.97×10^{14}	1.82×10^{34}	4.51×10^{14}	1.67×10^{34}
7.72×10^{14}	6.22×10^{34}		
1.23×10^{14}	2.21×10^{35}		

RMF	
ρ (g/cm^3)	P (dyn/cm^2)
3.36×10^{14}	9.69×10^{33}
4.60×10^{14}	3.39×10^{34}
6.25×10^{14}	9.04×10^{34}
8.46×10^{14}	1.99×10^{35}
1.14×10^{15}	3.73×10^{35}
2.25×10^{15}	1.12×10^{36}
7.42×10^{15}	5.03×10^{36}

[a] From Baym, G., Pethick, C., and Sutherland, P. (1971). *Astrophysical Journal* **170,** 199. Reprinted with permission of *The Astrophysical Journal*, published by the University of Chicago Press; ©1971 The American Astronomical Society.

[b] From Friedman, B., and Pandharipande, V. R. (1981). *Nuclear Physics* **A361,** 502. Reprinted with permission of *Nuclear Physics*: ©1981 North-Holland Publishing Co., Amsterdam.

of the constituent nuclei. As the composition changes from one form to another, a first-order phase transition is involved and, over the density range where phase transition occurs, the pressure remains constant. Therefore, in a detailed plot of the equation of state in this density domain, pressure rises with density except at regions of phase transition, where it remains constant. The equation of state appears to rise in steps with increasing density, but since the steps are quite narrow, the equation of state in this density domain may be approximated by a smooth curve.

The pressure of the matter system is due entirely to the degenerate electrons. In establishing the ground-state composition of the matter system, electrostatic interaction energy in the form of lattice energy has been included, but it is quite negligible as far as the pressure of the system is concerned. The pressure in this density domain does not rise as rapidly with density as it does in the first density domain because the ground-state composition of the system shows a gradual decline in the Z/A values with density. The composition is quite well established up to a density of 4×10^{11} g/cm^3, which marks the onset of neutron drip. The composition consists of nuclei found under normal laboratory conditions or their nearby isotopes, and their masses can be either measured or extrapolated from known masses with reasonable certainty. The ground-state equation of state at zero temperature in this density range is shown in Fig. 6. The numerical values of the equation of state before neutron drip, together with the atomic number Z and mass number A of the constituent nuclei at these densities, are listed in Table III under the heading BPS.

C. The Third Density Domain

The third density domain (10^{12}–10^{15} g/cm^3) begins properly with the onset of neutron drip, which occurs at a density of 4×10^{11} g/cm^3. With the onset of neutron drip, matter is composed of giant nuclei immersed in a sea of neutrons. Nucleons inside the nuclei coexist with nucleons outside the nuclei forming a two-phase system. Theoretical studies of such a system rely heavily on the nuclear liquid drop model which gives a proper account of the different energy components in a nucleus. The nuclear liquid drop model is described in Section III. At each density, matter is composed of a particular species of nuclei characterized by certain Z and A numbers, which exist in phase equilibrium with the neutron sea outside which has a much lower Z/A ratio. Such a two-phase system constitutes the ground-state composition of matter in this density range.

Once the ground-state composition is established, the equation of state can be found as before by establishing the electron fraction Y_e of the system and proceeding to

evaluate the electron pressure. The electron pressure remains the main pressure component of the system until neutron pressure takes over at higher densities. Electron number density, however, is kept fairly constant throughout the neutron drip region. It is the neutron number density and not the electron number density that is rising with increasing matter density. The suppression in electron number density with increasing density has kept the system pressure relatively constant in the density range between 4×10^{11}–10^{12} g/cm^3. This is shown in Fig. 6. Eventually, enough neutrons are produced, and the system pressure is taken over entirely by the neutron pressure. Then the equation of state shows a rapid rise subsequent to 10^{12} g/cm^3.

The numerical values of this portion of the equation of state are listed in Table III under the heading BPS. They are computed by means of the so-called Reid soft-core potential, which is determined phenomenologically by fitting nucleon–nucleon scattering data at energies below 300 MeV, as well as the properties of the deuteron. It is considered one of the best realistic nuclear potentials applicable to nuclear problems. The computation is done in the elaborate pair approximation which includes correlation effects between pairs of nucleons. In practice, this usually means solving the Brueckner–Bethe–Goldstone equations for the nucleon energy of each quantum state occupied by the nucleons. The summation of the nucleon energies for all occupied states yields the energy density of the system. Computations with the realistic nuclear potential are very involved, and several additional corrections are needed to achieve agreement with empirical results. Extension of the method to include finite temperature calculations has not been attempted.

A second form of approach is described as the independent particle approximation, in which case all particles are uncorrelated and move in the system without experiencing the presence of the others except through an overall nuclear potential. The success of the method depends on the adequacy of the effective potential that is prescribed for interaction between each pair of nucleons. A large variety of nuclear effective potentials has been devised. Potentials are usually expressed in functional forms depending on the separation between the interacting pair of nucleons. These potentials depend not only on the particle distance but also on their spin orientations. Some even prescribe dependence on the relative velocity between the nucleons. One effective nuclear potential deserving special mention is the Skyrme potential which, like the Reid soft-core potential, belongs to the class of velocity-dependent potentials. Computations based on effective nuclear potentials seem to constitute the only viable method in dealing with the neutron drip problem at finite temperatures. Some results of finite temperature equations of state in the neutron drip region have been obtained by means of the Skyrme potential. The difficulty in working with a neutron drip system lies in the treatment of phase equilibrium, which must be handled with delicate care. The nuclear liquid drop model serves to reduce much of that work to detailed algebraic manipulations. Still, quantitative results of finite-temperature equations of state in the neutron drip region are scarce.

D. Neutron Matter Region

As matter density increases towards 10^{12} g/cm^3, the constituent nuclei become so large and their Z/A ratio so low that they become merged with the neutron sea at a density of approximately 2×10^{14} g/cm^3, where the phenomenon of neutron drip terminates, and the ground state of the matter system is represented by nearly pure neutron matter. Since neutron matter is so similar to nuclear matter, all successful theories that describe nuclear matter properties have been applied to predict the properties of neutron matter. The ground-state equation of neutron matter that is determined by pair approximation is listed in Table III under BPS. It terminates at a density of 5×10^{14} g/cm^3, which is the upper limit of its applicability. There are also results obtained by the constrained variational method and the independent particle method with the Skyrme potential. These differential equations of state are compared in Fig. 7. It gives us some idea as to how unsettled the issue remains at the present time.

The numerical results of the ground-state equation of state obtained by the constrained variational method are listed in Table III under the heading FP. They are evaluated from a form of realistic nuclear potential by a method that solves the nuclear many-body problem by means of a variational technique. The finite temperature equation of state evaluated by this method is also available.

The numerical results of the ground-state equation of state evaluated from the Skyrme potential are listed in Table III under the heading SKM. Being an effective potential, the Skyrme potential contains adjustable parameters that are established by fitting nuclear properties. As the potential is tried out by different investigators, new sets of potential parameters are being proposed. The results given here are based on a recent set of parameters designated as SKM in the literature.

There are also attempts to formulate the nuclear interaction problem in a relativistic formalism. This seems to be quite necessary if the method is to be applicable to matter density in the 10^{15} g/cm^3 region. A simplified nuclear interaction model based on the relativistic quantum field theory seems quite attractive. It is called the *relativistic mean field model*. In this model, nuclear interaction is described by the exchange of mesons. In its simplest form,

only two types of mesons, a scalar meson and a vector meson, both electrically neutral, are called upon to describe nuclear interaction at short range. The scalar meson is called the σ-meson and the vector meson the ω-meson. Employed on a phenomenological level, the model requires only two adjustable parameters to complete its description, and when it is applied to study the nuclear matter problem it successfully predicts nuclear saturation. These two parameters may then be adjusted so that saturation occurs at the right density and with the correct binding energy. The model is then completely prescribed and may then be extended to study the neutron matter problem. The equation of state thus obtained for neutron matter at zero temperature is plotted in Fig. 7, together with others for comparison. The numerical results are listed in Table III under the heading RMF.

The relativistic mean field model shows great promise as a viable model for dense matter study. In the present form it has some defects. For example, it over-predicts the value of the compression modulus of nuclear matter. Also, the inclusion of a few other types of mesons may be needed to improve on the description of nuclear interaction. In particular, a charge vector meson that plays no role in the nuclear matter system contributes nevertheless to the neutron matter system and should be included. Since empirical results on dense matter are limited, it is difficult to fix the added parameters due to the new mesons in a phenomenological way. Future improvements of this model may have to be based on new results coming from experimental heavy ion collision results.

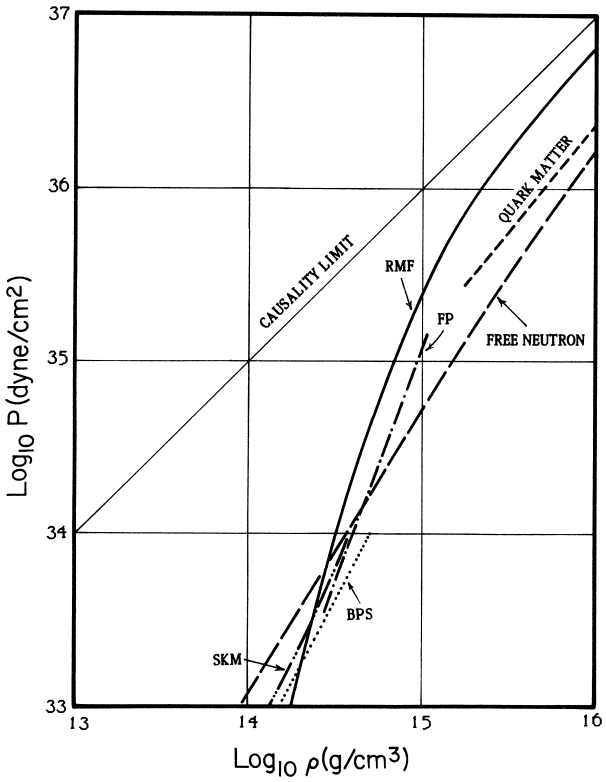


FIGURE 7 Different versions of the equation of state for ground-state neutron matter. Numerical values of curves indicated by BPS, FP, SKM, and RMF are listed in Table III under the same headings. See text for discussion of free neutron, quark matter, and causality limit.

The diagonal line labeled "causality limit" in Figure 7 represents an equation of state given by $P = c\rho$. Any substance whose equation of state extends above this line would give rise to a sound speed exceeding the speed of light, which is not allowed because it would violate causality, thus the causality limit defines an upper limit to all equations of state. It is drawn here to guide the eye.

The pressure of neutron matter rises rapidly with increasing density. Such an equation of state is called *stiff*. This feature is already exhibited by a free neutron system that is devoid of interaction. The equation of state of a free neutron system is plotted alongside the others in Fig. 7 for comparison and is labeled "free neutron." The stiffness of the free neutron equation of state had led scientists in the 1930s to suggest the possible existence of neutron stars in the stellar systems almost as soon as neutrons were discovered. Neutron stars were detected a quarter of century later.

E. Pion Condensation Region

The stiffness of the equation of state of neutron matter could be greatly modified should pions be found to form a condensate in neutron matter. A pion condensate occurs in a matter system only if the interaction between the nucleons and pions lowers the interaction energy sufficiently to account for the added mass and kinetic energy of the pions. The possibility of pion condensation has been suspected for a long time since pion–nucleon scattering experiments have revealed a strong attractive interaction between pions and nucleons, and this attractive mode could produce the pion condensation phenomenon. However, a quantitatively reliable solution of this problem is yet to be established. Nuclear interactions are difficult to deal with, and most theoretical results are not considered trustworthy unless they can be collaborated by some empirical facts. In the pion condensation problem, no empirical verification is available, and many of the estimates about pion condensation, though plausible, cannot be easily accepted.

The effect of pion condensation on the equation of state of neutron matter is to lower the pressure of the system, because it adds mass to the system without contributing the comparable amount of kinetic energy. The softening of the equation of state is a reflection of the presence of an attractive interaction that brings about pion condensation

and serves to lower the pressure. Some estimates have suggested that pion condensation should occur at a neutron matter density of around 5×10^{15} g/cm^3. No quantitative results, however, will be presented here for the pion condensation situation.

F. Baryonic Matter

As the density of neutron matter increases, the Fermi energy of the neutron system increases accordingly. If we neglect interactions among the particles, then as soon as the neutron Fermi energy reaches the energy corresponding to the mass difference of the neutron and Λ, the ground state of the matter system favors the replacement of the high-energy neutrons by Λ. Λ is the least massive baryon heavier than the nucleons. Based on an estimate in which interactions are neglected, Λ should appear in neutron matter at a density of 1.6×10^{15} g/cm^3. On a scale of ascending mass, the next group of baryons after the nucleons and Λ is the Σ, which consists of three charged species, and after that is the Δ, which consists of four charged species. Similar estimates put the appearance of Σ at 2.9×10^{15} g/cm^3 and the appearance of Δ at 3.7×10^{15} g/cm^3.

The above estimates emphasize the mass of the baryons. It turns out that the electric charge of the baryon is also important. For example, when a Σ^- appears, not only does it replace an energetic neutron but, because of its charge, it replaces an energetic electron as well. Again, based on estimates in which interactions are neglected, Σ^- appears at a density of 10^{15} g/cm^3 at Δ^- at 1.3×10^{15} g/cm^3, which are lower values than those given above.

To study baryonic matter properly, particle interactions must be included. Unfortunately, our knowledge of baryon interaction is still quite limited, and the interaction strengths among the different species of baryons are not yet established with sufficient accuracy. Therefore, much of our understanding of the baryonic matter is based on theoretical conjectures of the nature of baryon interaction.

One of the best methods appropriate to the study of baryonic matter is the relativistic mean field model mentioned before. The model considers all particle interactions to the mediated by two mesons, the scalar σ-meson and the vector ω-meson, both electrically neutral. Particle interaction is established once the coupling constants of these mesons with different baryons are known. There are theoretical justifications to believe that the ω-meson interacts with all baryons at equal strength, in the same manner that the electric field interacts with all electric charges equally no matter which particles carry them. In analogy with the electric charge, each baryon is associated with a unit of baryonic charge and its antiparticle with a negative unit of baryonic charge. The baryonic force mediated by the ω-meson behaves very much like the electric force in the sense that like charges repel and unlike charges attract. It is repulsive between baryons but attractive between baryons and antibaryons. If this is the case, the coupling constants of the ω-meson with all baryons in the relativistic mean field model may be taken to be the same as that adopted for the nucleons. One possible fallacy in this reasoning is that the coupling constants in the relativistic mean field model represent effective coupling constants that incorporate modifications due to the presence of neighboring particles. Therefore, even though the concept of a baryonic charge to which the ω-meson couples is correct, the effective ω-meson coupling constants with different baryons in the relativistic mean field model need not be the same.

For the σ-meson, the best estimate of its coupling constants is probably by means of the quark model which assigns definite quark contents to the baryons and σ-meson. If the quarks are assumed to interact equally among themselves, and given the fact that particle attributes such as electric charge and hypercharge must be conserved in the interaction process, it is possible to deduce that the σ-meson couples equally between nucleons and Δ, and also equally between Λ and Σ, but its coupling with Λ and Σ is only two-thirds of the coupling with the nucleon and Δ.

Incorporating these couplings in the relativistic mean field model, we can deduce the equation of state for baryonic matter. It shows a slight decrease in pressure compared with that due to neutron matter when matter density exceeds 10^{15} g/cm^3. The effect of admixing other baryons in a neutron matter system is relatively minor on the system's equation of state. On the other hand, the effect of particle interaction on the equation of state is quite significant. Therefore, a proper understanding of the equation of state of matter in the 10^{15} g/cm^3 range awaits better knowledge of particle interactions as well as methods in dealing with a many-body system.

G. Quark Matter

In a quark model of the elementary particles, the nucleon is viewed as the bound state of three quarks, which interact via the exchange of gluons as determined by quantum chromodynamics. The size of the nucleon is therefore determined by the confining radius of the quark interaction. When matter density reaches the point that the average separation of the nucleons is less than its confining radius, the individual identity of the nucleon is lost, and the matter system turns into a uniform system of quarks forming quark matter.

Quarks are fermions and like electrons and neutrons obey Pauli's exclusion principle. There are different species of quarks, and the known quarks are given the names of up quark (u), down quark (d), strange quark (s), and charm quark (c). There may be others. Inside

a nucleon, their effective masses are estimated to be (expressed in mc^2) under 100 MeV for the u and d quarks, approximately 100–200 MeV for the s quark, and approximately 2000 MeV for the c quark. If quark interaction is neglected, the equation of state of quark matter may be deduced in the same manner as it is for a degenerate electron system. The treatment of quark interaction turns out to be not as difficult as it is for neutrons. In a quark matter, the effective quark interaction is relatively weak, and perturbative treatment of the interaction is possible. This interesting feature of quantum chromodynamics prescribes that the effective interaction of the quarks should decrease in strength as the quarks interact in close proximity to each other, and this has been verified to be true in experiments. The equation of state of quark matter at zero temperature computed with quark interaction has been obtained. One version of it is plotted in Fig. 7 for illustration.

The transition from baryonic matter to quark matter is a first-order phase transition that may be established by a tangent construction method on the energy density of the baryonic matter and that of the quark matter in a manner similar to that discussed in Section III for the neutronization process and illustrated in Fig. 4. The onset of transition from neutron matter to quark matter has been estimated to begin at densities around 1.5 to 4×10^{15} g/cm^3.

In Table III, only the basic ground-state equation of state of dense matter is presented. It serves to suggest the possible behavior of matter compressed to high densities. It would be too elaborate to detail all aspects of the equation of state at finite temperatures. At the present time, the study of dense matter is being actively pursued.

V. TRANSPORT PROPERTIES OF DENSE MATTER

Properties of dense matter under nonequilibrium conditions, such as during the transfer of mass and conduction of heat and electricity, are of much physical interest. These are the transport properties, and knowledge of them is important in understanding stellar structure and stellar evolution. We discuss here the following transport properties of dense matter: electrical conductivity, heat conductivity, and shear viscosity.

A. Electrical Conductivity

The electrical conductivity of a substance is given by the ratio of the induced electric current density to the applied electric field. In the case of a metal, the electric current is due to the flow of conduction electrons. If all conduction electrons are given an average drift velocity v_d, then the current density j_e is given by:

$$j_e = nev_d \tag{45}$$

where n is the number density of the conduction electrons and e the electric charge of the electrons. The electrons acquire the drift velocity as they are accelerated by the electric field. When an electron collides with either the lattice or another electron, its drift velocity is redirected randomly, and the subsequent velocity is averaged to zero statistically. Thus, after each collision the electron may be assumed to be restored to thermal equilibrium and begins anew with zero drift velocity. The average drift velocity is therefore determined by the mean time between collisions τ, during which an electron is accelerated to its drift velocity by the electric field $\mathscr{E}$:

$$v_d = (e\mathscr{E}/m_e)\tau \tag{46}$$

τ is also called the relaxation time. Putting these two expressions together gives the electrical conductivity:

$$\sigma = \left(ne^2/m_e\right)\tau \tag{47}$$

The electrical conductivity is in units of inverse seconds, or sec^{-1}, in the cgs system of units and is related to the mks system of units by sec$^{-1} = (9 \times 10^9\ \Omega\ \text{meter})^{-1}$. When the electrons in the system are degenerate, the electron mass in Eq. (47) is replaced by p_F/c, where p_F is the Fermi momentum of the degenerate electron system. The relaxation time is the most crucial parameter in this investigation and must be related to the electron density, the average electron speed, the number and types of scatterers in the system, and the scattering cross sections of the electrons with different types of scatterers.

When the temperature is below the melting temperature T_m given by Eq. (41), matter in the density range of 10^2–10^{14} g/cm^3 is in a solid state possessing a crystalline structure. The constituent nuclei are organized into a lattice while the electrons are distributed more or less uniformly in the space between. In the neutron drip region, 10^{12}–10^{15} g/cm^3, neutrons are also outside the nuclei. The relaxation time of the electrons is determined by the frequency of scatterings with the lattice nuclei and with the other electrons and neutrons. When there are several scattering mechanisms present, the relaxation times due to different mechanisms are found by the inverse sum of their reciprocals:

$$\tau^{-1} = \tau_1^{-1} + \tau_2^{-1} + \cdots \tag{48}$$

Upon electron scattering, the lattice vibrates and the vibration propagates collectively like sound waves through the lattice. In quantized form the sound waves behave like particles, called *phonons*. Electron scattering from the lattice is usually described as electron–phonon scattering. Phonons increase rapidly in number with temperature. We therefore expect the electric conductivity due

to electron–phonon scattering to decrease with increasing temperature.

Electron–electron scatterings have minor effects on the electric conductivity, because in dense matter all the electrons are not bound to the nuclei. Electrons lose energy only if they are scattered from bound electrons. Elastic scattering of electrons does not alter the current being transported. Hence, this form of scattering does not affect electric conductivity and may be ignored.

Let us imagine that the dense matter in consideration is formed from a dynamical process, as in the formation of a neutron star, in which case the material substance is adjusting to reach the proper density while it is being cooled. If the solidification rate is faster than the nuclear equilibrium rate, there will be large admixtures of other nuclei with the equilibrium nuclei. The nonequilibrium nuclei act as impurities in the system. Also, there may be defects in the system due to rapid rates of cooling and rotation. Electrons and phonons are scattered by these impurities and defects which also limit the transport of electric charge by electrons.

There are estimates of the electric conductivity in dense matter that consider all the above discussed features. Typical results for a wide range of densities at a temperature of 10^8 K are shown in Fig. 8. Electric conductivities due to different scattering mechanisms are shown separately. The total conductivity should be determined by the lowest lying portions of the curves, since the total conductivity, like the relaxation time, is found by taking the inverse sum of the reciprocals of the conductivities due to different scattering mechanisms.

Electron–phonon scattering should be the dominant mechanism. The conductivity due to electron–phonon scattering has a temperature dependence of approximately T^{-1}. Electron–impurities scattering depends on the impurity concentration and the square of the charge difference between the impurity charge and the equilibrium nuclide charge. Let us define a concentration factor,

$$x_{\text{imp}} = \sum \left(\frac{n_i}{n_{\text{A}}}\right)(Z_i - Z)^2 \qquad (49)$$

where the summation is over all impurity species designated by subscript i; n_i and Z_i are, respectively, the number density and charge of each impurity species. The conductivity curve due to impurity scattering σ_{imp} is drawn in Fig. 8 with an assumed $x_{\text{imp}} = 0.01$. It is independent of T.

The electron–phonon curve would not be correct at the low-density end where the melting temperature is below 10^8 K. Above melting temperature, the lattice structure would not be there, and electrons would not be impeded by scattering from phonons but instead from the nuclei. Electric conductivity due to electron–nuclei scattering is shown in the possible melting region.

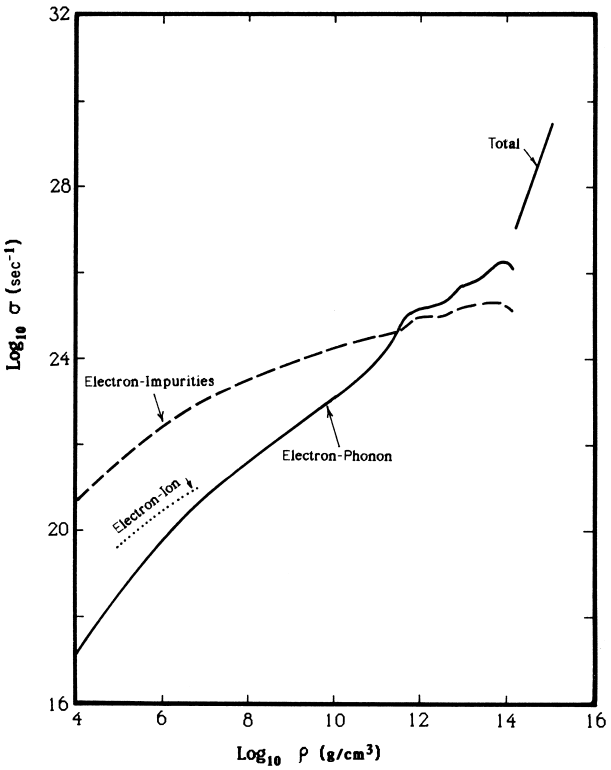


FIGURE 8 The electrical conductivity of dense matter at a temperature of $T = 10^8$ K. For densities below 2×10^{12} g/cm^3, the solid line corresponds to the electrical conductivity limited by the mechanism of electron–phonon scattering while the system is in a solid state. Possible melting occurs in the low-density region at this temperature, and electron–phonon scattering is replaced by electron-ion scattering. The electrical conductivity due to electron–ion scattering is drawn as a dotted line. The electrical conductivity due to electron–impurities scattering is drawn as a dashed line, taking an arbitrary impurity concentration factor of $x_{\text{imp}} = 0.01$. For densities above 2×10^{12} g/cm^3, the electrical conductivity is limited mainly by the mechanism of electron–proton scattering, and it is drawn as a solid line and labeled "total." [From Elliott, F., and Itoh, N., *Astrophys. J.* **206,** 218, 1976; **230,** 847, 1979. Reprinted with permission of *The Astrophysical Journal*, published by the University of Chicago Press; ©1976 and 1979. The American Astronomical Society.]

In the neutron matter region, 10^{14}–10^{15} g/cm^3, the system is composed basically of three types of particles—neutrons, protons, and electrons—of which both protons and electrons act as carriers for electrical conduction. The main difficulty in dealing with this situation is to take into full account the interactions among the particles. With the presence of a strong attractive interaction among the nucleons, it is quite likely that the protons will be paired to turn the system into a superconducting state when the temperature falls below the critical temperature, in which case the electric conductivity becomes infinite. The critical temperature T_{c} is estimated to be as high as 10^9–10^{10} K. Such a temperature, though it appears high, corresponds

actually to a thermal energy $k_B T$ that is small compared with the Fermi energies of the particles present.

If the system is not in a superconducting state, or in other words it is in a normal state, its electrical conductivity is due mainly to electrons as carriers. The electron relaxation time is determined by scattering by protons and has been evaluated. The electrical conductivity of neutron matter in the normal state is plotted in Fig. 8 in the density range 10^{14}–10^{15} g/cm^3.

B. Thermal Conductivity

The thermal conductivity of a substance is given by the ratio of the amount of heat transferred per unit area per unit time to the temperature gradient. Kinetic theory of dilute gas yields the following expression for thermal conductivity:

$$k = (1/3) n c_s \bar{v}^2 \tau \tag{50}$$

where n is the carrier number density, c_s its specific heat, $\bar{v}$ the average thermal velocity, and τ the relaxation time of the carriers. The thermal conductivity is expressed in units of erg/cm K sec. When there are several types of carriers participating in the transport of heat, the final thermal conductivity is the sum of individual conductivities due to different types of carriers. In the case of a solid, the important carriers are the electrons and phonons, but in general the phonon contribution to thermal conduction is negligible compared with the electron contribution.

In thermal conduction by electrons, no electric current is generated while energy is being transported. The type of electron–phonon scattering important for thermal conduction is different from that of electrical conduction. At each point in the system, the electron number density obeys the Fermi–Dirac distribution of Eq. (32), which assigns a higher probability of occupation of high-energy states when the temperature is high than when the temperature is low. When a thermal gradient exists in the system, neighboring points have different electron distributions. Electrons moving from a high-temperature point to a low-temperature point must lose some of their energy to satisfy the distribution requirement. If this can be accomplished over a short distance, the thermal resistivity of the substance is high, or its thermal conductivity is low. The most important mechanism of energy loss is through inelastic scatterings of electrons by phonons at small angles. Such scatterings constitute the major source of thermal resistivity. On the other hand, elastic scattering of electrons do not lead to energy loss and aid in thermal conduction. The frequency of inelastic scattering to that of elastic scattering depends on the thermal distribution of phonons. The temperature dependence of thermal conductivity due to electron–phonon scatterings is therefore complicated. For matter density below 10^8 g/cm^3, it is relatively independent of temperature, but above that it shows a decrease with increasing temperature.

Although electron–electron scatterings do not contribute to the electric conductivity, they contribute to the thermal conductivity by redistributing the electron energies. The thermal conductivity due to impurity scattering k_{imp} is directly related to the electric conductivity due to impurity scattering σ_{imp} by the Wiedemann–Franz rule:

$$k_{imp} = \pi^2 T / 3 (k_B/e)^2 \sigma_{imp} \tag{51}$$

The thermal conductivities due to different mechanisms at 10^8 K are shown in Fig. 9. The thermal conductivity due

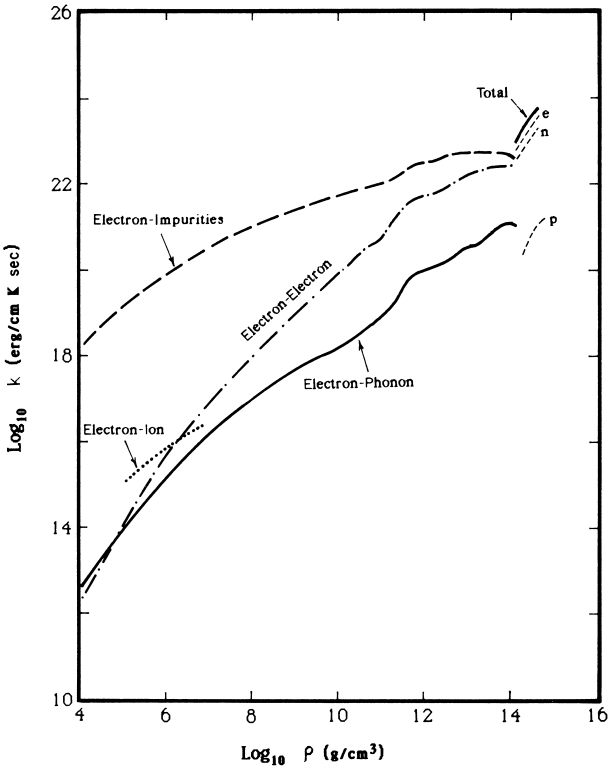


FIGURE 9 The thermal conductivity of dense matter at a temperature of $t = 10^8$ K. For densities below 2×10^{12} g/cm^3, the solid line corresponds to the thermal conductivity limited by the mechanism of electron–phonon scattering. If the lattice is melted, the solid line should be replaced by the dotted line, which is the thermal conductivity due to electron–ion scattering. The thermal conductivity due to electron–electron scattering is given by the dot-dashed line, and that due to electron–impurities scattering is given by the dashed line, where the impurity concentration factor is assumed to be $x_{imp} = 0.01$. For densities above 2×10^{12} g/cm^3, the thermal conductivity received contributions from the electrons, neutrons, and protons, which are drawn in thin dashed lines and marked by e, n, and p, respectively. The total thermal conductivity from these carriers is drawn in a solid line and labeled "total." [From Elliott, F., and Itoh, N., *Astrophys. J.* **206,** 218, 1976; **230,** 847, 1979. Reprinted with permission of *The Astrophysical Journal*, published by the University of Chicago Press; © 1976 and 1979. The American Astronomical Society.]

to impurity scattering is drawn with $x_{\rm imp} = 0.01$, and it has a temperature dependence linear in T. The thermal conductivity due to electron–electron scattering is expected to have a temperature dependence of T^{-1}. The thermal conductivity due to electron–ion scattering is shown for the region where 10^8 K is expected to be above the melting temperature. The total thermal conductivity of dense matter is determined by the mechanism that yields the lowest thermal conductivity at that density, since the total thermal conductivity, like the relaxation time, is found from the inverse sum of the reciprocals of the conductivities due to different mechanisms.

In the neutron matter region, 10^{14}–10^{15} g/cm^3, all three types of particles—neutrons, protons, and electrons—contribute to the thermal conductivity

$$k = k_{\rm e} + k_{\rm n} + k_{\rm p}$$

where the subscripts e, n, and p denote contributions to the thermal conductivity by electrons, neutrons, and protons, respectively. When the particles are in a normal stare (i.e., not in a superfluid or superconducting state), the thermal conductivity is determined primarily by the highly mobile electrons, whose motion is impeded largely by scatterings with the protons and other electrons and much less by scatterings with the neutrons. The neutron contribution to the thermal conductivity is substantial because of its high number density. Neutrons encounter neutron–proton scattering and neutron–neutron scattering in the process. The proton contribution to the thermal conductivity is small because the proton number density is low, but otherwise the protons contribute in a manner similar to the neutrons. The thermal conductivities due to these three components for a system in the normal state are shown in short dashed lines in Fig. 9. The total conductivity is drawn as a solid line there.

The system may also become superfluid when its temperature falls below the critical temperature of $T_{\rm c} \approx 10^9$–10^{10} K. The critical temperature of the protons is in general different from that of the neutrons, and therefore it is possible that while one turns superfluid, the other remains normal. Also, when the temperature falls below the critical temperature of a certain type of particles (say, the neutrons), there remains a normal component of neutrons in the system. This situation is usually described by a two-fluid model that consists of both the superfluid and normal fluid components. In general, scattering of particles off the superfluid component is negligible for transport purposes.

If only the protons turn superfluid (and superconducting) while the neutrons remain normal, the thermal conductivity due to the superfluid protons vanishes. The thermal conductivity found for the system in the normal state is basically unaltered, because the protons give a very small contribution to the thermal conductivity, as shown in Fig. 9.

If the neutrons turn superfluid while the protons remain normal, the superfluid component of the neutrons gives vanishing thermal conductivity, and the contribution by the normal component of the neutrons to the thermal conductivity diminishes rapidly with decreasing temperature below the critical temperature, because the number density of the normal neutrons decreases rapidly with decreasing temperature. The thermal conductivity in this case is therefore determined entirely by the electron contribution to the thermal conductivity, which is modified slightly from the normal matter case due to the absence of scattering by superfluid. The net result is that the thermal conductivity is only slightly reduced from the normal case shown in Fig. 9. However, if both the neutrons and protons turn superfluid, the thermal conductivity is due entirely to electron–electron scattering, and the general result is indicated by the extension of the electron–electron curve for densities below 10^{14} g/cm^3.

C. Shear Viscosity

When a velocity gradient exists in a fluid, a shearing stress is developed between two layers of fluid with differential velocities. The shear viscosity is given by the ratio of the shearing stress to the transverse velocity gradient. Elementary kinetic theory suggests that the shear viscosity of a dilute gas is given by:

$$\eta = \frac{1}{3} n m \bar{v}^2 \tau \tag{52}$$

where n is the molecular density, m the mass of each molecule, $\bar{v}$ the average thermal velocity of the molecules, and τ the relaxation time. Viscosity is expressed in units of g/cm/sec, which is also called *poise*. In appearance it is similar to the expression for thermal conductivity with the exception that the specific heat per particle in the thermal conductivity is being replaced by the particle mass. Consequently, we may suspect that the electron component of the viscosity should behave similarly to the thermal conductivity. This, however, is not true due to the fact that the relaxation time involved relates to different aspects of the scattering mechanism. Also, there is an additional component to the viscosity. The solid lattice can make a great contribution to the total viscosity. Unfortunately, the determination of the lattice viscosity is very difficult, and no adequate work has been performed to determine its properties at the present time. The following discussion relates only to the electron viscosity.

The electron relaxation time is determined by electron scatterings by phonons, impurities, electrons, and nuclei. Shearing stress is developed when electrons belonging to

fluid layers of different velocities are exchanged. Thus, viscosity is related to mass transfer or the transfer of electrons. This is similar to electric conduction where the transfer of electrons gives rise to charge transfer and is different from heat conduction which involves the adjustment of electron energy distributions. The evaluation of the relaxation times for the viscosity due to different scattering mechanisms is similar to that for the electrical conductivity.

The viscosity of dense matter at 10^8 K due to different scattering mechanisms is shown in the Fig. 10. The temperature dependence of the viscosity due to electron–phonon scattering is approximately T^{-1}, as in the case of the electrical conductivity. This is also true of the viscosity due to electron-impurities scattering, which is independent of temperature as in the case of the electrical conductivity. While electron–electron scatterings do not contribute to the electrical conductivity, they play a role in viscosity giving rise to a temperature dependence of T^{-2}.

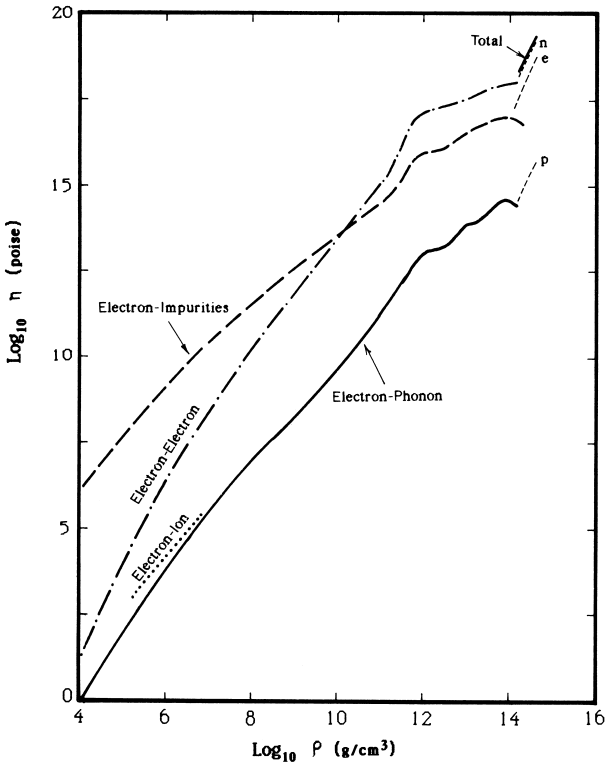


FIGURE 10 The viscosity of dense matter at a temperature of $T = 10^8$ K. For densities below 2×10^{12} g/cm^3, the viscosity due to electron–phonon scattering is drawn as a solid line. If the lattice is melted, the viscosity is due to electron–ion scattering, which is drawn as a dotted line. The viscosity due to electron–electron scattering is drawn as a dot-dashed line, and the viscosity due to electron-impurities scattering is drawn as a dashed line, where the impurity concentration factor is taken to be $x_{\text{imp}} = 0.01$. For densities above 2×10^{12} g/cm^3, the viscosity receives contributions from neutrons, electrons, and protons, and they are shown as thin dashed lines and labeled n, e, and p, respectively. The total viscosity from these three components is drawn as a solid line and labeled "total." [From Elliott, F., and Itoh, N., *Astrophys. J.* **206,** 218, 1976; **230,** 847, 1979. Reprinted with permission of *The Astrophysical Journal*, published by the University of Chicago Press; ©1976 and 1979 The American Astronomical Society.]

In the neutron matter region, 10^{14}–10^{15} g/cm^3, all three types of particles—neutrons, protons, and electrons—contribute to the viscosity. The contributions from neutrons, protons, and electrons to the viscosity are shown separately in Fig. 10 in this density range by dashed lines. The total viscosity is drawn as a solid line. They all have a temperature dependence of T^{-2}.

When the temperature drops below the critical temperature T_c, superfluid proton and/or neutron components appear. The behavior of the viscosity in the superfluid state is very similar to the thermal conductivity. If the protons turn superfluid, the viscosity is basically unaltered from the normal viscosity, because the proton contribution is small. If the neutrons turn superfluid, the superfluid component of the neutrons has vanishing viscosity. The viscosity is dominated by the electron contribution which is determined by the electron–electron scattering and electron-proton scattering mechanisms. When both protons and neutrons turn superfluid, then the viscosity is determined entirely by electron–electron scattering, and the general result is indicated by the extension of the electron–electron curve for densities below 10^{14} g/cm^3.

VI. NEUTRINO EMISSIVITY AND OPACITY

In a nonequilibrium situation where a temperature gradient exists in a substance, energy is transported not only by means of thermal conduction, as discussed in the last section, but also by radiation. Parameters of radiative transfer intrinsic to the matter system are its emissivity and opacity. There are two major forms of radiation. In one form, the radiative energy is transmitted by photons and in the other by neutrions.

In a dense matter system whose constituent electrons are highly degenerate (i.e., when the electron Fermi energy is high compared with the thermal energy $k_B T$), the degenerate electrons cannot avail themselves as effective scatterers for the passage of energy carriers created by the thermal gradient, and the thermal conductivity is correspondingly high. Therefore, energy transport is far more effective through heat conduction than it would be for radiative transfer. For this reason, the problem of photon emissivity and opacity in dense matter receives very little attention. Most astrophysical studies of photon emissivity and opacity are performed for relatively low-density

substances, such as those forming the interiors of luminous stars. Since our subject is dense matter, these parameters will not be discussed here. Interested readers are referred to astrophysical texts suggested in the bibliography. More relevant to dense matter physics is the topic of neutrino emissivity and opacity.

A. Neutrino Emissivity

The reaction described by Eq. (37) is a sample reaction in which a neutrino is produced. Indeed, neutrinos are produced throughout the second density domain whenever neutronization occurs, and in these processes neutrinos are produced even at zero temperature. At high temperature there are other reactions that are more effective in neutrino production. Reactions involving neutrinos belong to a class of interaction called the *weak interaction*. The coupling constant for the weak interaction is much smaller than that of the electromagnetic interaction. Neutrino interaction rates are in general slower than comparable photon interaction rates by a factor of at least 10^{20}. Neutrinos can pass through thick layers of substance and experience no interaction. For example, the neutrino mean free path through matter with density similar to our sun is about a billion solar radii. Hence, once they are produced in a star they are lost into space, and they serve as an efficient cooling mechanism for hot and dense stellar objects.

During a supernova process, the collapse of the stellar core raises the core temperature to as high as 10^{11} K. It quickly cools to a temperature of 10^9 K through neutrino emission as the core stabilizes into a neutron star. Neutrino emission dominates photon emission until the temperature drops to 10^8 K. It is estimated that neutrino cooling dominates for at least the first few thousand years of a neutron star after its formation; therefore, the neutron cooling mechanisms deserve attention. We are primarily interested in those neutrino emission mechanisms that supersede photon emission under similar conditions. They determine therefore the cooling rate of neutron stars in the early period, and they also play a major role in the dynamics of a supernova process.

An important neutrino production process is the modified Urca reactions which involve neutrons, protons, and electrons:

$$\mathrm{n} + \mathrm{n} \rightarrow \mathrm{n} + \mathrm{p} + \mathrm{e} + \bar{\nu} \tag{53a}$$

$$\mathrm{n} + \mathrm{p} + \mathrm{e} \rightarrow \mathrm{n} + \mathrm{n} + \nu \tag{53b}$$

where ν and $\bar{\nu}$ denote neutrino and antineutrino, respectively. In the following we shall not distinguish neutrinos from antineutrinos and address them collectively as neutrinos. These two reactions are very nearly the inverse of each other, and when they are occurring at equal rates, the number of neutrons, protons, and electrons in the system is unaltered, while neutrinos are being produced continuously. Urca is the name of a casino in Rio de Janeiro. Early pioneers of neutrino physics saw a parallel between nature's way of extracting energy from the stellar systems and the casino's way of extracting money from its customers, so they named the reactions after the casino. Reactions (53a,b) are modifications of the original Urca reactions by adding an extra neutron to the reaction. This increases the energy range over which neutrinos may be produced and thus improves the production rate.

Our current understanding of the weak interaction theory is provided by the Weinberg–Salam–Glashow theory. Even though most of the neutrino reactions to be discussed in this section have never been verified under laboratory conditions, they are nevertheless believed to be correct, and quantitative estimates of their reaction rates are reliable. When reactions (53a) occurs, a transition is made from a state of two neutrons to a state of a neutron, proton, electron, and neutrino. The theory evaluates the transition probability from an initial state of two neutrons that occupy quantum states of definite momenta to a final state of four particles of definite momenta. Whenever a transition is made, a neutrino of a specific energy is produced. The total neutrino energy emitted from the system per unit time is found by summing the neutrino energies of all allowed transitions multiplied by their respective transition probabilities. The neutrino emissivity is the total neutrino energy emitted per unit time per unit volume of the substance.

Consider a neutron matter system in the density range of 10^{14}–10^{15} g/cm^3 that is composed mainly of neutrons with a small admixture (about 1%) of protons and electrons. Quantum states occupied by the particles are given by the Fermi–Dirac distribution f_j of Eq. (32). Note that the temperature dependence of the final result comes from the temperature factor in the distribution. Designating the initial-state neutrons by subscripts 1 and 2, the final-state neutron by 3, and the proton, electron, and neutrino by p, e, and ν, respectively, we find that the emissivity of the neutron matter system is given by the following equation:

$$\begin{aligned} E_{\mathrm{Urca(a)}} = \int & \left[d^3 n_1 f_1\right]\left[d^3 n_2 f_2\right]\left[d^3 n_3 (1 - f_3)\right] \\ & \times \left[d^3 n_{\mathrm{p}} (1 - f_{\mathrm{p}})\right]\left[d^3 n_{\mathrm{e}} (1 - f_{\mathrm{e}})\right] \\ & \times \left[d^3 n (1 - f_\nu)\right] W(i \rightarrow \mathrm{f}) e_\nu \end{aligned} \tag{54}$$

where the summations over quantum states are represented by integrations over $d^3 n_j = V(d^3 p_j / h^3)$, $e_\nu = c|\mathbf{p}_\nu|$ is the neutrino energy, and $W(\mathrm{i} \rightarrow \mathrm{f})$ is the transition probability from initial state i to final state f per unit time and per unit volume of the system. It has the following structure:

$$W(\mathrm{i} \to \mathrm{f}) = (2\pi)^4 \, \delta(\mathbf{p}_\mathrm{i} - \mathbf{p}_\mathrm{f}) \, \delta(E_\mathrm{i} - E_\mathrm{f}) \times \sum |H(\mathrm{i} \to \mathrm{f})|^2 \qquad (55)$$

where $\mathbf{p}_\mathrm{i}$ and $\mathbf{p}_\mathrm{f}$ denote the total momenta of the initial and final states, respectively; E_i and E_f, the total energies of the initial and final states, respectively; and $H(\mathrm{i} \to \mathrm{f})$, the matrix element of the Hamiltonian describing the interaction: The summation symbol $\sum$ indicates the summations over all spin orientations of the particles. The mathematical delta functions are here to ensure that only energy–momentum-conserving initial and final states are included in this evaluation. In Eq. (54), the initial state particles are assigned distributions f_j while the final state particles are assigned distributions $(1 - f_j)$ because the particles produced in the reaction must be excluded from states that are already occupied; therefore, they must take up states that are not occupied, which are expressed by $(1 - f_j)$. The chemical potentials in the distributions f_j determine the particle numbers in the system. They are related by $\mu_1 = \mu_2 = \mu_3$ and $\mu_\mathrm{p} = \mu_\mathrm{e}$, while in most cases $\mu_\nu = 0$. The quantity μ_ν will be different from zero only in circumstance where the neutrino opacity is so high that the neutrinos are trapped momentarily and become partially degenerate.

If the neutrons and protons in the system are assumed to be noninteracting, the emissivity due to be noninteracting, the emissivity due to the reaction (53a) can be easily evaluated, and the result may be expressed conveniently as:

$$E_{\mathrm{Urca(a)}} = (6.1 \times 10^{19} \, \mathrm{erg/cm^3 \, sec})(\rho/\rho_0)^{2/3}(T_9)^8 \qquad (56)$$

where $\rho_0 = 2.8 \times 10^{14}$ g/cm^3 is the density of nuclear matter, and the symbol T_9 stands for temperature in units of 10^9 K. Here, the emissivity (in units of ergs per cubic centimeter per second) is evaluated for a dense system of neutron matter whose proton and electron contents are determined by the ground-state requirement under reaction Eq. (37). It is expressed in this form because neutron matter exists only with densities comparable to nuclear matter density, and T_9 is a typical temperature scale for neutron star and supernova problems.

An interesting point to note is that this emissivity is given by eight powers of temperature. This comes about for the following reason. For the range of temperature considered, the thermal energy is still small compared with the degenerate or Fermi energy ε_F of the fermions in the system (except neutrinos), and most of the easily accessible quantum states are already occupied, leaving only a small fraction of quantum states in each species contributing to the reaction (of the order of $k_\mathrm{B}T/\varepsilon_\mathrm{F}$ per species). Since there are two fermion species in the initial state and three (not counting neutrinos) in the final state, a factor of T^5 is introduced. The allowed neutrino states are restricted only by energy conservation, and their number is given by the integration over $d^3 n_\nu \delta(E_\mathrm{i} - E_\mathrm{f})$, which is proportional to the square of the neutrino energy e_ν^2. This, together with the neutrino energy term in Eq. (54), gives e_ν^3. Since e_ν must be related to $k_\mathrm{B}T$, which is the only energy variable in the problem, all together they give the T^8 dependence to the expression. The above deduction has general applicability, and if there were one less fermion in both the initial and final states, then the emissivity from such a process should be proportional to T^6.

The emissivity due to reaction (53b), $E_{\mathrm{Urca(b)}}$, can be shown to be of comparable magnitude to that evaluated above, and the total emissivity due to the modified Urca process is simply twice that of Eq. (56):

$$E_{\mathrm{Urca}} = (1.2 \times 10^{20} \, \mathrm{erg/cm^3 \, sec})(\rho/\rho_0)^{2/3}(T_9)^8 \qquad (57)$$

At high density, muons would appear alongside electrons, and neutrino production reactions similar to those of Eqs. (53a) and (53b) become operable, but with the electrons there replaced by muons. Similar results are obtained, but muons do not appear in dense matter until its density exceeds 8×10^{14} g/cm^3, and the emissivity due to the additional muon processes is only a minor correction.

A more significant consideration is the inclusion of nuclear interaction, which has been neglected in the above evaluation. Nuclear interaction appears in the evaluation of the matrix element $H(\mathrm{i} \to \mathrm{f})$. When nuclear interaction is included, the total emissivity due to modified Urca process is changed to

$$E_{\mathrm{Urca}} = (7.4 \times 10^{20} \, \mathrm{erg/cm^3 \, sec})(\rho/\rho_0)^{2/3}(T_9)^8 \qquad (58)$$

which is a factor of six higher than that evaluated without the inclusion of nuclear interaction.

Other important neutrino production mechanisms in neutron matter are

$$\mathrm{n} + \mathrm{n} \to \mathrm{n} + \mathrm{n} + \nu + \bar{\nu} \qquad (59a)$$

$$\mathrm{n} + \mathrm{p} \to \mathrm{n} + \mathrm{p} + \nu + \bar{\nu} \qquad (59b)$$

in each of which a pair of neutrinos is produced as the nucleons scatter from each other. Neutrino emissivities for these processes are evaluated to be

$$E_{\mathrm{nn}} = (1.8 \times 10^{19} \, \mathrm{erg/cm^3 \, sec}) \, (\rho/\rho_0)^{1/3}(T_9)^8 \qquad (60a)$$

$$E_{\mathrm{np}} = (2.0 \times 10^{19} \, \mathrm{erg/cm^3 \, sec}) \, (\rho/\rho_0)^{2/3}(T_9)^8 \qquad (60b)$$

where the subscript nn denotes reaction (59a) and np denotes (59b). These emissivities are evaluated with nuclear interaction taken into consideration. The processes are, however, not as effective as the modified Urca processes in neutrino production.

If neutron matter turns superfluid after its temperature falls below the critical temperature T_c, then the neutrino production rates evaluated above must be reduced; T_c

should be in the range 10^9–10^{10} K. Superfluidity is explained by the fact that an energy gap appears in the energy spectrum of the particle, just above its Fermi energy. What that means is that a group of quantum states whose energies fall in the energy gap are excluded from the system. Normally, in an inelastic scattering process, the neutrons or protons are scattered into these states, but since the states are absent they must be excited into much higher energy states, and scattering becomes difficult and less likely to occur. The consequence is that neutrons and protons may move about relatively freely without being impeded by scatterings. This is the explanation of superfluidity. By the same token, the above neutrino production mechanisms that depend on the scattering of neutrons and protons are similarly reduced. Qualitatively, if superfluidity occurs in both the neutron and proton components, $E_{\rm Urca}$ and $E_{\rm np}$ are reduced by a factor of $\exp[-(\Delta_{\rm n}+\Delta_{\rm p})/k_{\rm B}T]$, where $\Delta_{\rm n}$ and $\Delta_{\rm p}$ are the width of the energy gaps for neutron and proton superfluidity, respectively. For $E_{\rm nn}$, the reduction is $\exp[-2\Delta_{\rm n}/k_{\rm B}T]$. If superfluidity occurs in just one component, the reductions are obtained by setting the energy gap of the normal component in the above expressions to zero.

In the case of pion condensation in neutron matter, interactions involving pions and nucleons also modify the neutrino production rate. Some estimates have shown that the neutrino production rates thus modified can be significant. However, due to the uncertainty in our present understanding of the pion condensation problem, no emissivity for this situation will be quoted here.

There are also neutrino production mechanisms not involving the direct interaction of two nucleons, such as the following:

1. Pair annihilation

$$\mathrm{e}+\mathrm{e}^+ \rightarrow \nu+\bar{\nu} \tag{61a}$$

2. Plasmon decay

$$\text{plasmon} \rightarrow \nu+\bar{\nu} \tag{61b}$$

3. Photoannihilation

$$\mathrm{e}+\gamma \rightarrow \mathrm{e}+\nu+\bar{\nu} \tag{61c}$$

4. Bremsstrahlung

$$\mathrm{e}+(Z,A) \rightarrow (Z,A)+\mathrm{e}+\nu+\bar{\nu} \tag{61d}$$

5. Neutronization

$$\mathrm{e}+(Z,A) \rightarrow (Z-1,A)+\nu \tag{61e}$$

where e^+ denotes a positron, plasmon a photon propagating inside a plasma, and (Z, A) a nucleus of atomic number Z and mass number A. A photon in free space cannot decay into a neutrino pair, because energy and momentum cannot be conserved simultaneously in the process. However, when a photon propagates inside a plasma, its relation between energy and momentum is changed in such a way that the decay becomes possible. Quanta of the electromagnetic wave in a plasma are called *plasmons*. The neutrino production rates for these processes have been found to be relatively insignificant at typical neutron star densities and temperatures and will not be listed here.

B. Neutrino Opacity

When a radiation beam of intensity $I(\mathrm{erg/cm^2/sec})$ is incident on a substance of density $\rho(\mathrm{g/cm^2})$, the amount of energy absorbed from the beam per unit volume per unit time $E(\mathrm{erg/cm^3/sec})$ is proportional to the opacity $\kappa = E/\rho I$. The opacity is expressed in units of $\mathrm{cm^2/g}$. Each neutrino emission process has an inverse process corresponding to absorption. In addition to absorption, scattering can also impede the passage of neutrinos through the medium. Both absorption and scattering contribute to the opacity. Some of the more important processes are listed below.

1. Scattering by neutrons

$$\nu+\mathrm{n} \rightarrow \nu+\mathrm{n} \tag{62a}$$

2. Scattering by protons

$$\nu+\mathrm{p} \rightarrow \nu+\mathrm{p} \tag{62b}$$

3. Scattering by electrons

$$\nu+\mathrm{e} \rightarrow \nu+\mathrm{e} \tag{62c}$$

4. Scattering by nuclei

$$\nu+(Z,A) \rightarrow \nu+(Z,A) \tag{62d}$$

5. Absorption by nucleons

$$\nu+\mathrm{n} \rightarrow \mathrm{p}+\mathrm{e} \tag{62e}$$

6. Absorption by nuclei

$$\nu+(Z,A) \rightarrow (Z+1,A)+\mathrm{e} \tag{62f}$$

Similar processes occur for antineutrinos, which shall not be displayed here.

For each of these reactions, a reaction cross section is evaluated from the Weinberg–Salam–Glashow theory of weak interactions. The cross section represents an area effective in obstructing the incident beam of neutrinos. The opacity may be expressed in terms of the reaction cross sections as follows:

$$\kappa = \rho^{-1}\sum n_j\sigma_j \tag{63}$$

where n_j denotes the number density of the particles that react with the neutrino and σ_j their reaction cross sections, and the summation $\sum$ is over all the reactions listed above. Often, neutrino opacity is expressed by a neutrino mean free path, which is defined as:

$$\lambda = (\rho\kappa)^{-1} \tag{64}$$

Among these reactions, the contribution of reaction (62f) to the opacity of dense matter is quite negligible because it entails the production of electrons; since electrons in the system are already highly degenerate, it is difficult to accommodate the newly produced electrons.

The most important reaction in this regard is reaction (62d) where the neutrino is scattered by the nuclei. This is the result of coherent scattering, in which all nucleons in a nucleus participate as a single entity in the process. The cross section of a coherent process involving A nucleons is proportional to A^2 times the cross section of scattering from a single nucleon. This reaction therefore dominates all others when the matter system is composed of giant nuclei, which is the case when matter density is below nuclear density, that is, before nuclei dissolve into neutron matter.

For neutron matter the important reactions are scatterings by neutrons, protons, and electrons as indicated by reactions (62a), (62b), and (62c). Neutrinos are scattered elastically by the nucleons and nuclei since the scatterers are massive. The neutrinos may change directions after scattering but do not lose their energies to the scatterers. They lose energy only if they are scattered by electrons. Electron scattering is therefore an important process in lowering the energies of the high-energy neutrinos, bringing them into thermal equilibrium with all neutrinos should the neutrinos be trapped in the system for a duration long enough for this to happen. Even though neutrinos interact very weakly and are therefore very difficult to confine, neutrino trapping is in fact believed to occur at the moment when the collapsing stellar core reaches the point of rebound initiating the explosive supernova process. Therefore, a great deal of attention has been given to the problem of neutrino opacity and the issue of neutrino trapping.

The cross sections for these processes in the reference frame of the matter system are evaluated to be as follows:

1. Neutrino–electron scattering

$$\sigma_{\mathrm{e}} \approx (1/4)\sigma_0\left(e_\nu / m_{\mathrm{e}}c^2\right)(\varepsilon_{\mathrm{f}}/mc), \qquad e_\nu \ll \varepsilon_{\mathrm{F}} \tag{65}$$

2. Neutrino–nucleon scattering

$$\sigma_{\mathrm{N}} \approx (1/4)\sigma_0\left(e_\nu / m_{\mathrm{e}}c^2\right)^2, \qquad e_\nu \ll m_{\mathrm{n}}c^2 \tag{66}$$

3. Neutrino–nucleus scattering

$$\sigma_{\mathrm{A}} \approx (1/16)\sigma_0\left(e_\nu / m_{\mathrm{e}}c^2\right)^2 A^2[1-(Z/A)], \qquad e_\nu \ll 300A^{-1/3}\,\mathrm{MeV} \tag{67}$$

where $\sigma_0 = 1.76 \times 10^{-44}\,\mathrm{cm}^2$ is a typical weak interaction cross section, e_ν the neutrino energy, and ε_{F} the Fermi enegy of the electrons. The total neutron opacity of the substance is given by:

$$\kappa = \rho^{-1}[n_{\mathrm{e}}\sigma_{\mathrm{e}} + n_{\mathrm{N}}\sigma_{\mathrm{N}} + n_{\mathrm{A}}\sigma_{\mathrm{A}}] \tag{68}$$

According to the Weinberg–Salam–Glashow theory, neutrinos also interact with quarks, and dense quark matter also emits and absorbs neutrinos. However, since our understanding of quark matter is still far from complete, no results related to neutrino emissivity and opacity in quark matter will be quoted at this time.

SEE ALSO THE FOLLOWING ARTICLES

ATOMIC PHYSICS • BINARY STARS • DARK MATTER IN THE UNIVERSE • NEUTRINOS • NEUTRON STARS • NUCLEAR PHYSICS • PARTICLE PHYSICS, ELEMENTARY • PLASMA SCIENCE AND ENGINEERING • QUANTUM THEORY • STELLAR STRUCTURE AND EVOLUTION • SUPERCONDUCTIVITY

BIBLIOGRAPHY

Chandrasekhar, S. (1939). "An Introduction to the Study of Stellar Structure," Dover, New York.

Clayton, D. D. (1968). "Principles of Stellar Evolution and Nucleosynthesis," McGraw-Hill, New York.

Gasiorowicz, S. (1979). "The Structure of Matter: A Survey of Modern Physics," Addision-Wesley, Reading, MA.

Leung, Y. C. (1984). "Physics of Dense Matter," Science Press, Beijing, China.

Schwarzschild, M. (1958). "Structure and Evolution of the Stars," Dover, New York.

Shapiro, S. L., and Teukolsky, S. A. "Black Holes, White Dwarfs, and Neutron Stars, The Physics of Compact Objects," Wiley, New York.

Zeldovich, Ya. B., and Novikov, I. D. (1971). "Relativistic Astrophysics," Chicago University Press, Chicago, IL.



Field Theory and the Standard Model

Giampiero Passarino
University of Turin

I. Quantum Field Theory (QFT) of Fundamental Interactions
II. QFT of Electromagnetism
III. Concept of Renormalization
IV. Gauge Invariance in QFT
V. Full Standard Model Lagrangian
VI. Confronting the Infinities
VII. Building Blocks for Renormalizing the Standard Model
VIII. Phenomenology of the Standard Model
IX. Conclusion

GLOSSARY

Born approximation The lowest order term in some coupling constant in the perturbative expansion for physical observables.

Box diagram A particular Feynman diagram with one loop and four external legs.

Covariant derivative A generalization of the ordinary derivative with special properties of transformation under a gauge transformation. Unlike ordinary derivatives, covariant derivatives of fields transform as the field itself.

Cross section A measure of the probability of transition between two configurations.

Cutoff Any parameter such as an energy scale (Λ) or the dimension of space-time ($n \neq 4$) that regularizes the ultraviolet (UV) divergences in QFT.

Electron Together with the electron neutrino, the up- and down-quark forms the stable matter.

Feynman diagrams A representation in terms of graphs of scattering and decay processes. They systematize the problem of calculation of physical observables.

Feynman rules The collection of rules that translate $\mathcal{S}$-matrix elements into Feynman diagrams.

Fields The quantum analogues of generalized coordinates.

Gauge fixing A procedure according to which Feynman

rules and diagrams can be derived in a theory that embodies a gauge symmetry.

Gauge theories Field theories based on a *gauge* symmetry. In electromagnetism the transformation amounts to adding a constant to the electric potential. The non-Abelian variants consist of transformations whose order matters.

Generations Experiments with the Large Electron Positron collider have shown that three generations, replica of the pairs with the lightest mass, electron, electron neutrino, the up- and down-quark.

Ghosts Fictitious particles that are needed to create a consistent quantum version of any gauge theory.

Gluons As electrically charged particles interact by exchanging photons (quantum electrodynamics), quarks interact by exchanging eight color-charged gluons. Gluons are electrically neutral.

Group A set of abstract elements with a rule for multiplication that has particular properties.

Higgs boson In the standard model all particles acquire their mass through their interaction with the Higgs field. The strength of the interaction is a measure of the mass.

Higgs mechanism In field theory this manifests itself by producing massive vector particles in a gauge-invariant Lagrangian.

Hilbert space The abstract space of states that gives physical information in a quantum theory.

Lagrangian A function of the fields and their first derivatives that describes the dynamics of the theory.

LEP The Large Electron Positron collider at CERN.

Leptons Fundamental fermions that are not subjected to the strong force.

Minimal standard model The standard model with only one physical Higgs boson.

Muon decay Because the electron is lighter than the muon, the muon may decay into a muon neutrino, an electron, and an electron antineutrino.

Parity A transformation that reverses the spatial components of vectors but leaves their time component unchanged.

Perturbation theory The organization of the calculation for any physical observables through a series in some coupling constant.

Photons The particles that carry the electromagnetic force. They are massless.

Propagator The simplest example of Feynman diagram representing the propagation of a particle between two sources.

QCD Quantum chromodynamics, the theory that governs the interaction of quarks and gluons.

QED Quantum electrodynamics, the quantum mechanical formulation of electricity and magnetism.

QED (QCD) radiation The emission of massless quanta from charged particles (bremsstrahlung process).

QFT Quantum field theory.

Quarks Fundamental fermions that carry the color charge. There are six varieties: up, down, charm, strange, top, and bottom.

Radiative corrections The terms of higher order in some coupling constant that give information on the physical observables.

Regularization Intermediate step to define a predictive theory that is applied before renormalization. In dimensional regularization space-time has dimension $n \neq 4$.

Renormalization A procedure that in a certain class of quantum field theories allows one to absorb certain infinite terms into the parameters of the Lagrangian such as masses and coupling constants.

Renormalization scheme One of the many options to implement the procedure of renormalization.

$\mathcal{S}$-matrix Its absolute value squared is the probability for in-state α to be observed as out-state β.

Scattering The process according to which some initial state of n particles transforms into m final particles.

Self-energy A particular Feynman diagram with one loop (or more) and two external legs.

SLC The Stanford Linear Collider.

Spin The intrinsic angular momentum of a particle.

Standard model A quantum field theory that incorporates the theory of weak and electromagnetic interactions (electroweak) and the theory of strong interactions (quantum chromodynamics, QCD).

Strong force The force responsible for keeping quarks bound in composite particles as protons, neutrons, and mesons.

Top quark The heaviest fundamental particle found so far.

Tree diagrams Feynman diagrams where internal momenta are completely determined by external momenta.

Unitarity A property of the quantum theory that guarantees the conservation of probabilities.

UV divergences Manifestations of a pathology of the theory, giving infinitely large theoretical predictions; see renormalization.

Vertex (diagram) A particular Feynman diagram with one loop (or more) and three external legs.

Weak force The force responsible for the decay of neutrons, pions, and muons and in general for radioactivity.

Weak-mixing angle A parameter of the standard model that, at lowest order in perturbation theory, is related to the ratio of the W and Z boson masses.

Width For an unstable particle this parameter represents the decay probability per unit time. Its inverse is the lifetime of the particle itself.

W* and *Z The particles that carry the weak force. The W is electrically charged, the Z is neutral. They are massive.

THE GREAT SUCCESS of modern particle physics is based on the possibility of describing the fundamental structure and behavior of matter within a theoretical framework called the standard model. This model incorporates all the known particles and the forces through which they interact, with the exception of gravity. It is currently the best description we have of the world of quarks and other particles. In recent years quantum field theory has been very successful, not only in the critical, but also in the popular sence. To some degree we may celebrate the fact that the theory agrees with the data and currently the numerical agreement is unequivocal.

I. QUANTUM FIELD THEORY (QFT) OF FUNDAMENTAL INTERACTIONS

In 1864, Maxwell explained electricity and magnetism as two manifestations of a single, unified electromagnetic force. The objects in our surroundings are all built up of atoms, which consist of electrons and atomic nuclei. In the nuclei there are protons and neutrons, which in turn are made up of quarks. The discovery and study of radioactivity and the subsequent development of atomic physics during the first half of the 20th century produced the concepts of strong and weak interaction. In simple terms, the strong interaction holds the atomic nucleus together, and the weak interaction allows certain nuclei to decay radioactively.

One hundred years after Maxwell, Glashow, Salam, and Weinberg independently discovered how the electromagnetic force could be described in the same theory as the weak force. In this way they linked a new class of phenomena, such as radioactivity, with the more familiar effects of electricity and magnetism.

The *electroweak* theory, shortly to become the standard model (SM), predicted that there must be a neutral carrier for the weak force (called Z) as well as charged carriers (called W^+ and W^-), and that the Z must give rise to weak neutral reactions, previously unseen. The first few of these *interactions* were observed at CERN in the Gargamelle bubble chamber in 1973.

Ten years later, the UA1 and UA2 experiments at CERN proved the existence of both the neutral and charged vector bosons. Rubbia and van der Meer from CERN were awarded the Nobel Prize for physics in 1984, respectively for the discovery of these particles and for the development of "stochastic cooling" method which made this discovery possible.

Electroweak theory now forms a central part of the description of forces in the standard model. The strong force, meanwhile, is described by a mathematically similar theory, in which eight different kinds of particles, the gluons, carry the force between quarks very much as the photon carries the force between electrically charged particles. This theory is called quantum chromodynamics (QCD). Therefore, in the standard model the forces between the constituents are described by quantum field theories, all of the non-Abelian gauge theory type. However, the theoretical foundation of the standard model was at first incomplete and it was totally unclear whether the theory could be used for detailed prediction of physical observables. 't Hooft (2000) and Veltman (2000a) placed this theory on a firmer foundation. The term gauge relates to a particular feature of these theories, a (*gauge*) symmetry that represents one of the most fundamental features of physics. Maxwell's theory of electromagnetism, in modern terminology, is a gauge theory.

The order in which one performs two gauge transformations is immaterial for electromagnetism and we say that electromagnetism is an Abelian gauge theory. In moving from the classical theory to its quantum mechanical counterpart we find new problems. Immediately after the formulation of quantum mechanics attempts were made to create a quantum field theory of electromagnetism. But the new quantum electrodynamics (QED) became complicated and predictions for physical observables often gave unreasonable results. The problem was solved in the 1940s by Tomonaga, Schwinger, and Feynman (who shared the 1965 Nobel Prize in physics). The method developed by these three is called renormalization and its meaning will be described below.

As early as the 1930s a first quantum field theory for weak interaction was formulated. This theory suffered from problems that were even worse than those of QED and not even the renormalization method could solve them.

In the mid-1950s Yang and Mills formulated the first example of a new field theory, a non-Abelian gauge theory. As opposed to QED, the result of the non-Abelian transformations depend on the order.

The new possibilities of the theory were not fully exploited until the 1960s when a number of researchers collaborated in the development of a non-Abelian gauge theory that unites electromagnetism and weak interaction into an electroweak interaction.

In the following the basic aspects of the standard model are summed up briefly. For a more detailed description see Bardin and Passarino (1999).

II. QFT OF ELECTROMAGNETISM

The Maxwell equations describe electricity and magnetism and we must face the problem of their quantization. While well understood classically, they present difficulties

at the quantum level, the problem being with gauge invariance. The classical equations are

$$\partial_\mu F^{\mu\nu} = j^\nu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad (1)$$

where A_μ is the potential and j_μ describes currents and charges. If some A_μ is a solution of these equations, then also $A_\mu + \partial_\mu \Lambda(x)$ is a solution of the same classical problem, where Λ is an arbitrary differential function. In quantum mechanics the four components of A_μ correspond to four different kinds of photons, each described by a wavefunction, and gauge invariance tell us that adding an arbitrary quantity to the wavefunctions will not change the result of a measurement of the system. In describing electromagnetism there will be unphysical photons that are, nevertheless, part of our formalism. To restrict the solutions in the classical theory, we impose a subsidiary condition, such as the Lorentz gauge $\partial_\mu A^\mu = 0$.

In quantum theory we introduce a Hilbert space with *in* and *out* states describing configurations of widely separated particles. An *in* vector corresponds to a physical system characterized by some configuration at time $t = -\infty$. Similarly, an *out* state correspond to time $t = +\infty$. There will be a matrix that transforms the *in* basis into the *out* basis for any configuration. Such a matrix $\mathcal{S}$ contains all physical information for any scattering process. Then one should formulate a subsidiary condition, such as the Lorentz one, for the $\mathcal{S}$-matrix. This turns out to be rather difficult and the adopted solution is to observe that, for all practical purposes, the Feynman diagrams represent the true content of the theory. Knowing the diagrams and the rules needed to construct them, we can derive all predictions for the theory and compare with the data. Furthermore, the $\mathcal{S}$-matrix is unitary by construction, i.e., probabilities are conserved, and we must have some other means to establish unitarity of the theory directly in terms of Feynman diagrams.

III. CONCEPT OF RENORMALIZATION

Consider some theory, supposed described by a Lagrangian depending on certain parameters. Suppose for simplicity there is only one parameter called x: $\mathcal{L} = \mathcal{L}(x)$. In the tree approximation there is no ambiguity and theoretical predictions from this Lagrangian can be compared with experiment. One data point is needed to fix x; after that any other comparison is a test of the theory. Of course, ideally one would like to combine all data and express the result as a probability that all the data are consistent with one free parameter, but we will not dwell on that.

Now suppose that one wants to go beyond the tree approximation. Then radiative corrections must be calculated. The relation between the parameter x and the experimental data becomes much more complicated. Nonetheless it remains precisely true: one measurement is needed to fix the free parameter x, the rest is a test. Of course the value of x as determined using only the tree approximation will be different from the value determined taking into account radiative corrections. As it happens, this difference is usually infinitely large because the radiative corrections contain infinities. Such infinities must be well defined and understood, but nowadays everybody uses the same regularization scheme, i.e., dimensional regularization, and there is as yet no real problem there. In still higher order there is the problem of how to define γ^5; however, that is not the issue here, but it must be mentioned because it is a potential source for scheme diversification.

Because of the awkward situation that the *corrected* x and the *tree* x are so different one introduces the notion of a counter-term. Thus in the Lagrangian one writes $x(1 + \delta x)$ instead of x, and δx is chosen in some well-defined manner such that now x remains in the neighborhood of the tree x. It is, however, purely a matter of convenience; the only thing that ever emerges in the confrontation with the data is $x(1 + \delta x)$. In order to have meaningful communication it is necessary, when talking about x, to specify what δx is used. Stating one's conventions on this matter is what is usually termed the *subtraction scheme* or *renormalization scheme*. Two essentially different approaches may be distinguished:

1. Prescribe precisely what x is.
2. Prescribe precisely what δx is.

Again, only the combination $x(1 + \delta x)$ appears in the confrontation with the data, and we are discussing here a matter of convention. As a matter of terminology we will call quantities such as δx *counter terms* and quantities such as X *bare parameters*.

In the older days of QED method 1 was the preferred. The convention was to prescribe x and to use for that some very well defined experimental quantity. The quantity δx is then obtained from the data including radiative corrections. A case in point is the electron mass. The quantity $m(1 + \delta m)$ was called the *bare mass* and m itself the *experimental mass*. This method also reflected some vague intuition about the physical meaning of the bare mass: if the interactions could be switched off, that is what one would see. Also, the mass of the electron is very well known, and the scheme is well understood. Convention 1 has the advantage of not being dependent on the choice of regularization scheme, but it offers a problem when there is no clear, precisely known experimental quantity that can play the role of defining x. Such is the case of QCD with respect to the coupling constant g of that theory. That g, at least as seen experimentally, is a function

of the scale, and moreover not easy to measure due to confinement. Consequently theorists have more or less settled on method 2. The quantity δx is prescribed and x is determined from some experiment depending on x. The δx defined by different schemes differ from each other by some finite amount.

The renormalization idea is, therefore, the following. Experimentally one never observes the lowest order alone, but the sum of all orders. Up to first order, the mass squared in a propagator is $m^2 - \delta m$ and that is what the experimenter observes. Therefore, $m^2 - \delta m$ is the observed mass squared and the theory makes no predictions about the mass. It is a free parameter, and it must be fixed by comparing the results of the theory with the observed data. The most important question is the following: Do all infinities of the theory appear in combination with a few parameters? If this is the case, we call the theory renormalizable, else nonrenormalizable. We start by assuming the existence of some cutoff Λ above which the theory eventually changes. The question now is what Λ-dependent effects could we expect at low energy, characterized by some energy scale $E \ll \Lambda$.

In working out perturbation theory (in some coupling constant g) we will encounter series in the variable $g\Lambda^2/E^2$. In a nonrenormalizable theory any measurable quantity will correspond to a series that at sufficiently high order diverges as $\Lambda \to \infty$,

$$g^l \left[a_0 + a_2 g^2 + \cdots + a_k g^k \left(\frac{\Lambda}{E}\right)^2 + a_{k+2} g^{k+2} \left(\frac{\Lambda}{E}\right)^4 + \cdots \right]. \tag{2}$$

For example, consider the leptonic part of the Fermi theory of weak interactions,

$$\begin{aligned} &\mathcal{L} = G_{\mathrm{F}} j_\mu j_\mu^\dagger, \qquad G_{\mathrm{F}} = \frac{g^2}{m_p^2}, \qquad g^2 \approx 10^{-5} \\ &J_\alpha = \sum_l \bar{\nu}_l \gamma_\alpha (1+\gamma_5) l. \end{aligned} \tag{3}$$

Consider $\nu_{\mathrm{e}} e$ elastic scattering. In lowest order the result is proportional to g^2; however, in next order we have three diagrams proportional to $g^4\Lambda^2/E^2$. The Λ effect is not measurable at low energies simply because through renormalization of g this effect can be transformed away. Another amplitude suffering large corrections is that for μ decay. The situation is precisely as before but the series has different coefficients and the renormalization of g on the basis of $\nu_{\mathrm{e}} e$ scattering will not neutralize the series for μ decay. Thus, now the corrections become observable and we can rule out values of Λ larger than E/g.

In a renormalizable theory the cutoff dependence is not observable and can be absorbed in the parameters of the theory, e.g., coupling constants and masses.

IV. GAUGE INVARIANCE IN QFT

The quantum mechanical counterpart of the subsidiary condition that restricts the solutions in the classical theory, for example, $\partial_\mu A^\mu = 0$, is that $\partial_\mu A^\mu$ is a free field that decouples, that is, does not interact with matter. To get rules for diagrams in a gauge theory, including an Abelian one, difficulties manifest in the fact that the matrix that defines the propagator of the theory has no inverse. Consider, for instance, the following Lagrangian:

$$\mathcal{L} = -\frac{1}{4} F_{\mu_n u} F^{\mu\nu} - \frac{1}{2} M^2 A_\mu A^\mu. \tag{4}$$

The propagator for the field A^μ, the basic block to construct Feynman diagrams, is given by the inverse of $V_{\mu\nu} = -(p^2 + M^2)\delta_{\mu\nu} + p_\mu p_\nu$, which has a simple solution

$$V^{-1}_{\mu\nu} = \frac{1}{(2\pi)^4 i} \frac{1}{p^2 + M^2} \left(\delta_{\mu\nu} + \frac{p_\mu p_\nu}{M^2} \right). \tag{5}$$

The gauge-invariant theory corresponding to $M = 0$ is therefore singular since V is singular.

If, due to gauge invariance, a Lagrangian is *singular*, then a good Lagrangian can be obtained by adding a term $-1/2C^2$, where C behaves nontrivially with respect to the gauge transformation, $C \to C + t\Lambda$. Here t is an operator that may contain derivatives and be field dependent. C will appear to be a free field and successively we must introduce the so-called Faddeev–Popov ghost fields to compensate for its introduction.

A gauge-fixed Lagrangian for QED is given by

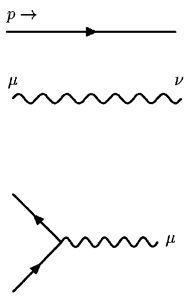
$$\begin{aligned} \mathcal{L}_{\mathrm{QED}} = &-\frac{1}{4} F_{\mu\nu} F_{\mu\nu} - \frac{1}{2} (\mathcal{C}_A)^2 \\ &- \sum_f \bar{\psi}_f (\not{\partial} - ie Q_f \not{A} + m_f) \psi_f, \end{aligned} \tag{6}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad \mathcal{C}_A = -\frac{1}{\xi} \partial_\mu A_\mu, \tag{7}$$

and where the sum runs over the fermion fields f. Each fermion has a charge eQ_f, with e being the charge of the positron, and mass m_f. Within the SM we have leptons with charge $Q_l = -1$, up-quarks with $Q_f = 2/3$, and down-quarks with charge $Q_f = -1/3$.

The Feynman rules of QED are particularly simple. They can be summarized as follows:



$$\begin{aligned} &\frac{1}{(2\pi)^4 i}\,\frac{-i\not{p}+m_f}{p^2+m_f^2-i\epsilon}, \\ &\frac{1}{(2\pi)^4 i}\,\frac{1}{p^2-i\epsilon} \\ &\quad\times\left[\delta_{\mu\nu}+(\xi^2-1)\frac{p_\mu p_\nu}{p^2}\right], \\ &(2\pi)^4 i\cdot ieQ_f\gamma_\mu. \end{aligned} \tag{8}$$

Starting from these rules, the full content of the theory can be analyzed and accurate predictions are formulated, for example, the $g-2$ of electrons and muons, the Lamb shift, etc.

V. FULL STANDARD MODEL LAGRANGIAN

In this section we give the basic rules for constructing the standard model (SM) Lagrangian. Since masses arise from a Higgs mechanism we must include some set of scalar fields and we assume the simplest (minimal) scalar sector.

The electroweak theory is based on the invariance group $SU(2)\otimes U(1)$ and we must discuss the field content of this theory in terms of representations of the group itself. Fields will be given by matrices augmented with a rule for group multiplication and the matrices themselves are considered as a matrix representation of the abstract group.

Within the SM Lagrangian there is a triplet of vector bosons B_μ^a, a singlet B_μ^0, a complex scalar field K, fermion families, and Faddeev–Popov ghost fields (hereafter FP) $X^\pm, Y^Z, Y^A$. The physical fields Z and A are related to B_μ^3 and B_μ^0 by a rotation in terms of the so-called weak-mixing angle.

The scalar field in the minimal realization of the SM is

$$K=\frac{1}{\sqrt{2}}\begin{pmatrix}\chi\\ \sqrt{2}i\phi^-\end{pmatrix},\qquad \chi=H+2\frac{M}{g}+i\phi^0, \tag{9}$$

where by H we denote the physical Higgs boson and M and g are Lagrangian parameters corresponding to the bare W mass and to the $SU(2)$ bare coupling constant. The total Lagrangian will be the sum of various pieces. The first is $\mathcal{L}_{\rm YM}+\mathcal{L}_{\rm S}$, with the standard Yang–Mills terms given by

$$\mathcal{L}_{\rm YM}=-\frac{1}{4}F_{\mu\nu}^a F_{\mu\nu}^a-\frac{1}{4}F_{\mu\nu}^0 F_{\mu\nu}^0, \tag{10}$$

where the first line gives the simplest generalization of the QED Lagrangian [$U(1)$ Abelian invariance] to a non-Abelian group, $SU(2)$. The *minimal* Higgs sector is specified by

$$\mathcal{L}_{\rm S}=-(D_\mu K)^+ D_\mu K-\mu^2 K^+K-\frac{1}{2}\lambda(K^+K)^2, \tag{11}$$

where $\lambda>0$ and symmetry breaking requires $\mu^2<0$. By the last statement we mean the mechanism of introducing masses for the vector bosons through the shift in the scalar field that allows for zero vacuum expectation value of the physical Higgs field. The remaining degrees of freedom in the K-doublet will be nonphysical, i.e., absent from the asymptotic *in* and *out* bases, and connected with the longitudinal polarizations of the spin-1 particles. Generically, spontaneous symmetry breaking refers to any situation where the system has a set of degenerate ground states related by continuous symmetry transformations.

Moreover, we use standard definitions for

$$\begin{aligned} F_{\mu\nu}^a &= \partial_\mu B_\nu^a-\partial_\nu B_\mu^a+g\varepsilon_{abc}B_\mu^b B_\nu^c, \\ F_{\mu\nu}^0 &= \partial_\mu B_\nu^0-\partial_\nu B_\mu^0, \end{aligned} \tag{12}$$

and the covariant derivative for the scalar field assumes the following form:

$$D_\mu K=\left(\partial_\mu-\frac{i}{2}gB_\mu^a\tau^a-\frac{i}{2}gg_1B_\mu^0\right)K, \tag{13}$$

with the standard Pauli matrices τ^a and $g_1=-s_\theta/c_\theta$. They follow from the fact that K, as defined in Eq. (9), belongs to a doublet representation of the symmetry group. The covariant derivative that we have just introduced is the natural generalization of the familiar concept of QED,

$$\bar\psi(\not{D}+m)\psi,\qquad D_\mu=\partial_\mu-igA_\mu, \tag{14}$$

which makes the Lagrangian invariant even in presence of matter fields.

Before going on, we split the Lagrangian into $\mathcal{L}_{\rm YM}-(D_\mu K)^+D_\mu K$ and $\mathcal{L}_{\rm S}^{\rm I}$, the latter containing the interactions of the scalar sector, and write

$$\begin{aligned} &\mathcal{L}_{\rm YM}-(D_\mu K)^+D_\mu K=\mathcal{L}_0 \\ &\quad+M\left(\frac{1}{c_\theta}Z_\mu\partial_\mu\phi^0+W_\mu^+\partial_\mu\phi^-+W_\mu^-\partial_\mu\phi^+\right), \end{aligned} \tag{15}$$

where the charged fields have been introduced as

$$\begin{aligned} &W_\mu^\pm=\frac{1}{\sqrt{2}}\left(B_\mu^1\mp iB_\mu^2\right),\qquad \phi^\pm=\frac{1}{\sqrt{2}}(\phi^1\mp i\phi^2), \\ &\phi^0\equiv\phi^3. \end{aligned} \tag{16}$$

This part of the Lagrangian contains $Z-\phi^0$, $W^\pm-\phi^\mp$ mixing terms; they are of the zeroth order in the coupling constant and their contribution must be summed up if we want to develop perturbation theory. There we discover the singularity of the Lagrangian. The Lagrangian is invariant under a set of transformations that are the generalization of the well-known QED example $A_\mu\to A_\mu+\partial_\mu\Lambda$.

The $SU(2) \otimes U(1)$ transformation laws of the various fields are as follows:

$$
\begin{aligned}
&B_\mu^a \to B_\mu^a + g\varepsilon_{abc}\Lambda^b B_\mu^c - \partial_\mu \Lambda^a, \quad B_\mu^0 \to B_\mu^0 - \partial_\mu \Lambda^0,\\
&K \to \left(1 - \frac{i}{2} g \Lambda^a \tau^a - \frac{i}{2} g g_1 \Lambda^0\right) K, \quad \text{with} \quad g_1 = -\frac{s_\theta}{c_\theta},\\
&H + i\phi^0 \to H + i\phi^0 - \frac{i}{2} g \Big[\big(\Lambda^3 + g_1 \Lambda^0\big) \\
&\qquad \times \left(H + 2\frac{M}{g} + i\phi^0\right) + 2i\Lambda^+ \phi^- \Big],\\
&\phi^0 \to \phi^0 - \frac{1}{2} g \big(\Lambda^3 + g_1 \Lambda^0\big)\left(H + 2\frac{M}{g}\right)\\
&\qquad + \frac{i}{2} g (\Lambda^- \phi^+ - \Lambda^+ \phi^-),\\
&\phi^- \to \phi^- - \frac{1}{2} g \Lambda^- \left(H + 2\frac{M}{g} + i\phi^0\right)\\
&\qquad - \frac{i}{2} g \big(- \Lambda^3 + g_1 \Lambda^0\big)\phi^-,
\end{aligned}
\tag{17}
$$

where we have taken the appropriate combinations of gauge parameters Λ^a ($a = 1, 2, 3$) and Λ^0. The construction of the SM continues as follows. First we add a *gauge-fixing* piece to the Lagrangian (called $\mathcal{L}_{\text{gf}}$ in the following) that cancels these mixing terms. However, it breaks the gauge invariance and successively we must introduce the so-called Faddeev–Popov ghost fields to compensate for this breaking. The gauge-fixing term transforms as

$$
\mathcal{C}^i \to \mathcal{C}^i + (\mathrm{M}^{ij} + g\mathrm{L}^{ij})\Lambda^j. \tag{18}
$$

M^{ij} must have an inverse and we thus have a permissible gauge. $g\mathrm{L}^{ij}$ defines the interaction with the gauge bosons. We now specify a set of gauges R_ξ depending on a single parameter ξ. We have a renormalizable gauge for finite ξ and the physical (unitary) gauge is obtained for $\xi \to \infty$. That these two gauges belong to the same family and are connected through a continuous parameter is vital in proving renormalizability and unitarity of the theory. The gauge-fixing piece is

$$
\mathcal{L}_{\text{gf}} = -\frac{1}{2}\mathcal{C}^a \mathcal{C}^a - \frac{1}{2}(\mathcal{C}^0)^2 = -\mathcal{C}^+\mathcal{C}^- - \frac{1}{2}[(\mathcal{C}^3)^2 + (\mathcal{C}^0)^2], \tag{19}
$$

where we can write

$$
\mathcal{C}^a = -\frac{1}{\xi}\partial_\mu B_\mu^a + \xi M \phi^a. \tag{20}
$$

The various components are given in the following equations: First

$$
\begin{aligned}
\mathcal{C}^\pm &= -\frac{1}{\xi}\partial_\mu W_\mu^\pm + \xi M \phi^\pm,\\
\mathcal{C}^0 &= -\frac{1}{\xi}\partial_\mu B_\mu^0 + \xi \frac{s_\theta}{c_\theta} M \phi^0.
\end{aligned}
\tag{21}
$$

Then, in the Z–A basis, we obtain

$$
\mathcal{C}_A = -\frac{1}{\xi}\partial_\mu A_\mu, \qquad \mathcal{C}_Z = -\frac{1}{\xi}\partial_\mu Z_\mu + \xi \frac{M}{c_\theta}\phi^0. \tag{22}
$$

In the R_ξ gauge we have that

$$
\begin{aligned}
&\mathcal{L}_{\text{YM}} - (D_\mu K)^+ D_\mu K - \mathcal{C}^+\mathcal{C}^- - \frac{1}{2}\mathcal{C}_Z^2 - \frac{1}{2}\mathcal{C}_A^2\\
&\quad = \mathcal{L}_{\text{prop}} + \mathcal{L}^{\text{bos,I}}.
\end{aligned}
\tag{23}
$$

The quadratic part of the Lagrangian, $\mathcal{L}_{\text{prop}}$, now reads

$$
\begin{aligned}
\mathcal{L}_{\text{prop}} = &-\partial_\mu W_\nu^+ \partial_\mu W_\nu^- + \left(1 - \frac{1}{\xi^2}\right)\partial_\mu W_\mu^+ \partial_\nu W_\nu^-\\
&- \frac{1}{2}\partial_\mu Z_\nu \partial_\mu Z_\nu + \frac{1}{2}\left(1 - \frac{1}{\xi^2}\right)(\partial_\mu Z_\mu)^2\\
&- \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu + \frac{1}{2}\left(1 - \frac{1}{\xi^2}\right)(\partial_\mu A_\mu)^2\\
&- \frac{1}{2}\partial_\mu H \partial_\mu H - \partial_\mu \phi^+ \partial_\mu \phi^- - \frac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0\\
&- M^2 W_\mu^+ W_\mu^- - \frac{1}{2}\frac{M^2}{c_\theta^2} Z_\mu Z_\mu - \xi^2 M^2 \phi^+ \phi^-\\
&- \frac{1}{2}\xi^2 \frac{M^2}{c_\theta^2}\phi^0 \phi^0 - \frac{1}{2} M_H H^2.
\end{aligned}
\tag{24}
$$

The quadratic part of the Lagrangian allows us to derive propagators. Those for the gauge fields are as follows:

$$
\begin{aligned}
\mathcal{L}_{\text{prop}} \to\ & W^\pm \quad \frac{1}{P^2 + M^2}\left\{\delta_{\mu\nu} + (\xi^2 - 1)\frac{p_\mu p_\nu}{p^2 + \xi^2 M^2}\right\}\\
&= \frac{1}{P^2 + M^2}\left(\delta_{\mu\nu} + \frac{p_\mu p_\nu}{M^2}\right)\\
&\quad - \frac{p_\mu p_\nu}{M^2(p^2 + \xi^2 M^2)}\\
&= \frac{1}{p^2 + M^2}\left(\delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right)\\
&\quad + \frac{\xi^2}{p^2 + \xi^2 M^2}\frac{p_\mu p_\nu}{p^2},\\
& Z \quad \text{is obtained from } W^\pm \text{ by replacing}\\
&\quad M \to \frac{M}{c_\theta},\\
& A \quad \frac{1}{p^2}\left\{\delta_{\mu\nu} + (\xi^2 - 1)\frac{p_\mu p_\nu}{p^2}\right\}.
\end{aligned}
\tag{25}
$$

The scalar field propagators are given by

$$
\begin{aligned}
&\text{----}\blacktriangleright\text{----}_{\phi^+} \quad \frac{1}{p^2 + \xi^2 M^2},\\
&\text{--------}_{\phi^0} \quad \frac{1}{p^2 + \xi^2 M^2/c_\theta^2}.
\end{aligned}
\tag{26}
$$

Having fixed the propagators, we can spell out the weak Lagrangian, describing the vector bosons and their interactions including interactions with the scalar system. The interested reader should consult any appropriate textbook for further details (e.g., Bardin and Passarino, 1999).

Having derived the first part of the Lagrangian, we now discuss the coupling of vector bosons with fermions. For most of the phenomenological applications this part of the Lagrangian is basically all we need.

Fermions will be arranged into two-column vectors (isodoublets). A generic fermion isodoublet will be denoted by

$$\psi = \begin{pmatrix} u \\ d \end{pmatrix}, \qquad \psi_{\mathrm{L,R}} = \frac{1}{2}(1 \pm \gamma_5)\psi, \tag{27}$$

where $u = \nu_l\,(l = e, \mu, \tau), u, c, t$-quark and $d = l\,(l = e, \mu, \tau), d, s, b$-quark. Furthermore, we distinguish between left and right fields since a theory of weak interactions cannot be purely vectorial, in contrast with QED (and QCD). The covariant derivative for the L-fields is

$$D_\mu \psi_{\mathrm{L}} = \left(\partial_\mu + g B_\mu^i T^i\right)\psi_{\mathrm{L}}, \qquad i = 0, \ldots, 3, \tag{28}$$

and which is written in terms of the following generators of $SU(2) \otimes U(1)$:

$$T^a = -\frac{i}{2}\tau^a, \qquad T^0 = -\frac{i}{2} g_2 I. \tag{29}$$

For the R-fields we have instead

$$D_\mu \psi_{\mathrm{R}} = \left(\partial_\mu + g B_\mu^i t^i\right)\psi_{\mathrm{R}}, \qquad i = 0, \ldots, 3, \tag{30}$$

$$t^a = 0, \qquad t^0 = -\frac{i}{2}\begin{pmatrix} g_3 & 0 \\ 0 & g_4 \end{pmatrix}. \tag{31}$$

This part of the Lagrangian can be written as

$$\mathcal{L}_{\mathrm{V}}^{\mathrm{fer,I}} = -\bar{\psi}_{\mathrm{L}} \not{D} \psi_{\mathrm{L}} - \bar{\psi}_{\mathrm{R}} \not{D} \psi_{\mathrm{R}}, \qquad g_i = -\frac{s_\theta}{c_\theta}\lambda_i. \tag{32}$$

The parameters g_2, g_3, and g_4 are arbitrary constants. However, one can prove that $g_3 = g_1 + g_2$. In other words, these constants are not completely free if we want to generate fermion masses with the help of the Higgs system.

Thus, ψ_{L} transforms as a doublet under $SU(2)$ and the ψ_{R} as a singlet. The parameters λ_i are then fixed by the requirement that the electromagnetic current has the conventional structure, $iQ_f e \bar{f}\gamma_\mu f$, without parity-violating terms and with the right normalization. We put $e = g s_\theta$ and derive the solution as

$$\begin{aligned} \lambda_2 &= 1 - 2Q_u = -1 - 2Q_d, \\ \lambda_3 &= -2Q_u, \qquad \lambda_4 = -2Q_d, \end{aligned} \tag{33}$$

where the charge is

$$Q_f = 2I_f^{(3)} |Q_f|. \tag{34}$$

$W^\pm$ always couples to a $V + A$ current, as expected, if the theory has to explain μ-decay and $\mathcal{L}_{\mathrm{V}}^{\mathrm{fer,I}}$ reads

$$\begin{aligned} \mathcal{L}_{\mathrm{V}}^{\mathrm{fer,I}} = \sum_f \Bigg[& i g s_\theta Q_f A_\mu \bar{f}\gamma_\mu f + i\frac{g}{2c_\theta} Z_\mu \bar{f}\gamma_\mu \\ & \times \left(I_f^{(3)} - 2Q_f s_\theta^2 + I_f^{(3)}\gamma_5\right) f \Bigg] \\ & + \sum_{\mathrm{d}} \Bigg[i\frac{g}{2\sqrt{2}} W_\mu^+ \bar{u}\gamma_\mu(1+\gamma_5) d \\ & + i\frac{g}{2\sqrt{2}} W_\mu^- \bar{d}\gamma_\mu(1+\gamma_5) u \Bigg], \end{aligned} \tag{35}$$

where the first sum runs over all fermions f and the second over all doublets, d, of the SM.

For the Higgs-fermion sector, in the presence of quarks, we need not only the field K, but its conjugate K^{c}; that is, we need both K and K^{c} in order to give mass to the up- and down-partners of the fermionic isodoublet. The K^{c} is

$$K^{\mathrm{c}} = -\frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{2} i\phi^+ \\ \chi^* \end{pmatrix}, \tag{36}$$

with the corresponding part of the Lagrangian

$$\mathcal{L}_{\mathrm{S}}^{\mathrm{fer}} = -\alpha_f \bar{\psi}_{\mathrm{L}} K u_{\mathrm{R}} - \beta_f \bar{\psi}_{\mathrm{L}} K^{\mathrm{c}} d_{\mathrm{R}} + \mathrm{h.c.} \tag{37}$$

The solution for the Yukawa couplings gives

$$\alpha_f = \frac{1}{\sqrt{2}} g \frac{m_u}{M}, \qquad \beta_f = -\frac{1}{\sqrt{2}} g \frac{m_d}{M}. \tag{38}$$

The last part of the Lagrangian is now

$$\mathcal{L}_{\mathrm{S}}^{\mathrm{fer}} = -\sum_f m_f \bar{f} f + \mathcal{L}_{\mathrm{S}}^{\mathrm{fer,I}}, \tag{39}$$

with an interaction Lagrangian given by

$$\begin{aligned} \mathcal{L}_{\mathrm{S}}^{\mathrm{fer,I}} = \sum_{\mathrm{d}} \Bigg\{ & i\frac{g}{2\sqrt{2}} \phi^+ \left[\frac{m_u}{M}\bar{u}(1+\gamma_5)d - \frac{m_d}{M}\bar{u}(1-\gamma_5)d\right] \\ & + i\frac{g}{2\sqrt{2}} \phi^- \left[\frac{m_d}{M}\bar{d}(1+\gamma_5)u - \frac{m_u}{M}\bar{d}(1-\gamma_5)u\right] \Bigg\} \\ & + \sum_f \left(-\frac{1}{2} g H \frac{m_f}{M}\bar{f} f + i g I_f^{(3)} \phi^0 \frac{m_f}{M}\bar{f}\gamma_5 f\right), \end{aligned} \tag{40}$$

which completes the construction of the SM Lagrangian.

VI. CONFRONTING THE INFINITIES

This short section will be devoted mainly to introducing some of the building blocks that are needed in order to discuss radiative corrections in any field theory. Beyond

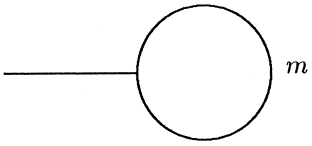


FIGURE 1 The one-point Green function.

the Born level, loops will appear and they will depend on several variables, internal and external masses.

To cope with the complications of the standard model in the general case, we must derive a complete set of formulas valid for arbitrary internal and external masses (Passarino and Veltman, 1979). One has to deal with expressions for scalar diagrams with one, two, three, and four external lines ('t Hooft and Veltman, 1979). Besides scalar functions we also need tensor integrals with up to four external legs and as many powers of momentum as allowed in a renormalizable theory. These tensor structures can be reduced to linear combinations of scalar functions. The one-point function is given in Fig. 1 and the corresponding expression will be discussed below.

We start by introducing the one-point scalar integrals

$$i\pi^2 A_0(m) = \mu^{4-n} \int d^n q \frac{1}{q^2 + m^2 - i\epsilon}, \qquad (41)$$

where μ is an arbitrary mass scale and we adopted dimensional regularization defining an analytical continuation of the $\mathcal{S}$-matrix in the complex n-plane. Note the presence of a factor i as a consequence of a Wick rotation. Within dimensional regularization one obtains a consistent theory if it can be shown that the poles for $n = 4$ can be removed order by order in perturbation theory.

This integral can be easily evaluated in terms of the Euler Γ function, giving

$$A_0(m) = \pi^{n/2-2}\Gamma\left(1 - \frac{n}{2}\right) m^2 \left(\frac{m^2}{\mu^2}\right)^{n/2-2}. \qquad (42)$$

If we introduce $\varepsilon = 4 - n$ and expand around $n = 4$, then we derive the following expression:

$$A_0(m) = m^2\left(-\frac{2}{\varepsilon} + \gamma + \ln\pi - 1 + \ln\frac{m^2}{\mu^2}\right) + \mathcal{O}(\varepsilon). \qquad (43)$$

where $\gamma = 0.577216$ is the Euler constant. It is customary to define a quantity $1/\bar{\varepsilon}$ by

$$\frac{1}{\bar{\varepsilon}} = \frac{2}{\varepsilon} - \gamma - \ln\pi, \qquad (44)$$

and to write

$$A_0(m) = m^2\left(-\frac{1}{\bar{\varepsilon}} - 1 + \ln\frac{m^2}{\mu^2}\right) + \mathcal{O}(\varepsilon). \qquad (45)$$

Explicit expressions for two- and higher point scalar functions will not be discussed here. For the two-point function we have an expression that contains logarithms at most, while for three- and four-point functions the final expression contains 12 and 108 (in the most general case) di-logarithms (Figs. 2–4).

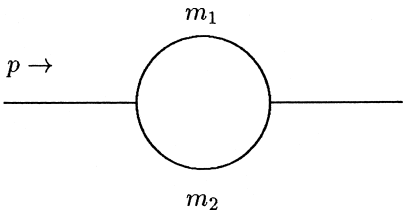


FIGURE 2 The two-point Green function.

VII. BUILDING BLOCKS FOR RENORMALIZING THE STANDARD MODEL

The renormalization of the QED Lagrangian has to be extended to cover the full standard model. Once this step is taken, one will be able to make complete theoretical predictions and compare them with experimental measurements.

A definition of the renormalization procedure is of basic importance:

- The renormalization procedure comprises the specification of the gauge-fixing term including, together with the corresponding FP Lagrangian, the choice of the regularization scheme—typically dimensional regularization—the prescription for the renormalization scheme, and the choice of a input parameter set.

For one loop we may say that the Lagrangian of the SM contains three parameters, which we are usually chosen to be the bare coupling constant g, the bare W mass M, and the bare weak-mixing angle s_θ. The caveat about one-loop is appropriate. In full generality, in the SM, we have many free parameters, for example, the coupling constant g, the weak mixing angle θ, the mass of the charged vector boson M, the mass of the Higgs boson M_{H}, the three

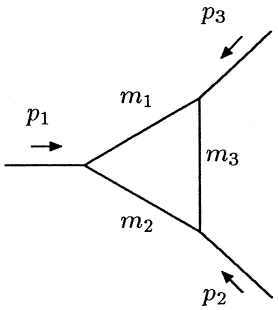


FIGURE 3 The three-point Green function.

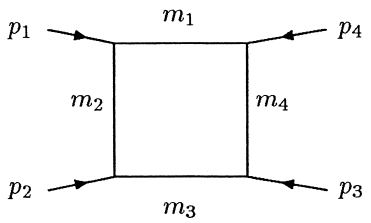


FIGURE 4 The four-point Green function.

angles θ_1, θ_2, and θ_3 in the Cabibbo–Kobayashi–Maskawa (CKM) matrix (Kobayashi and Maskawa, 1973) as well as a phase δ, and the masses of the quarks and leptons.

As far as the Higgs boson mass or the top quark mass are concerned, we immediately note that they will appear only inside loops and, for a one-loop renormalization devoted to studying the observables related to light fermion-pair production, these masses are simply the pole masses, i.e., the pole of the corresponding two-point function. In principle, we should pay special attention to internal masses in any diagram, but we always understand that bare masses and couplings are inserted in the loop calculations. Alternatively, we could insert renormalized quantities inside the loops, but then additional counter-term contributions would have to be taken into account, which become relevant at the two-loop level. In the same spirit there is no problem associated with lepton masses and with light quark masses apart from QCD corrections, which come into play whenever final-state light quarks are produced. There, we improve upon one-loop calculations by absorbing higher order QCD logarithms into a running mass. For consistency, therefore, light quark masses are everywhere treated as running masses.

Owing to the complexity of the Lagrangian as compared with the relatively easier case of QED, the discussion about differences and similarities among the available renormalization schemes is considerably more involved.

The QED Lagrangian in the Feynman gauge can be derived from Eq. (6), setting $\xi = 1$ and $Q_e = -1$. It is unambiguous at the tree level. Moving to higher orders, we assume that it is made of bare fields and parameters labeled with super or subindices 0 and specifies the renormalization constants for the two fields—A_μ and ψ—and the two QED parameters—the electron mass m and the charge e:

$$\begin{aligned} A_\mu^0 &= Z_A^{1/2} A_\mu, \qquad & \psi^0 &= Z_\psi^{1/2}\psi, \\ e_0 &= Z_e e, & m_0 &= Z_m m = m + e^2\delta m + \mathcal{O}(e^4), \\ Z_i &= 1 + e^2\delta Z_i + \mathcal{O}(e^4). \end{aligned} \tag{46}$$

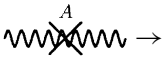
The Lagrangian can now be rewritten, up to terms $\mathcal{O}(e^2)$, as

$$\mathcal{L}_{\text{QED}}^{\text{ren}} = \mathcal{L}_{\text{QED}} + \mathcal{L}_{\text{ct}}, \tag{47}$$

with a counter-term Lagrangian:

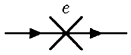
$$\begin{aligned} \mathcal{L}_{\text{ct}} &= e^2\mathcal{L}_{\text{ct}}^{(2)} + \mathcal{O}(e^4), \\ \mathcal{L}_{\text{ct}}^{(2)} &= -\frac{1}{4}\delta Z_A F_{\mu\nu}F_{\mu\nu} - \frac{1}{2}\delta Z_A(\partial_\mu A_\mu)^2 - \delta Z_\psi \bar\psi \not\partial \psi \\ &\quad - (\delta Z_\psi m + \delta m)\bar\psi\psi - i\left(\delta Z_e + \delta Z_\psi + \frac{1}{2}\delta Z_A\right) \\ &\quad \times e A_\mu \bar\psi \gamma_\mu \psi. \end{aligned} \tag{48}$$

We may say that the counter-term part of the Lagrangian generates a new set of QED Feynman rules, to be denoted by a cross. With their help we fix the counter-terms. First, the δZ_A counter-term:



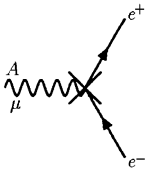
$$\rightarrow -e^2\delta Z_A. \tag{49}$$

Then the δZ_ψ and δm counter-terms:



$$\rightarrow -e^2(\delta Z_\psi i\not p + \delta Z_\psi m + \delta m). \tag{50}$$

Finally, the remaining combinations:



$$\rightarrow -ie\gamma_\mu e^3\left(\delta Z_e + \delta Z_\psi + \frac{1}{2}\delta Z_A\right). \tag{51}$$

After a relatively simple calculation one derives the following expressions:

$$\delta Z_A = \frac{1}{12\pi^2}\left(-\frac{1}{\bar\varepsilon} + \ln\frac{m^2}{\mu^2}\right). \tag{52}$$

$$\begin{aligned} \delta m &= \frac{m}{16\pi^2}\left(-\frac{3}{\bar\varepsilon} + 3\ln\frac{m^2}{\mu^2} - 4\right), \\ \delta Z_\psi &= \frac{1}{16\pi^2}\left(-\frac{1}{\bar\varepsilon} + \frac{2}{\hat\varepsilon} + 3\ln\frac{m^2}{\mu^2} - 4\right). \end{aligned} \tag{53}$$

Finally,

$$\delta Z_e \equiv -\frac{1}{2}\delta Z_A. \tag{54}$$

At this point the renormalization procedure can be carried through order by order. With the one-loop renormalized Lagrangian and with the one-loop counter Lagrangian we construct all two-loop diagrams and introduce $\mathcal{O}(e^2)$ new counter-terms. One obtains the correct result consistent with unitarity, provided that one has shown that overlapping diagrams contain new divergences behaving as local counter-terms.

A. Renormalization Schemes for the SM

There are many renormalization schemes in the literature that have been used to discuss the standard model,

for instance, the on-shell schemes in various implementations (Passarino and Veltman, 1979; Marciano and Sirlin, 1981). The most complete implementations of the OMS scheme can be found in the work of Bardin *et al.* (1980, 1982, 1989) for the nonminimal version and Fleischer and Jegerlehner (1981) or Böhm *et al.* (1986). Additional work can be found in Aoki *et al.* (1982) and Consoli *et al.* (1983). Next we have the GMS scheme first introduced in Passarino and Pittau (1989) and further discussed by Passarino (1991), sometimes called the G_F-scheme in the literature (Jegerlehner, 1986). Finally, one has the $\overline{MS}$ scheme with application developed originally in Marciano and Sirlin (1981) and more recently in Degrassi *et al.* (1991; Degrassi and Sirlin, 1991). A discussion of scheme dependence may be found in Hollik and Timme (1986).

To further the discussion, we need to develop the right procedure for renormalization of the electric charge, the Fermi constant, and of Z-boson mass.

In the OMS scheme we introduce multiplicative renormalization for the parameters and for the fields, a procedure that allows us to make finite all the Green functions. Note that multiplicative parameter renormalization is enough to render finite the $\mathcal{S}$-matix elements, the truly measurable quantities, once wavefunction renormalization factors are properly introduced.

Actually there are many equivalent parametrizations, depending on the set of fields—mass eigenstates or not—that we use, and on the choice between Yukawa parameters or fermion masses. They also depend on selecting the $SU(2)$ coupling constant g and its $U(1)$ partner g' versus g and the weak mixing angle s_θ. All the parametrizations of the counter-terms in the OMS scheme have in common that not only the $\mathcal{S}$-matrix elements, but also the Green's functions are finite. After having introduced this requirement, we can proceed according to the following two alternatives:

1. *The minimal OMS scheme*, $\phi_{0i} = Z_{1i}^{-1/2}\phi_i$. Within this scheme the renormalization constants refer to any given weak multiplet. They may be chosen such that finiteness of the Green functions is obtained. Consequently, the renormalized propagators are finite but not all of them have residue one (Hollik, 1990; Böhm *et al.*, 1986).

2. *The nonminimal OMS scheme*, $\phi_{0i} = (Z_1^{-1/2})_{ij}\phi_j$. This allows us to normalize to 1 all residues in the diagonal propagators and to prescribe values for the nondiagonal parts. Therefore the wavefunctions are properly normalized for absorption and emission of a particle or antiparticle.

A convenient choice is to forbid mixing for on-shell particles. In this case, no additional wavefunction renormalization factor is required.

In the minimal formulation of the OMS scheme we write

$$\begin{aligned}
B^a_{0\mu} &= Z_{2_W}^{1/2} B^a_\mu, & B_{0\mu} &= Z_{2_B}^{1/2} B_\mu,\\
\psi^j_{0\mathrm{L}} &= Z_{\mathrm{L}j}^{1/2}\psi^j_{\mathrm{L}}, & \psi^j_{0\mathrm{R}} &= Z_{Rj}^{1/2}\psi^j_{\mathrm{R}},\\
K_0 &= Z_{\mathrm{K}} K, & g_0 &= Z_{1_W} Z_{2_W}^{-3/2} g,\\
g'_0 &= Z_{1_B} Z_{2_B}^{-3/2} g', & g_{0,j\sigma} &= Z_K^{-1/2} Z_{1,j\sigma}\, g_{j\sigma},\\
\lambda_0 &= Z_K^{-2} Z_\lambda \lambda, & \mu_0^2 &= Z_K^{-1}(\mu^2 - \delta\mu^2),
\end{aligned} \tag{55}$$

where j is an isospin index and $g_{j\sigma}$ is a Yukawa coupling; furthermore, λ and μ are the parameters in the scalar self-interaction and K is the complex scalar isodoublet; finally, $\sigma = \pm\frac{1}{2}$.

As an example of a nonminimal OMS renormalization scheme, we mention the implementation used in the unitary gauge, where the independent parameters of the scheme are the electric charge, the masses of all particles, and the fields of the Lagrangian. We start with nonmini mal field renormalizations by nondiagonal matrices:

$$\begin{aligned}
\psi^i_{0\mathrm{L}} &= \left(Z_{\mathrm{L}}^{1/2}\right)_{ij}\psi^j_{\mathrm{L}}, & \psi^i_{0\mathrm{R}} &= \left(Z_{\mathrm{R}}^{1/2}\right)_{ij}\psi^j_{\mathrm{R}},\\
W_{0\mu} &= Z_W^{1/2} W_\mu, & Z_{0\mu} &= Z_Z^{1/2} Z_\mu,\\
H_0 &= Z_H^{1/2} H, & A_{0\mu} &= Z_A^{1/2} A_\mu + Z_M^{1/2} Z_\mu,
\end{aligned} \tag{56}$$

where, for instance, the renormalization constant $Z_M^{1/2}$ is fixed by the requirement that the γ–Z mixing vanishes on the two mass shells, $p^2 = 0$ and $p^2 = -M_Z^2$. The bosonic masses are renormalized in the usual way:

$$\begin{aligned}
M^2 &= Z_{M_W} Z_W^{-1} M_W^2, & M_0^2 &= Z_{M_Z} Z_Z^{-1} M_Z^2,\\
M_{0H}^2 &= Z_{M_H} Z_H^{-1} M_H^2,
\end{aligned} \tag{57}$$

while fermionic mass renormalization is more involved due to mixing, and requires the introduction of one more matrix of renormalization constants, Z_{m_f}. The latter has the dimensionality of mass and the corresponding part of the counter-term Lagrangian looks as follows:

$$\mathcal{L}_{\mathrm{ct}} \sim -\left(\bar\psi_{\mathrm{L}} Z_{m_f}\psi_{\mathrm{R}} + \bar\psi_{\mathrm{R}} Z^+_{m_f}\psi_{\mathrm{L}} - \bar\psi m_f \psi\right). \tag{58}$$

In this scheme there are more constants Z_i than physical parameters, and the essence of this particular implementation of the OMS scheme is to fix *all but one* of the renormalization constants by requiring that the residue of *all* the propagators be exactly 1. This remaining renormalization constant is associated with the renormalization of the electric charge, which we write as

$$e_0 = Z_e Z_A^{-1/2} e. \tag{59}$$

Another way of formulating the concept is to use the additive renormalization of the electric charge, expressed by

$$e_0^2 = e^2 + \delta e^2. \tag{60}$$

In the one-loop approximation the link between these two approaches is given by the following equality:

$$\frac{\delta e^2}{e^2} = 2(Z_e - 1) - (Z_A - 1). \tag{61}$$

The requirement that the relevant Ward identity be satisfied implies that

$$Z_e \equiv 1. \tag{62}$$

In any implementation of OMS we start with two rather important definitions: *The OMS weak-mixing angle*, to all orders in perturbation theory, is

$$M_Z^2 c_W^2 = M_W^2, \tag{63}$$

and we also have *the OMS weak charge*. In all the implementations of the OMS scheme we assign a fundamental role to the equation

$$g^2 = \frac{e^2}{s_W^2}, \tag{64}$$

which is therefore taken as the *definition* of the coupling constant g, valid to all orders.

There is a very simple meaning in the fact that both relations are considered as valid to *all orders*. Indeed, in the OMS schemes the *parameters* of the scheme are the complete set of masses and the electromagnetic coupling e, with the consequence that s_θ and g cannot be treated simultaneously as independent parameters, i.e., they must be related. For this definition we postulate that the lowest order relations are valid to all orders.

The full $\mathcal{L}_{\rm ct}$ of the SM may be derived in complete analogy with what is usually done for pure QED. The expression for $\mathcal{L}_{\rm ct}$ is rather long, even in the unitary gauge, where the number of fields is minimal, and it will not be presented here.

Alternatively, we will develop another strategy for renormalization where it will be understood that we prescribe counter-terms in the $\overline{MS}$ scheme and fix, from some experiment, the quantities $g^2_{\overline{MS}}$, $M_{\overline{MS}}$, and $(s_\theta)_{\overline{MS}}$. Indeed, in any renormalizable theory, the infinities cancel after renormalization in any physical observable. Therefore, we can reformulate the theory by setting everywhere $1/\bar{\varepsilon}$ to zero and by promoting the bare parameters to $\overline{MS}$ parameters. In other words: *defining an $\overline{MS}$ parameter is equivalent to adopting the heuristic rule* (*valid at one-loop*)

$$\frac{1}{\bar{\varepsilon}} + \ln \mu^2 \to \ln \mu^2_{\overline{MS}}, \tag{65}$$

in the relation expressing the bare parameters in terms of the renormalized ones. For instance,

$$\begin{aligned} e^2_{\overline{MS}}(\mu^2) &= 4\pi\alpha(0)\left[1 - \frac{\alpha(0)}{3\pi}\ln\frac{\mu^2_{\overline{MS}}}{m^2}\right]^{-1} \\ &\approx 4\pi\alpha(0)\left[1 + \frac{\alpha(0)}{3\pi}\ln\frac{\mu^2_{\overline{MS}}}{m^2}\right]. \end{aligned} \tag{66}$$

We refer to this renormalization procedure as the GMS scheme, generalized minimal subtraction. Here we consider g, s_θ, and M as the most significant parameters defining the SM. These three quantities can be fixed by choosing experimental data points, measured with high accuracy. There is one remaining unknown in the theory, namely the Higgs boson mass. As usual, finite predictions for measurable quantities will be functions of $M_{\rm H}$ and, in principle, this parameter can be constrained through its effect in radiative corrections.

The SM is a system that, like QED and unlike the four-fermion theory, needs no cutoff. It is conceivable that this model may describe physics up to a very large scale, even though it is plausible that at some definite scale new physics may come in. A procedure that is at the heart of the GMS scheme can be summarized as follows: it is possible to avoid the explicit construction of a renormalization procedure since the relations between physical quantities are finite and by taking as many input data points as required to fit the free parameters, we are able to obtain predictions for other physical processes. The essence of this procedure consists in writing a set of three renormalization (fitting) equations (Veltman, 1977),

$$d_i^{\rm exp} = d_i(g, M, s_\theta), \qquad i = 1, 2, 3, \tag{67}$$

or more in a nonminimal model, i.e., a model with higher representation for the Higgs field. The remaining masses are fixed by the one-to-one relations $m^{\rm exp} = m$, where m stands for the mass of the Higgs boson, of leptons, of the top quark.

These renormalization equations can be solved, in one approximation or another, and we obtain the bare parameters of the theory. Whenever bare parameters are shown, they must be understood as strictly defined in the $\overline{MS}$ sense. However, any definition of $1/\bar{\varepsilon}$ suffices since our goal is to write renormalized quantities and to study their impact on physical observables:

$$g = g\big(d_1^{\rm exp}, d_2^{\rm exp}, d_3^{\rm exp}\big), \ldots. \tag{68}$$

As explained above, the bare parameters so determined contain infinities and a dependence on the unit of mass μ. We can define the $\overline{MS}$ parameters by throwing away infinities and choosing the appropriate scale, but in full generality we can also give any value to the quantity $1/\bar{\varepsilon}$ and proceed to discussing measurable quantities.

In contrast to QED, where we have a well-defined and unique correspondence between bare parameters and experimental data points, here, for the first time, we encounter a rather different situation. To be more specific, we observe that the tree level s_θ of the minimal standard model (MSM) is expressible through at least two different ratios (Passarino and Veltman, 1990): the ratio between the $SU(2)$ and the $U(1)$ couplings in the

covariant derivative of the Higgs doublet—$s_\theta = e/g$—or the mass ratio $c_W = M_W/M_Z$. In the definition of the on-shell scheme we assume that there is a θ_W with the relation $s_W^2 = 1 - M_W^2/M_Z^2$ valid to all orders, while the complementary relation, $M_W^2 s_W^2 = \pi\alpha/(\sqrt{2}\,G_F)$, is modified by radiative corrections. However, the s_θ^2 that appears here could also be identified with the ratio e^2/g^2, which is not in a one-to-one correspondence with s_W^2. This is the main reason why the differences between renormalization schemes are more involved in the SM than in QED.

With this caveat we may briefly summarize the whole procedure for renormalizing the SM. We will use the fine structure constant defined as the residue of the pole at zero momentum transfer of charged particle scattering and show that the renormalization of the electric charge in the model is gauge parameter independent. In order to write a second renormalization condition we consider next the μ-lifetime, showing that, also in this case, gauge parameter independence is achieved. Our set of input parameters is finally specified by selecting the on-shell Z mass. Note that gauge parameter independence of the $\mathcal{S}$-matrix follows from a general theorem and here we are simply stating that all the formal manipulations inherent in the renormalization conditions do not spoil this property.

A crucial point in developing high-precision calculations is represented by the possibility of improving upon first-order perturbation theory, and the keyword here is *resummation*. This is a procedure that allows for the correct treatment of potentially large one-loop terms.

Resummation, in other words, deals with higher order reducible diagrams. Moreover, resummation allows us to express the results in the convenient language of effective couplings where the theory maintains its Born-level structure and all the couplings are promoted to become functions of the scale at which the phenomenon is observed. It can be shown that this improved Born approximation contains the bulk of some large effect arising in higher orders and the nonfactorizable part is usually small, although essential for high-precision predictions. Unfortunately, gauge invariance of the resummed result is not always respected, in particular when we start treating one-loop diagrams in different ways, for example, resummation of self-energies, first order for vertices and boxes.

VIII. PHENOMENOLOGY OF THE STANDARD MODEL

The standard model is a predictive theory where observables are computed to a very high degree of accuracy. Precise electroweak measurements performed at the Large Electron Positron Collider (LEP), the Stanford Linear Collider (SLD), and elsewhere can be used to check the validity of the model and, within its framework, to infer information about its fundamental parameters.

In the following we review the most important applications to the phenomenology of fundamental interactions. The outcome of this analysis can be summarized by saying that within the standard model, and using the available data, we can determine the top quark and W masses indirectly and to high accuracy. Furthermore, the number of light neutrino species is $N_\nu = 2.9835 \pm 0.0083$, which is only two standard deviations below the expected value of 3. Finally, the Higgs boson mass is constrained by comparing data with radiative corrections and $M_{\rm H} < 203\,\text{GeV}$ at 95% confidence level. Therefore, we may say that whichever the *new* theory will be, at low energies (below 200 GeV) it behaves precisely as the standard model with a minimal structure of the Higgs component.

Within the context of the SM we define realistic observables (RO) that are described in terms of some set of amplitudes

$$A_{\rm SM} = A_\gamma + A_Z + \text{nonfactorizable}, \tag{69}$$

for the process $e^+e^- \to f\bar{f}$, which, at the Born level, is depicted in Fig. 5. Note that this family of processes is the relevant one for LEP 1 energy range, i.e., around the Z resonance. At higher energies additional channels will open, and above the $WW(ZZ)$ thresholds, $e^+e^- \to 4$ fermions must be included, too. The last term in Eq. (69) is due to all those contributions that do not factorize into the Born-like amplitude, for example, weak boxes. Once the matrix element $A_{\rm SM}$ is computed, squared, and

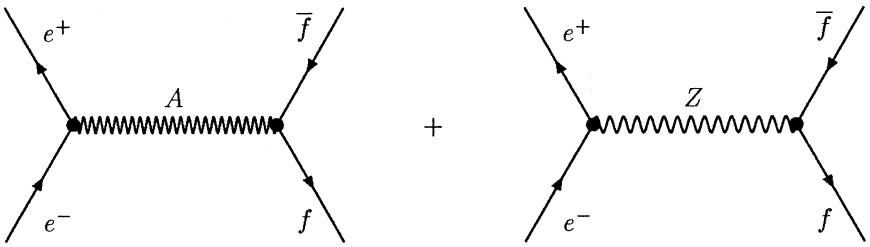


FIGURE 5 The process $e^+e^- \to (Z, A) \to f\bar{f}$ in the Born approximation.

integrated to obtain the cross section, a convolution with initial- and final-state QED and final-state QCD radiation follows:

$$\sigma(s) = \int dz\, H_{\rm in}(z,s) H_{\rm fin}(z,s)\, \hat{\sigma}(zs), \qquad (70)$$

where $H_{\rm in}(z,s)$ and $H_{\rm fin}(z,s)$ are so-called *radiator* or *flux* functions accounting for *initial-* (*final-*) *state radiation* (*ISR*, *FSR*), respectively, and $\hat{\sigma}(zs)$ is the kernel cross section of the hard process, evaluated at the reduced center-of-mass energy $s' = zs$.

In QED, and more generally in the SM, the most important terms in radiative corrections to various processes at high energy or at large momentum transfer are those that contain large logarithms of the type $\ln(Q^2/m^2)$, where Q^2 stands for a large kinematic variable. The quantity m is the mass of a light, charged particle, which emits photons, for example, the leptons, or the light quarks.

In QED or in the SM the theory becomes almost massless at high energies, that is, almost mass singular, due to the presence of low-mass fermions. In those cases where we need a very high accuracy of theoretical predictions, the presence of large logarithms calls for summing up the perturbation series, in spite of the low value of the coupling constant. The method for such a summation was actually developed within QCD and is based on the factorization theorems, which allow us to split the contributions of large and small distances. As a result of this resummation procedure one starts with the cross section for the so-called hard process, that is, the process with large kinematic variables, which is subsequently convoluted with the structure function of the initial (final) particles.

Suppose that we are considering the QED process $e^+e^- \to f\bar{f}$ and that the only kinematic cut imposed is that on the invariant mass of the $f\bar{f}$ pair, $\hat{s} > s_0$. This allows us to substitute the result for $D(x;s)$ into the folding of the hard scattering cross-section, to define $\hat{s} = (1-x)s$ and to obtain the result of Eq. (70).

It is well known that the structure of the matrix element for the process $e^+(p_+)e^-(p_-) \to \bar{f}(q_+) f(q_-)$ changes after inclusion of higher order electroweak corrections. One needs the introduction of complex-valued form-factors which depend on the two Mandelstam variables $s = -(p_+ + p_-)^2$ and $t = -(p_- - q_-)^2$. The separation into insertions for the γ exchange and for the Z exchange is lost.

The weak boxes are present as nonresonant insertions to the electroweak form-factors. At the Z resonance, the one-loop weak box terms are small, with relative contributions $\leq 10^{-4}$. If we neglect them, the t dependence is turned off. The t dependence would also spoil factorization of the form-factors into products of effective vector and axial–vector couplings.

Full factorization is re-established by neglecting various terms that are $\mathcal{O}(\alpha\Gamma_Z/M_Z)$, where Γ_Z is the Z width. The resulting effective vector and axial–vector couplings are complex valued and dependent on s. The factorization is the result of a variety of approximations which are valid at the Z resonance to the accuracy needed.

After the above-mentioned series of approximations we arrive at the so-called Z-boson pole approximation, which is actually equivalent to setting $s = M_Z^2$ in the form-factors. After deconvoluting ROs of QED and QCD radiation the set of approximations transform realistic observables into pseudo-observables, the ground where theory and data are usually confronted.

A. Definition of Pseudo-Observables

The experiments *measure* some primordial (basically cross sections and thereby asymmetries also) quantities which are then reduced to secondary quantities under some set of specific assumptions. Within these assumptions, the secondary quantities, the pseudo-observables (PO), also deserve the label of *observability*.

Independent of the particular realization of the effective couplings, they are complex-valued functions, due to the imaginary parts of the diagrams. Imaginary parts, although not next-to-leading in a strict sense, are sizeable two-loop effects. These are enhanced by factors π^2 and sometimes also by a factor N_f, with N_f being the total number of fermions (flavor $\otimes$ color) in the SM. Once we include the best of the two-loop terms, then imaginary parts should also come in.

The explicit formulas for the $Zf\bar{f}$ vertex are always written starting from a Born-like form of a prefactor $\times$ fermionic current, where the Born parameters are promoted to effective, scale-dependent parameters,

$$\begin{aligned}\rho_Z^f \gamma_\mu \big[\big(I_f^{(3)} + i a_L\big)\gamma_+ - 2Q_f \kappa_Z^f s^2 + i a_Q \big] \\ = \gamma_\mu \big(\mathcal{G}_V^f + \mathcal{G}_A^f \gamma_5\big), \end{aligned} \qquad (71)$$

where $\gamma_+ = 1 + \gamma_5$ and $a_{Q,L}$ are the SM imaginary parts.

By definition, the total and partial widths of the Z boson include also QED and QCD corrections. The partial decay width is therefore described by the following expression:

$$\begin{aligned}\Gamma_f &\equiv \Gamma(Z \to f\bar{f}) \\ &= 4 c_f \Gamma_0 \big(|\mathcal{G}_V^f|^2 R_V^f + |\mathcal{G}_A^f|^2 R_A^f \big) + \Delta_{\rm EW/QCD}, \end{aligned} \qquad (72)$$

where $c_f = 1$ or 3 for leptons or quarks ($f = l, q$), and R_V^f and R_A^f describe the final-state QED and QCD corrections and take into account the fermion mass m_f. The last term,

$$\Delta_{\rm EW/QCD} = \Gamma^{(2)}_{\rm EW/QCD} - \frac{\alpha_S}{\pi}\Gamma^{(1)}_{\rm EW}, \qquad (73)$$

accounts for the nonfactorizable corrections. The standard partial width Γ_0 is

$$\Gamma_0 = \frac{G_F M_Z^3}{24\sqrt{2}\pi} = 82.945(7) \quad \text{MeV}. \tag{74}$$

The peak hadronic and leptonic cross sections are defined by

$$\sigma_h^0 = 12\pi \frac{\Gamma_e \Gamma_h}{M_Z^2 \Gamma_Z^2} \qquad \sigma_\ell^0 = 12\pi \frac{\Gamma_e \Gamma_l}{M_Z^2 \Gamma_Z^2}, \tag{75}$$

where Γ_Z is the total decay width of the Z boson, i.e., the sum of all partial decay widths.

The effective electroweak mixing angles (*effective sines*) are always defined by

$$4|Q_f|\sin^2\theta_{\text{eff}}^f = 1 - \frac{\operatorname{Re}\mathcal{G}_V^f}{\operatorname{Re}\mathcal{G}_A^f} = 1 - \frac{g_V^f}{g_A^f}, \tag{76}$$

where we define

$$g_V^f = \operatorname{Re}\mathcal{G}_V^f, \qquad g_A^f = \operatorname{Re}\mathcal{G}_A^f. \tag{77}$$

The forward-backward asymmetry A_{FB} is defined via

$$A_{\text{FB}} = \frac{\sigma_{\text{F}} - \sigma_{\text{B}}}{\sigma_{\text{F}} + \sigma_{\text{B}}}\,, \qquad \sigma_{\text{T}} = \sigma_{\text{F}} + \sigma_{\text{B}}\,, \tag{78}$$

where σ_{F} and σ_{B} are the cross sections for forward and backward scattering, respectively. Before analyzing the forward–backward asymmetries we have to describe the inclusion of imaginary parts. A_{FB} is calculated as

$$A_{\text{FB}} = \frac{3}{4}\frac{\sigma_{\text{VA}}}{\sigma_{\text{T}}}, \tag{79}$$

where

$$\begin{aligned}\sigma_{\text{VA}} = {} & \frac{G_F M_Z^2}{\sqrt{2}}\sqrt{\rho_e\rho_f}\,Q_e Q_f \operatorname{Re}\bigl[\alpha^*\bigl(M_Z^2\bigr)\mathcal{G}_V^e\mathcal{G}_A^f\chi(s)\bigr] \\ & + \frac{G_F^2 M_Z^4}{8\pi}\rho_e\rho_f \operatorname{Re}\bigl[\mathcal{G}_V^e\bigl(\mathcal{G}_A^e\bigr)^*\bigr] \\ & \times \operatorname{Re}\bigl[\mathcal{G}_V^f\bigl(\mathcal{G}_A^f\bigr)^*\bigr] s|\chi(s)|^2. \end{aligned} \tag{80}$$

This result is valid in the realization where ρ_f is a real quantity, i.e., the imaginary parts are not resummed in ρ_f. In this case

$$\begin{aligned}\mathcal{G}_V^f &= \operatorname{Re}\bigl(\mathcal{G}_V^f\bigr) + i\,\operatorname{Im}\bigl(\mathcal{G}_V^f\bigr) = g_V^f + i\,\operatorname{Im}\bigl(\mathcal{G}_V^f\bigr) \\ \mathcal{G}_A^f &= I_f^{(3)} + i\,\operatorname{Im}\bigl(\mathcal{G}_A^f\bigr)\end{aligned} \tag{81}$$

Otherwise $\mathcal{G}_A^f = I_f^{(3)}$ is a real quantity but ρ_f is complex valued and Eq. (80) has to be changed accordingly, i.e., we introduce

$$g_V^f = \sqrt{\rho_f}\,v_f\,, \qquad g_A^f = \sqrt{\rho_f}\,I_f^{(3)}\,, \tag{82}$$

with

$$v_f = I_f^{(3)} - 2Q_f \sin^2\theta_{\text{eff}}^f. \tag{83}$$

For the peak asymmetry, the presence of the ρ's is irrelevant since they will cancel in the ratio. We have

$$\hat{A}_{\text{FB}}^{0\text{f}} = \frac{3}{4}\hat{A}_e\hat{A}_f, \qquad \hat{A}_f = \frac{2\operatorname{Re}\bigl[\mathcal{G}_V^f\bigl(\mathcal{G}_A^f\bigr)^*\bigr]}{\bigl(\bigl|\mathcal{G}_V^f\bigr|^2 + \bigl|\mathcal{G}_A^f\bigr|^2\bigr)}. \tag{84}$$

The question is what to do with imaginary parts in Eq. (84)? For partial widths, as they absorb all corrections, the convention is to use

$$\bigl|\mathcal{G}_{V,A}^f\bigr|^2 = \bigl(\operatorname{Re}\mathcal{G}_{V,A}^f\bigr)^2 + \bigl(\operatorname{Im}\mathcal{G}_{V,A}^f\bigr)^2. \tag{85}$$

On the contrary, the PO peak asymmetry $A_{\text{FB}}^{0\text{f}}$ will be defined by an analogy of equation (84) where *conventionally* imaginary parts are not included,

$$A_{\text{FB}}^{0\text{f}} = \frac{3}{4}\mathcal{A}_e\mathcal{A}_f, \qquad \mathcal{A}_f = \frac{2\bigl(g_V^f g_A^f\bigr)}{\bigl(g_V^f\bigr)^2 + \bigl(g_A^f\bigr)^2}. \tag{86}$$

We note that Eq. (86) is not an approximation of Eq. (84). Both are POs and both could be used as the *definition*. Numerically, they give very similar results.

In contrast to POs, which are defined, it is impossible to avoid imaginary parts for ROs without spoiling the comparison between the theoretical prediction and the experimental measurement. Then one has to start with Eq. (80). We will develop Eq. (80) in the realization where imaginary parts are added linearly. For the ZZ part of the VA cross section one derives:

$$\operatorname{Re}\bigl[\mathcal{G}_V^e\bigl(\mathcal{G}_A^e\bigr)^*\bigr]\operatorname{Re}\bigl[\mathcal{G}_V^f\bigl(\mathcal{G}_A^f\bigr)^*\bigr]. \tag{87}$$

This collapses to a familiar expression if the axial–vector coefficients are real; however, one cannot factorize and simplify the ρ's especially away from the pole because of the $Z\gamma$ component. For the $Z\gamma$ part of the VA cross section one has the following result:

$$\operatorname{Re}[\alpha^*(s)\chi(s)]\operatorname{Re}\bigl(\mathcal{G}_A^e\mathcal{G}_A^f\bigr) - \operatorname{Im}[\alpha^*(s)\chi(s)]\operatorname{Im}\bigl(\mathcal{G}_A^e\mathcal{G}_A^f\bigr). \tag{88}$$

A definition of the PO heavy quark forward–backward asymmetry parameter which would include mass effects is

$$\mathcal{A}_b = \frac{2g_V^b g_A^b}{\frac{1}{2}(3-\beta^2)\bigl(g_V^b\bigr)^2 + \beta^2\bigl(g_A^b\bigr)^2}\beta, \tag{89}$$

where β is the b-quark velocity. The difference is very small due to an accidental cancellation of the mass corrections between the numerator and denominator of Eq. (89). This occurs for down-quarks, where $(g_V^b)^2 \approx (g_A^b)^2/2$ and where

$$\begin{aligned}A_{\text{FB}}^{0\text{b}} &\approx \frac{3}{4}\frac{2g_V^e g_A^e}{\bigl(g_V^e\bigr)^2 + \bigl(g_A^e\bigr)^2}\frac{2g_V^b g_A^b}{\bigl(g_V^b\bigr)^2 + \bigl(g_A^b\bigr)^2}(1+\delta_{\text{mass}}), \\ \delta_{\text{mass}} &\approx 4\frac{m_q^2}{s}\frac{\bigl(g_A^b\bigr)^2/2 - \bigl(g_V^b\bigr)^2}{\bigl(g_V^b\bigr)^2 + \bigl(g_A^b\bigr)^2}.\end{aligned} \tag{90}$$

Note that for the c-quark this difference is even bigger, one more example that for b-quarks we meet an accidental cancellation. Note that the mass effect should be even smaller since running quark masses seem to be the relevant quantities instead of the pole ones. Therefore, our definition of the PO forward–backward asymmetry and coupling parameter will be as in Eq. (86).

The most important higher order effects to be inserted in the PO calculations consist in the inclusion of higher order QCD corrections, mixed electroweak–QCD corrections (Czarnecki *et al*., 1996; Harlander *et al*., 1998), and next-to-leading two-loop corrections of $\mathcal{O}(\alpha^2 m_t^2)$ (Degrassi *et al*., 1991, 1996, 1997; Degrassi and Sihrlin, 1996).

Degrassi *et al*. incorporated the two-loop $\mathcal{O}(\alpha^2 m_t^2)$ corrections in the theoretical calculation of M_W and $\sin^2\theta_{\text{eff}}^{\text{lept}}$. More recently the complete calculation of the decay rate of the Z has been made available (Degrassi and Gambino, 2000). The only case that is not covered is the one of final b-quarks because it involves nonuniversal $\mathcal{O}(\alpha^2 m_t^2)$ vertex corrections.

Another development in the computation of radiative corrections to the hadronic decay of the Z is contained in two papers which together provide complete corrections of $\mathcal{O}(\alpha\alpha_S)$ to $\Gamma(Z \to q\bar{q})$ with $q = u, d, s, c$, and b. Czarencki *et al*. (1996) treat the decay into light quarks. Harlander *et al*. (1998) consider the remaining diagrams contributing to the decay into bottom quarks and thus the mixed two-loop corrections are complete.

B. Pseudo-Observables and Experimental Data

Usually, 25 of such *pseudo-observables* are introduced and discussed; namely, the mass of the W (M_W), the hadronic peak cross section (σ_h), the partial leptonic and hadronic widths (Γ_f, $f = \nu, e, \mu, \tau, u, d, c, s, b$), the total width ($\Gamma_Z$), the total hadronic width (Γ_h), the total invisible width (Γ_{inv}), various ratios (R_l, R_b, R_c), and the asymmetries and polarizations (A_{FB}^{μ}, A_{LR}^{e}, A_{FB}^{b}, A_{FB}^{c}, P^{τ}, P^{b}). Finally, we have effective sines ($\sin^2\theta_e$, $\sin^2\theta_b$).

The effective weak-mixing angle is definable, in principle, for all fermions, but we know that the largest difference will be in $\sin^2\theta_{\text{eff}}^{b} - \sin^2\theta_{\text{eff}}^{e}$ due to large flavor-dependent corrections. However, only $\sin^2\theta_{\text{eff}}^{e}$ is usually reported, which permits the following definition:

$$\sin^2\theta_{\text{eff}}^{\text{lept}} = \sin^2\theta_{\text{eff}}^{e}. \tag{91}$$

There is more with respect to the Z partial widths: By definition, the total and partial widths of the Z boson include final-state QED and QCD radiation.

Moreover, we have defined

$$\Gamma_h = \Gamma_u + \Gamma_d + \Gamma_c + \Gamma_s + \Gamma_b,$$
$$\Gamma_{\text{inv}} = \Gamma_Z - \Gamma_e - \Gamma_\mu - \Gamma_\tau - \Gamma_h,$$

$$R_l = \frac{\Gamma_h}{\Gamma_e}, \qquad R_{b,c} = \frac{\Gamma_{b,c}}{\Gamma_h}, \qquad \sigma_h = 12\pi\frac{\Gamma_e\Gamma_h}{M_Z^2\Gamma_Z^2}.$$

To this end, we emphasize that usually $\Gamma_{\text{inv}} = 3\Gamma_\nu$ is assumed. Then the total Z width becomes $\Gamma_Z = 3\Gamma_\nu + \Gamma_e + \Gamma_\mu + \Gamma_\tau + \Gamma_h$.

C. Fitting the Experimental Data

Finally, a fit is performed to the dataset. The aim of the fit is to check the validity of the SM and, within its framework, to establish information about its basic parameters. In what follows we present the most relevant aspects in the fitting procedure.

1. There is an error induced by the uncertainty in $\alpha(M_Z^2)$, the fine structure constant of QED, arising from the contribution of light quarks.
2. The theoretical error on $\alpha_S(M_Z^2)$, the QCD running coupling constant, comes from missing higher order electroweak corrections and uncertainties in the interplay of electroweak and QCD corrections. The value of $\alpha_S(M_Z^2)$ depends essentially on R_l, Γ_Z, and σ_h. The current average is $\alpha_S(M_Z^2) = 0.118 \pm 0.003$.
3. The top quark mass has a double role in the fit. The direct experimental information of $m_t = 174.3 \pm 5.1$ GeV can be used in the fit either by introducing the appropriate penalty function or by performing the fit to all data except m_t.
4. Once the χ^2_{min} has been determined, we use the parameters at the minimum and evaluate the list of observables within the SM. A useful quantity in this respect is represented by the *pull* associated with each observable. The pull is defined to be the difference between the measured quantity and the SM prediction at the minimum of the χ^2 in units of the total measurement error (Fig. 6).
5. There are basically two procedures as far as the Higgs boson mass is concerned. In the first, M_{H} is assumed to be an external parameter to be varied arbitrarily between 60 GeV and 1 TeV with a central value of 300 GeV. The fit is repeated for the three values and the differences in the results are assigned to the parameters as asymmetrical errors. The alternative procedure consists in creating a $\chi^2(M_{\text{H}})$ curve, where for each value of M_{H} the minimum is found as a function of the remaining parameters, and in determining one-sided upper bounds on M_{H} from consideration of $\chi^2_{\text{min}} + \Delta\chi^2$. Actually, it has become customary to report the analysis in terms of $\ln M_{\text{H}}$. Indeed, to the first order, the radiative corrections in the SM are proportional to $\ln M_{\text{H}}$ (Veltman screening theorem). The most recent estimate gives $\ln(M_{\text{H}}/\text{GeV}) = 1.79^{+0.27}_{-0.28}$ and $M_{\text{H}} < 203$ GeV at 95% confidence level.

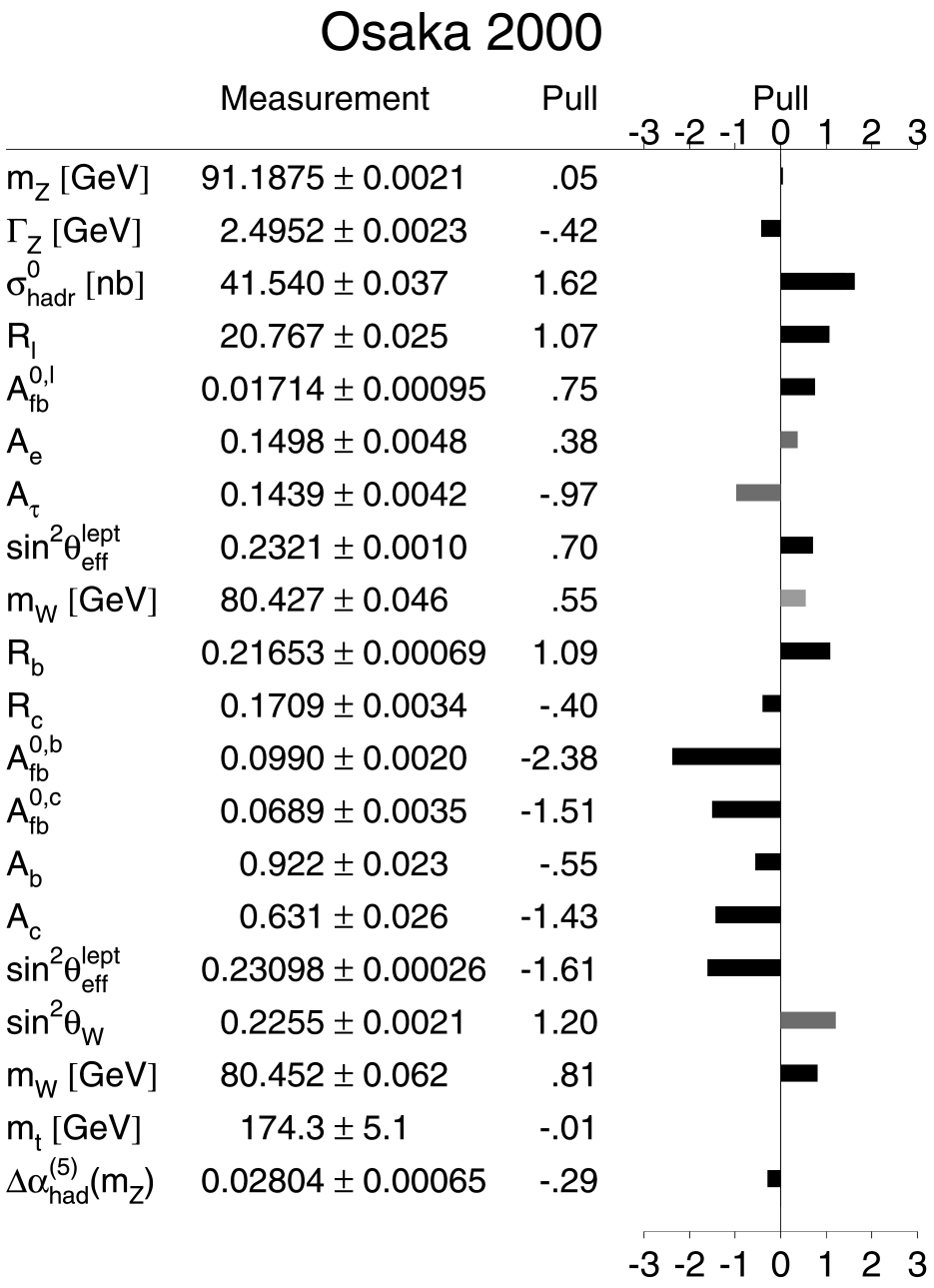


FIGURE 6 Pulls in the standard model. [Courtesy of the LEP Electroweak Working Group.]

IX. CONCLUSION

As mentioned by Peskin (2000), physics at its microscopic scale has a basis as rational as chemistry. Everything moves because we have a mechanism at work and this is encouraging for those aspects of fundamental interactions that are still unsolved.

If we look at the standard theory as we know it, one ingredient is still missing experimental detection. As Veltman (2000b) put it, the Higgs force may be the door to understanding other mysteries of particle physics. It may have enormous consequences for our understanding of this world, including the structure of the whole universe.

SEE ALSO THE FOLLOWING ARTICLES

ELECTRODYNAMICS, QUANTUM • ELECTROMAGNETICS • GREEN'S FUNCTIONS • PERTURBATION THEORY • QUANTUM CHROMODYNAMICS • UNIFIED FIELD THEORIES

BIBLIOGRAPHY

Aoki, K. I., Hioki, Z., Kawabe, R., Konuma, M., and Muta, T. (1982). *Suppl. Prog. Theor. Phys.* **73,** 1.

Bardin, D., and Passarino, G. (1999). "The Standard Model in the Making: Precision Study of the Electroweak Interactions," Clarendon, Oxford.

Bardin, D., Christova, P. C., and Fedorenko, O. M. (1980). *Nucl. Phys. B* **175,** 235.

Bardin, D., Christova, P. C., and Fedorenko, O. M. (1982). *Nucl. phys. B* **197,** 1.

Bardin, D., Bilenky, M. S., Mitselmakher, G. V., Riemann, T., and Sachtwitz, M. (1989). *Z. Phys. C* **44,** 493.

Böhm, M., Hollik, W., and Spiesberger, H. (1986). *Fortschr. Phys.* **34,** 687.

Consoli, M., LoPresti, S., and Maiani, L. (1983). *Nucl. Phys. B* **233,** 474.

Czarnecki, A., and Kühn, J. H. (1996). *Phys. Rev. Lett.* **77,** 3955.

Degrassi, G., and Gambino, P. (2000). *Nucl. Phys. B* **567,** 3.

Degrassi, G., and Sirlin, A. (1991). *Nucl. Phys. B* **352,** 342.

Degrassi, G., Fanchiotti, S., and Sirlin, A. (1991). *Nucl. Phys. B* **351,** 49.

Degrassi, G., Gambino, P., and Vicini, A. (1996). *Phys. Lett. B* **383,** 219.

Degrassi, G., Gambino, P., and Sirlin, A. (1997). *Phys. Lett. B* **394,** 188.

Fleischer, J., and Jegerlehner, F. (1981). *Phys. Rev. D* **23,** 2001.

Harlander, R., Seidensticker, T., and Steinhauser, M. (1998). *Phys. Lett. B* **426,** 125.

Hollik, W. (1990). *Fortschr. Phys.* **38,** 165.

Hollik, W., and Timme, H. J (1986). *Z Phys. C* **33,** 125.

Jegerlehner, F. (1986). *Z. Phys. C* **32,** 425.

Kobayashi, M., and Maskawa, T. (1973). *Progr. Theor. Phys.* (Kyoto) **49,** 652.

Marciano, W. J., and Sirlin, A. (1981). *Phys. Rev. Lett.* **46,** 163.

Passarino, G. (1991). *Nucl. Phys. B* **361,** 351.

Passarino, G., and Pittau, R. (1989). *Phys. Lett. B* **228,** 89.

Passarino, G., and Veltman, M. (1979). *Nucl. Phys. B* **160,** 151.

Passarino, G., and Veltman, M. (1990). *Phys. Lett. B* **237,** 537.

Peskin, M. (2000). Tampere 99, *High Energy Physics*, 319–342, Institute of Physics Publishing.

't Hooft, G. (2000). *Rev. Mod. Phys.* **72,** 333.

't Hooft, G., and Veltman, M. (1979). *Nucl. Phys. B* **153,** 365.

Veltman, M. (1977). *Nucl. Phys. B* **123,** 89.

Veltman, M. (2000a). *Rev. Mod. Phys.* **72,** 341.

Veltman, M. (2000b). Speech at the official ceremony for LEP fest 2000.



Green's Functions

G. Rickayzen

University of Kent at Canterbury

I. Green's Theorem and Green's Functions in Electrostatics
II. Hermitian Operators and Green's Functions
III. Examples of Green's Functions
IV. Conversion from Differential to Integral Equation
V. Green's Functions in the Complex Plane
VI. Perturbation Theory
VII. Feynman Diagrams
VIII. Field Theory

GLOSSARY

Dirac delta function Improper function $\delta(x)$ that satisfies the equations $\delta(x) = 0$, $x \neq 0$, $\int_{-\infty}^{\infty} dx\, \delta(x) = 1$, and $\int_{-\infty}^{\infty} dx' f(x')\delta(x - x') = f(x)$.

Eigenvalues and eigenfunction Eigenvalues α_n and eigenfunctions $\psi_n(x)$ of an operator $\mathcal{L}$ are the allowed solutions to the equation $\mathcal{L}\psi_n(x) = \alpha_n \psi_n(x)$.

Hermitian operator Operator $\mathcal{H}$ in an n-dimensional space x, which, for any two complex functions $f(x)$, $g(x)$, satisfies $[\int dx\ f(x)^* \mathcal{H} g(x)]^* = \int dx\ g(x)^* \mathcal{H} f(x)$.

Linear homogeneous boundary conditions Boundary conditions satisfied by the solution $\psi(x_1, \ldots, x_N)$ of a partial differential equation are linear and homogeneous if they can be written in the form $D_i \psi(x) = 0$, where the operators D_i are linear.

GREEN'S FUNCTION relates the output or linear response of a system β to an input or weak disturbance α through an equation of the form

$$\beta(x_1, \ldots, x_N) = \int dx'_1, \ldots, dx'_N G(x_1, \ldots, x_N; x'_1, \ldots, x'_N) \times \alpha(x'_1, \ldots, x'_N). \quad (1)$$

The input α and the output β are both functions of the variables $(x_1, \ldots, x_N)$, which are often position and time coordinates. The function $G(x_1, \ldots, x_N; x'_1, \ldots, x'_N)$ is a Green's function for the system and is sometimes called the influence function.

Green's functions are exploited in physics in four main ways:

1. To relate the response of a system to the cause, as in Eq. (1)
2. To solve ordinary and partial differential equations when certain boundary conditions are imposed (the boundary-value problem)
3. To convert a linear or partial differential equation into an integral equation that incorporates given boundary conditions
4. To expand the solutions of both linear and nonlinear equations in powers of a small parameter.

I. GREEN'S THEOREM AND GREEN'S FUNCTIONS IN ELECTROSTATICS

In his famous paper of 1828 entitled "The Application of Mathematical Analysis to the Theories of Electricity and Magnetism," George Green introduced the theorem and functions that now bear his name and used them to solve Laplace's and Poisson's equations of electrostatics. Poisson's equation relates the electrostatic potential $V(\mathbf{r})$ to the charge density present $\rho(\mathbf{r})$ through

$$\nabla^2 V\mathbf{r} = \frac{-\rho(\mathbf{r})}{\varepsilon_0}. \tag{2}$$

Laplace's equation is the special case of this when no charge density is present.

Green's theorem states that for any two continuous and differentiable functions $u(\mathbf{r})$ and $v(\mathbf{r})$ in a region of space bounded by a surface or surfaces S,

$$\int [u(\mathbf{r})\nabla^2 v(\mathbf{r}) - v(\mathbf{r})\nabla^2 u(\mathbf{r})]\,\mathrm{d}v = \int (u\nabla v - v\nabla u)\cdot dS. \tag{3}$$

Here, the first integral is taken over the whole of the bounded space and the second over the whole of the bounding surface. For $v(\mathbf{r})$, Green took the potential $V(\mathbf{r})$ which satisfies Eq. (2); and for $u(\mathbf{r})$, he took the potential due to a unit point charge at the position $\mathbf{r}'$ in free space. As the latter potential depends on both the field point $\mathbf{r}$ and the position of the charge $\mathbf{r}'$, we write

$$u(\mathbf{r}) \equiv G(\mathbf{r}, \mathbf{r}'). \tag{4}$$

This is the Green's function for the problem. Since it is the potential due to a unit charge at $\mathbf{r}'$, it satisfies Poisson's equation for this case, namely,

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = \frac{-\delta(\mathbf{r} - \mathbf{r}')}{\varepsilon_0}, \tag{5}$$

where the charge density $\delta(\mathbf{r} - \mathbf{r}')$ satisfies

$$\delta(\mathbf{r} - \mathbf{r}') = 0, \qquad \mathbf{r} \neq \mathbf{r}', \tag{6}$$

because no charge is present other than at $\mathbf{r}'$, and

$$\int d^3r\, \delta(\mathbf{r} - \mathbf{r}') = 1 \tag{7}$$

for any volume that includes $\mathbf{r}'$, because the total charge present is 1. The function $\delta(\mathbf{r} - \mathbf{r}')$ is Dirac's delta function in three dimensions.

It follows from the definition of the delta function that

$$\int V(\mathbf{r})\nabla^2 G(\mathbf{r}, \mathbf{r}')\, dv = \frac{1}{\varepsilon_0}\int V(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')\, dv = \frac{-V(\mathbf{r}')}{\varepsilon_0}. \tag{8}$$

If, therefore, one substitutes for $u(\mathbf{r})$ and $v(\mathbf{r})$ in Eq. (3), one obtains

$$\frac{V(\mathbf{r}')}{\varepsilon_0} = \int \frac{\rho(\mathbf{r})}{\varepsilon_0} d^3r G(\mathbf{r}, \mathbf{r}') + \int [G(\mathbf{r}, \mathbf{r}')\nabla V(\mathbf{r}) - V(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}')] \cdot dS. \tag{9}$$

Equation (9) yields the potential at any position in terms of the given charge density and the potential and its normal gradient on the bounding surface.

Choosing $G(\mathbf{r}, \mathbf{r}')$ to satisfy different given boundary conditions, one is led to a number of important special cases.

1. If the boundary tends to infinity and $V(\mathbf{r}) \to 0$ as $\mathbf{r} \to \infty$, one can choose $G(\mathbf{r}, \mathbf{r}') \to 0$ as $\mathbf{r} \to \infty$. Then the surface integral vanishes and

$$V(\mathbf{r}') = \int \rho(\mathbf{r})G(\mathbf{r}, \mathbf{r}')\, d^3r. \tag{10}$$

In this case,

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}'|}, \tag{11}$$

the Coulomb potential.

2. If $G(\mathbf{r}, \mathbf{r}')$ is chosen so that

$$G(\mathbf{r}, \mathbf{r}') = 0, \qquad \mathbf{r} \in S, \tag{12}$$

then

$$V(\mathbf{r}') = \int \rho(\mathbf{r})G(\mathbf{r}, \mathbf{r}')\, d^3r - \int V(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}') \cdot dS \tag{13}$$

These examples show that whatever the charge density and the values of V and the normal derivative of V on the bounding surface, it is possible to use Green's functions to write the potential explicitly in terms of these quantities. Equation (10) is in the form of Eq. (1), where the input is the charge density and the potential is the output. Equation (13) is an example of the use of Green's functions in the solution of a boundary-value problem in physics and engineering.

II. HERMITIAN OPERATORS AND GREEN'S FUNCTIONS

For a wide class of differential operators, it is possible to derive a generalization of Green's theorem and Green's functions and from these construct solutions of the relevant differential equations. In physics, an important subset of these operators includes the Hermitian operators $\mathcal{H}$. When the boundary conditions are linear and homogeneous, the Green's functions take on a simple form.

Consider the eigenvalue equations

$$\mathcal{H}\psi_n(x) = E_n\psi_n(x), \tag{14}$$

subject to linear homogeneous boundary conditions; here, the coordinate x may stand symbolically for many coordinates $x_1, \ldots, x_N$. If $\mathcal{H}$ is Hermitian, the eigenvalues E_n of the equation are all real. We assume additionally that the corresponding eigenfunctions $\psi_n(x)$ form a complete set; that is, any allowed function $f(x)$ can be expanded linearly in terms of the eigenfunctions as

$$f(x)\sum_n f_n\psi_n(x), \tag{15}$$

where the coefficients f_n are unique. The different eigenfunctions $\psi_n(x)$ may be chosen to satisfy the relations

$$\int dx\,\psi_m^*(x)\psi_n(x) = \delta_{m,n}, \tag{16}$$

where

$$\delta_{m,n} = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases} \tag{17}$$

In that case, the coefficients in Eq. (15) are given by

$$f_n \int \psi_n^*(x)\, f(x)\, dx, \tag{18}$$

where the integral is over the N-dimensional space $x_1, \ldots, x_N$. Further,

$$\sum_n \psi_n^*(x)\psi_n(x') = \delta(x - x'). \tag{19}$$

The inhomogeneous equation

$$(H - E)g(x) = -f(x) \tag{20}$$

can now be solved by substituting for $f(x)$ from Eq. (15) and for $g(x)$ from

$$g(x) = \sum_n g_n\psi_n(x). \tag{21}$$

Thus,

$$(H - E)\sum_n g_n\psi_n(x) = -\sum_n f_n\psi_n(x) \tag{22}$$

If one now uses Eq. (14), multiplies by $\psi_m^*(x)$, and integrates over x, one obtains

$$(E_m - E)g_m = -f_m. \tag{23}$$

Hence, provided that $E \neq E_n$,

$$g(x) = -\sum_n \frac{f_n\psi_n(x)}{E_n - E} = -\sum_n \int \frac{dx'\psi_n^*(x')\, f(x')\psi_n(x)}{E_n - E}$$

$$= \int dx' G(x, x')\, f(x'), \tag{24}$$

where

$$G(x, x') = \sum_n \frac{\psi_n(x)\psi_n^*(x')}{E - E_n}. \tag{25}$$

Equation (24) is of the standard form in Eq. (1) with the Green's function given by Eq. (25).

The equation satisfied by the Green's function is

$$(H - E)G(x, x') = -\sum \psi_n(x)\psi^*(x') = -\delta(x - x'), \tag{26}$$

where Eq. (19) has been used. Equation (26) is of the standard form of a Green's function equation and shows that G provides the response to a unit source. Note that in this case the Green's function obeys the symmetry relation

$$G(x, x')^* = G(x', x). \tag{27}$$

Much use is made of these results in wave mechanics because the Hamiltonian operator, which governs the behavior of microscopic systems, is Hermitian. When $\mathcal{H}$ in Eq. (14) is the Hamiltonian, the equation is Schrödinger's equation for the stationary states of the system.

Equation (20) expresses $f(x)$ explicitly in terms of $g(x)$ and its derivatives, while Eq. (24) expresses $g(x)$ in terms of $f(x)$. Thus, the Green's function inverts the relationship and is an inverse of the differential operator $(\mathcal{H} - E)$. Since the delta function plays the role of a unit operator in function space, Eq. (26) embodies this inverse relationship explicitly.

III. EXAMPLES OF GREEN'S FUNCTIONS

A. Retarded Electromagnetic Potential

A varying current density $\mathbf{j}(\mathbf{r}, t)$ generates an electromagnetic vector potential $\mathbf{A}(\mathbf{r}, t)$, which satisfies

$$\nabla^2\mathbf{A} - \frac{\partial^2\mathbf{A}}{c^2\partial t^2} = -\mu\mathbf{j}. \tag{28}$$

This is the generalization of Eq. (2) for time-varying phenomena. The related Green's function satisfies the equation

$$\left(\nabla^2 - \frac{\partial^2}{c^2 \partial t^2}\right) G(\mathbf{r}, t; \mathbf{r}', t') = -\delta(\mathbf{r} - \mathbf{r}')\delta(t - t'). \quad (29)$$

For physical solutions, the boundary conditions for the Green's function are

$$G(\mathbf{r}, t; \mathbf{r}', t') \to 0 \qquad \text{as } |\mathbf{r} - \mathbf{r}|' \to \infty, \quad (30)$$

$$G(\mathbf{r}, t; \mathbf{r}', t') = 0 \qquad t < t'. \quad (31)$$

The latter condition ensures that the effect follows the cause.

Since Eq. (29) is translationally invariant in space and time, it can be solved using the Fourier transform:

$$\tilde{G}(\mathbf{k}, \omega) = \int d^3 r \int dt\, e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')+i\omega(t-t')} G(\mathbf{r} - \mathbf{r}', t - t'). \quad (32)$$

Then

$$\left(k^2 - \frac{\omega^2}{c^2}\right)\tilde{G}(\mathbf{k}, \omega) = 1 \quad (33)$$

$$\tilde{G}(\mathbf{k}, \omega) = \frac{1}{k^2 - \omega^2/c^2} \quad (34)$$

and

$$G(\mathbf{r}, t; \mathbf{r}', t') = \frac{1}{(2\pi)^3} \int \frac{d^3 k\, d\omega\, e^{+i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')-i\omega(t-t')}}{k^2 - \omega^2/c^2}. \quad (35)$$

Because of the singularity in the integrand at $\omega = \pm ck$, the integral is ambiguous. However, the Green's function must satisfy Eq. (31), and this requirement removes ambiguity. To ensure that Eq. (31) is met, one replaces ω by $\omega + i\varepsilon$, where ε is small and tends to zero at the end of the calculation. The result, after integration, is

$$G(\mathbf{r}, t, \mathbf{r}', t') = \frac{c}{4\pi |\mathbf{r} - \mathbf{r}'|} \delta[|\mathbf{r} - \mathbf{r}'| - c(t - t')]. \quad (36)$$

Then, for any current density, the vector potential is given by

$$A(\mathbf{r}, t) = \frac{c\mu}{4\pi} \int \frac{d^3 r'\, dt'}{|\mathbf{r} - \mathbf{r}'|} \delta[|\mathbf{r} - \mathbf{r}'| - c(t - t')]\mathbf{j}(\mathbf{r}', t')$$

$$= \frac{c\mu}{4\pi} \int \frac{d^3 r'}{|\mathbf{r} - \mathbf{r}'|} \mathbf{j}\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right). \quad (37)$$

Because the current is evaluated at the earlier time $t - (1/c)|\mathbf{r} - \mathbf{r}'|$, this is known as the retarded potential. The time difference is simply the time taken for light to travel from the current point to the field one.

B. Free Nonrelativistic Particle

Schrödinger's stationary equation for a free particle of mass m is

$$\left[-\frac{\hbar^2}{2m}\nabla^2 - E\right]\psi(\mathbf{r}) = 0, \quad (38)$$

and the corresponding Green's function equation is

$$\left(-\frac{\hbar^2}{2m}\nabla^2 - E\right) G(\mathbf{r}, \mathbf{r}', E) = -\delta(\mathbf{r} - \mathbf{r}'). \quad (39)$$

Again, the equation is translationally invariant and can be solved with the use of the Fourier transform:

$$\tilde{G}(\mathbf{k}, E) = \int d^3 r\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} G(\mathbf{r}, \mathbf{r}', E). \quad (40)$$

This is equivalent to the expansion in terms of eigenfunctions given in the last section because, in this case, the eigenfunctions are the plane waves $\exp(i\mathbf{k}\cdot\mathbf{r})$. The result is

$$\tilde{G}(\mathbf{k}, E) = \frac{1}{E - \hbar^2 k^2/2m} \quad (41)$$

and

$$G(\mathbf{r}_1, \mathbf{r}', E) = \frac{1}{(2\pi)^3} \int d^3 k \frac{e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}}{E - \hbar^2 k^2/2m}. \quad (42)$$

This result is in the form of Eq. (25), except that, because the spectrum is continuous, the sum is replaced by an integral. This again leads to an ambiguity in the definition of the integral that can only be resolved by reference to the boundary conditions imposed by a particular physical problem. For example, in scattering problems, one often requires $G(\mathbf{r}, \mathbf{r}', E)$ to represent outgoing waves as $\mathbf{r} \to \infty$. This is ensured if one replaces E by $E + i\varepsilon$ and allows $\varepsilon \to 0$ at the end of the calculation. The result is

$$G(\mathbf{r}, \mathbf{r}', E) = \frac{e^{ik_0|\mathbf{r}-\mathbf{r}'|}}{4\pi |\mathbf{r} - \mathbf{r}'|}, \qquad k_0 = \sqrt{\frac{2mE}{\hbar^2}}. \quad (43)$$

C. The Dirac Equation

This equation governs the behavior of a free electron and other spin-$\frac{1}{2}$ particles and takes the form

$$\left[\gamma_\mu \frac{\partial}{\partial x_\mu} + (mc/\hbar)\right]\psi \equiv \left[\frac{\gamma_4}{ic}\frac{\partial}{\partial t} + \gamma \cdot \nabla + K\right]\psi = 0, \quad (44)$$

where the quantities γ_μ ($\mu = 1, 2, 3, 4$) are 4×4 matrices that satisfy

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu}. \quad (45)$$

The wave functions ψ are column vectors with four components; the corresponding Green's function satisfies

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + K\right) G(x, x') = -\delta(x - x'), \quad (46)$$

where x_μ stands for the coordinate and the time variables x_1, x_2, x_3, and $x_4 = ict$. Since ψ has four components, any source will have four components, and the Green's function, which relates source to response, will be a 4×4 matrix. The right-hand side of Eq. (46) is assumed to be multiplying the unit 4×4 matrix.

As Eq. (46) is again translationally invariant, the solution may be found using Fourier transforms. The result is (with $k_4 = \omega/c$)

$$\tilde{G}(k_\mu) = \frac{1}{\gamma_4 k_4 - i\gamma \cdot \mathbf{k} + K} = \frac{\gamma_4 k_4 - i\gamma \cdot \mathbf{k} - K}{\mathbf{k}^2 + K^2 - k_4^2} \quad (47)$$

and shows explicitly that $\tilde{G}(k_\mu)$ is a 4×4 matrix.

IV. CONVERSION FROM DIFFERENTIAL TO INTEGRAL EQUATION

Sometimes an equation of physics may be written in the form

$$\mathcal{L}y(x) = -f[x, y(x)], \quad (48)$$

where $\mathcal{L}$ is a differential operator and the right-hand side depends on $y(x)$ and may even be a nonlinear function of $y(x)$. If $y(x)$ satisfies linear homogeneous boundary conditions and if one can find a Green's function of $\mathcal{L}$ that satisfies these conditions, then Eq. (48) can be rewritten in the form

$$y(x) = \int dx' G(x, x') f[x', y(x')]. \quad (49)$$

The solutions of this integral equation automatically satisfy Eq. (48) and the boundary conditions.

V. GREEN'S FUNCTIONS IN THE COMPLEX PLANE

It was seen in Section II that the Green's functions for Hermitian operators can be written in the form of Eq. (25). It we allow for the possibility of a continuous spectrum, this can be written

$$G(x, x', E) = \int_{-\infty}^{\infty} \frac{dy}{E - y} A(x, x', y), \quad (50)$$

where

$$A(x, x', y) = \sum_n \psi_n(x)\psi_n^*(x')\delta(E_n - y) + \int dE' \psi_{E'}(x)\psi_{E'}^*(x')\delta(E' - y). \quad (51)$$

The function $A(x, x', y)$ is called the spectral function of the Green's function.

Equation (50) provides a definition for the Green's function as a function of the complex variable E. It is analytic throughout the complex plane except at the eigenvalues of the Hermitian operator $\mathcal{H}$. The singular points of $G(x, x', E)$ are therefore these eigenvalues.

If an eigenvalue E_n is nondegenerate, the residue of $G(x, x, E)$ at E_n is $\psi_n^*(\mathbf{r})\psi_n(\mathbf{r})$. Since nondegenerate eigenfunctions can be chosen to be real, this determines $\psi_n(\mathbf{r})$. In fact, all the physical properties can be determined once $G(x, x', E)$ is known. Thus, it is possible to work entirely with Green's functions rather than with eigenfunctions and eigenvalues. This is particularly valuable in problems involving many degrees of freedom.

When the eigenvalues have a continuous spectrum, the Green's function defined in Eq. (50) is ambiguous. We have already seen this in the special cases discussed in Section III. In fact, the most general inverse of $(E - y)$ is

$$\frac{1}{E - y} = \frac{P}{E - y} + \alpha(y)\delta(E - y), \quad (52)$$

where $\alpha(y)$ is an arbitrary function and where the symbol P indicates that the principal value of any integral including this term is to be taken. Thus, every function $\alpha(y)$ defines a different solution of Eq. (26). The actual function $\alpha(y)$ required is determined by the boundary conditions as illustrated by the examples given in Section III.

In problems of quantum statistical mechanics, it turns out to be easiest to calculate the Green's function, Eq. (50), at the discrete imaginary points $E = i2\pi lkT$ ($l = 0, \pm 1, \pm 2, \ldots$) if the Green's function describes the behavior of particles that obey Bose statistics, or the points $E = i(2l + 1)\pi kT$ if it describes particles that obey Fermi–Dirac statistics. By analytic continuation, the complete Green's function can be determined once these values are known.

VI. PERTURBATION THEORY

The conversion of a differential equation to an integral equation provides a useful basis for obtaining a solution as a power series, a perturbation series. Consider Schrödinger's equation for a particle of mass m moving in a potential $u(\mathbf{r}) + \lambda v(\mathbf{r})$:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + u(\mathbf{r}) + \lambda v(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (53)$$

where λ is small and the Green's function $G_0(r, r', E)$ when λ is zero is known. Then, the equation can be converted to

$$\psi(\mathbf{r}) = \phi(\mathbf{r}) + \lambda \int d^3r' G_0(\mathbf{r}, \mathbf{r}', E) v(\mathbf{r}')\psi(\mathbf{r}'), \quad (54)$$

where $\phi(\mathbf{r})$ is the corresponding solution when λ is zero. One obtains a series solution by writing

$$\psi(\mathbf{r}) = \sum_{j-0}^{\infty} \lambda^j \psi^{(j)}(\mathbf{r}), \qquad \psi^{(0)}(\mathbf{r}) = \phi(\mathbf{r}). \quad (55)$$

If this series is substituted in Eq. (54), one finds a recurrence relation for the coefficients, namely,

$$\psi^{(j+1)}(\mathbf{r}) = + \int d^3r' G_0(\mathbf{r}, \mathbf{r}', E) v(\mathbf{r}') \psi^{(j)}(\mathbf{r}'). \quad (56)$$

Provided that the series converges (and this will usually be the case if the problem is properly defined physically), one has a formal solution to the equation. Often, when λ is small, the first-order term in λ will suffice. This is the Born approximation for the wave function.

This method can also be used to produce a power series for the Green's function $G(\mathbf{r}, \mathbf{r}', E)$ that satisfies

$$\left[-\frac{\hbar^2}{2m} \nabla^2 u(\mathbf{r}) + \lambda v(\mathbf{r}) - E \right] G(\mathbf{r}, \mathbf{r}', E) = -\delta(\mathbf{r} - \mathbf{r}'). \quad (57)$$

The corresponding integral equation is

$$G(\mathbf{r}, \mathbf{r}', E) = G_0(\mathbf{r}, \mathbf{r}', E) + \lambda \int d^3r' G_0(\mathbf{r}, \mathbf{r}', E) v(\mathbf{r}') G(\mathbf{r}, \mathbf{r}', E), \quad (58)$$

and a powers series solution of the form

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_j \lambda^j G_j(\mathbf{r}, \mathbf{r}', E) \quad (59)$$

is easily generated from the zero-order term $G_0(\mathbf{r}, \mathbf{r}', E)$.

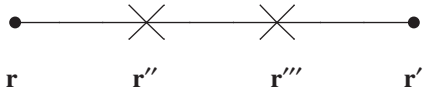
VII. FEYNMAN DIAGRAMS

In an elegant series of papers published in 1949 and for which he ultimately shared the Nobel Prize, Richard Feynman derived a perturbation expansion for the solutions of problems in quantum electrodynamics and showed how the solutions could be represented by diagrams that are simple to construct and make the result memorable. These diagrams now bear his name and can be illustrated by reference to the solution of Eq. (58).

The first few terms in the series expansion of Eq. (58) are shown explicitly in the result

$$\begin{aligned} G(\mathbf{r}, \mathbf{r}', E) = {} & G_0(\mathbf{r}, \mathbf{r}', E) \\ & + \lambda \int d^3r'' G_0(\mathbf{r}, \mathbf{r}'', E) v(\mathbf{r}'') G_0(\mathbf{r}'', \mathbf{r}, E) \\ & + \lambda^2 \int d^3r'' \, d^3r''' G_0(\mathbf{r}, \mathbf{r}'', E) v(\mathbf{r}'') \\ & \times G_0(\mathbf{r}'', \mathbf{r}''', E) v(\mathbf{r}''') G_0(\mathbf{r}''', \mathbf{r}', E) + \cdots . \end{aligned} \quad (60)$$

This is represented as follows. The Green's function $G_0(\mathbf{r}, \mathbf{r}', E)$ is represented by a straight line from the point $\mathbf{r}$ to the point $\mathbf{r}'$ and is conceived as the propagator of a particle from the point $\mathbf{r}$ to the point $\mathbf{r}'$. The potential $\lambda v(\mathbf{r})$ is represented by a cross (a vertex) at the point $\mathbf{r}$. Then, the second-order term in Eq. (60) is represented by the diagram

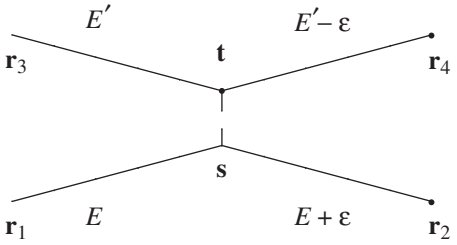


From the diagram, one can immediately construct the term in the series that it represents using the connection between lines and crosses (vertices) just given. In addition, one integrates over all the internal coordinates, $\mathbf{r}''$ and $\mathbf{r}'''$ in this case. To obtain the whole series, one draws all diagrams in which a straight line connects points $\mathbf{r}$ and $\mathbf{r}'$ and on which there can be any number of vertices. For each such diagram, one writes down the product of Green's functions and potential terms that the lines and vertices, respectively, represent, and one integrates over all internal coordinates. The sum of the contributions from all diagrams constructed according to these rules is then the expansion for $G(\mathbf{r}, \mathbf{r}', E)$.

The diagram illustrated can be interpreted in the following way. In propagating from $\mathbf{r}$ to $\mathbf{r}'$ in the presence of the full potential, the particle may propagate under the potential $u(\mathbf{r})$ from $\mathbf{r}$ to $\mathbf{r}''$, where it is scattered by the potential $\lambda v(\mathbf{r}'')$. It may then propagate from $\mathbf{r}''$ to $\mathbf{r}'''$, where it is scattered by the potential $\lambda v(\mathbf{r}''')$. Finally, it propagates to $\mathbf{r}'$. The total series combines all possible ways in which the particle can propagate from $\mathbf{r}$ to $\mathbf{r}'$ with scattering by $\lambda v(\mathbf{r})$ at any number of points between.

VIII. FIELD THEORY

At the present time, Green's functions find their widest applications in field theory, both in elementary particle physics and in the physics of condensed matter. The response of the system can be given in terms of an appropriate Green's function that can be calculated using perturbation theory. The various terms of the series can be represented by Feynman diagrams and there are well-defined rules for calculating the contribution of a particular diagram to the Green's function. An important difference between these Feynman diagrams and the ones described in the previous section arises from the form of the interaction. In field theory, the basic interaction is not with an external potential but between particles. This changes the vertices in the diagrams. For example, if the interaction is a direct interaction, $V(\mathbf{r} - \mathbf{r}')$, between the two particles at positions $\mathbf{r}$ and $\mathbf{r}'$, the vertex is

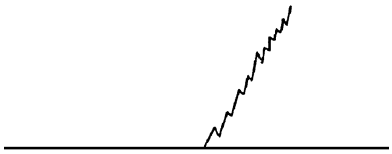


Note that energy can be exchanged between the two particles but it is conserved. The contribution of the diagram is

$$\int G_0(\mathbf{r}_1, \mathbf{s}, E)G_0(\mathbf{s}, \mathbf{r}_2, E + \varepsilon)V(\mathbf{t} - \mathbf{s})G_0(\mathbf{r}_3, \mathbf{t}, E') \times G_0(\mathbf{t}, \mathbf{r}_4, E' - \varepsilon)\, d^3t\, d^3s. \quad (61)$$

The Green's functions $G_0(\mathbf{r}_3, \mathbf{r}', E)$ are the appropriate Green's functions for the particles in the absence of the interaction $V(\mathbf{r})$.

Sometimes the interaction gives rise to the emission or absorption of a particle. For example, in elementary particle physics, it may relate to the emission or absorption of a photon or meson. In this case, the interaction is represented by the vertex



where the wavy line represents the Green's function of the emitted or absorbed particle.

Whatever the form of the interaction, one can in principal calculate any Green's function by drawing all Feynman diagrams consistent with the interaction and adding together all their contributions to give an infinite series. In practice, this series cannot be summed and it is necessary to resort to an approximation in which only a subset of all the contribution is included. In general, an infinite subset is required. One important and successful approach to field theory has been to identify sets of diagrams that provide valuable approximations to the Green's functions sought. These approximations depend on the particular problem under discussion.

SEE ALSO THE FOLLOWING ARTICLES

CALCULAS • DIFFERENTIAL EQUATIONS, ORDINARY • DIFFERENTIAL EQUATIONS, PARTIAL • ELECTROMAGNETICS • FIELD THEORY AND THE STANDARD MODEL • INTEGRAL EQUATIONS • PERTURBATION THEORY

BIBLIOGRAPHY

Barton, G. (1989). "Elements of Green's Functions and Propagation: Potentials, Diffusion and Waves," Clarendon Press, Oxford, U.K.

Economou, E. N. (1990). "Green's Functions in Quantum Physics," 2nd ed., Springer-Verlag, Berlin.

Gonis, A. (1992). "Green's Functions for Ordered and Disordered Systems," North-Holland, Amsterdam.

Mahan, G. D. (1990). "Many-Particle Physics," 2nd ed., Plenum, New York.

Morse, P. M., and Feshbach, H. (1953). "Methods of Theoretical Physics," Vol. 2, McGraw-Hill, New York.

Rickayzen, G. (1984). "Green's Functions and Condensed Matter," Academic Press, New York.

Stakgold, I. (1998). "Green's Functions and Boundary Value Problems," 2nd ed., Wiley, New York.

Van Bladel, J. (1964). "Electromagnetic Fields," McGraw-Hill, New York.

Zavialov, O. I. (1990). "Renormalized Quantum Field Theory," Klewer, Dordrecht, The Netherlands.



Heavy Ions (High-Energy Physics)

Karel Šafařík
CERN, European Laboratory for Particle Physics

I. Physics Motivation
II. QGP Signatures
III. Experimental Situation
IV. Interpretation
V. Future Outlook

GLOSSARY

Baryon Particle with nonzero baryon number, its quantum numbers are carried by three valence quarks; the constituents of an atomic nucleus, protons and neutrons, are baryons.

Chemical potential Thermodynamical variable which measures the change of free energy of a system when changing the amount of some of the substance in the system, at constant temperature and pressure; in heavy-ion physics it is often used with respect to baryon number—baryon chemical potential, or quark flavor—quark chemical potential.

Color Property of quarks and gluons—the charge of strong interactions, analogous to electric charge in electromagnetic interactions; however, it has three different states for positive charge and three different states for negative charge.

Entropy State function of the system which describes the unavailability of the energy of that system to provide work; it is proportional to the logarithm of the number of possible microscopical configurations of the system consistent with the system's thermodynamical state; a closed system tends toward the most probable state, i.e., the one with maximum number of microscopical configurations, and therefore it tends to maximize its entropy (and be unavailable for work).

Gluon (g) Carrier of strong interaction, i.e., quantum of strong field which has a similar role to the photon in the electromagnetic field, but, unlike the photon, which is electrically neutral, the gluon itself carries a strong charge called color.

Hadron Strongly interacting particle; the hadron family of particles consists of baryons and mesons.

Meson Strongly interacting particle with zero baryon number, its quantum numbers are carried by a valence quark and an antiquark; examples of mesons are π mesons or K mesons.

Order parameter Quantity which characterizes the order of the system below the critical temperature (for

an order–disorder transition), where it has a nonzero value; above the critical temperature the order parameter vanishes; the way in which the order parameter approaches zero signals the order of the phase transition, an abrupt discontinuous change means a first-order phase transition (with a latent heat), a continuous change with a discontinuous first derivative at the critical temperature means a second-order phase transition; if the first derivative is also continuous then the phase transition is of a smooth crossover type.

Phase transition Transformation of system into another distinct ensemble, usually characterized by change in its symmetry; a phase transition can be classified by its order, according the behavior of the order parameter near the transition point.

Quantum chromodynamics (QCD) Theory of strong interactions, similar to quantum electrodynamics—a theory of electromagnetic interactions; the main difference between the two is due to different underlying symmetry which in turn reflects the difference in the structure of interacting charges.

Quark (q) Basic constituent of strongly interacting matter; there exist six types, or "flavors" of quarks, named: up, down, strange, charm, beauty, and top; however, only the first two constitute normal matter, protons and neutrons; quarks are under normal conditions confined inside baryons or mesons and cannot be therefore observed as free particles.

Quark–gluon plasma (QGP) New state of matter predicted by quantum chromodynamics at finite temperature; in this state quarks and gluons are no longer confined inside hadrons, and instead can move freely over an extended region.

Strangeness Quantum number carried by strange quarks; it is conserved in strong and electromagnetic interactions, but not in weak interactions.

THE MAIN MOTIVATION for colliding heavy ions at the maximum achievable energy is to observe hadronic matter under extreme conditions of high-energy density. Atoms of heavy materials (i.e., those with high atomic number) are subsequently stripped of their electrons and accelerated to ultra-relativistic energies. Then the beam of heavy ions is directed against a target made of similar, material (fixed-target experiments) or against another heavy-ion beam (colliding-beam experiments). In such collisions we aim to reach sufficient temperature and energy density to produce a new state of matter, a Quark–Gluon Plasma (QGP). In this new state the quarks are no longer confined inside individual hadrons, but they are free to move within the interaction region. Intuitively, it is clear that if we were able to compress hadronic matter beyond the limit at which the mean distance between hadrons is comparable to the hadron size (of the order of 1 fm), the usual picture of quark confinement cannot hold any more. The main goals of the study of heavy-ion collisions are to look for a transition between ordinary hadronic matter and QGP, to measure the properties of hadronic matter under extreme conditions and those of QGP, and to investigate the corresponding equations of state.

The quarks are confined (inside hadrons) under normal conditions by strong interactions. It is only the residual (i.e., not completely screened) strong field which binds together protons and neutrons inside an atomic nucleus. The theory of strong interactions, Quantum ChromoDynamics (QCD), becomes, however, too complicated when applied to these collective phenomena at small relative momentum, because the coupling constant in this regime is larger than unity. The only rigorous results up to now are obtained from numerical calculations on a lattice for zero baryon density. In order to get predictions for realistic experimental conditions, where at today's energies a substantial baryon charge remains even in the central region (i.e., region with relatively slow particles in the center-of-mass system of the colliding ions) due to the stopping of incoming nucleons, one has to use extrapolations or models based on the underlying theory.

In addition to the interest of heavy-ion physics for QCD itself, this topic is also relevant to other fields, namely, cosmology and astrophysics. First, the transition from QGP to ordinary hadronic matter—hadro-synthesis—must have occurred some 10^{-5} sec after the Big Bang when the temperature went below about 10^{12} K ($\simeq$100 MeV), as schematically shown in Fig. 1. Depending on the nature of this transition, critical fluctuations may have developed, which in turn may be traced in the further evolution of the

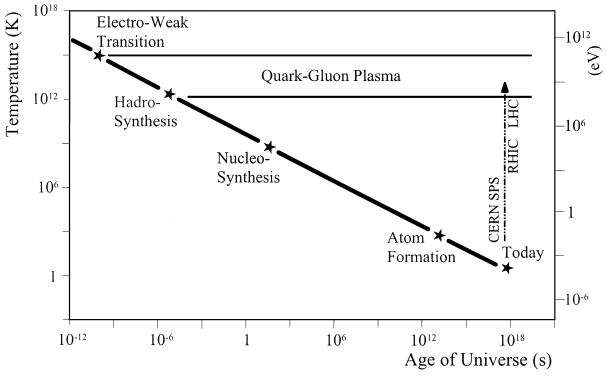


FIGURE 1 Temperature evolution of the universe as a function of time after Big Bang. The occurrence of various predicted transitions is indicated, as well as the path which heavy-ion experiments at different accelerators are exploring in order to recreate a QGP. (From CERN. Safarik, K. (2000). Heavy-ion physics. *In* "Proceedings of 1999 European School of High Energy," p. 267.)

universe. Second, the transition from hadronic matter to QGP may happen in the inner core of neutron stars. The presence of a QGP core may be detected by measuring the physical properties of the neutron stars, such as the mass, the luminosity, the surface temperature, and the revolution frequency.

We shall give a brief overview of physics motivations for the study of heavy-ion collisions in the next section. The expected signatures of a phase transition will be discussed in Section II. In the absence of rigorously calculable predictions for today's experimental conditions, this field of physics is naturally more driven experimentally than by theory. We will review the basic experimental results achieved at high energies in Section III and, in Section IV we will give their interpretation. The last section is devoted to the future experimental programs of heavy-ion colliders. First, the experiments just starting at the Relativistic Heavy-Ion Collider (RHIC) at Brookhaven National Laboratory (BNL) will be mentioned, and then the ALICE (A Large Ion Collider Experiment) at the Large Hadron Collider (LHC) machine, under construction at CERN, will be described.

We use here the system of units defined by equalizing universal physical constants to unity. We set the speed of light $c = 1$, Planck constant $\hbar = 1$, and Boltzmann constant $k = 1$. After this we need only one dimension. We choose to use the dimension of energy, namely, the energy of electron accelerated by electrostatic potential of 1 V, i.e., eV, and its multiple, MeV (10^6 eV), GeV (10^9 eV), etc. In this system of units mass, momentum, and temperature have the dimension of energy, i.e., eV; length and time have the dimension of inverse energy, i.e., eV^{-1}; speed, angular momentum, and entropy are dimensionless; pressure has the dimension of energy to the fourth power, i.e., eV^4; etc. The mass of the nucleon (proton or neutron) in these units is a little below 1 GeV and its size about 4 GeV^{-1}. In addition, for length and time the metric unit 1 fm ($=10^{-15}$ m $\simeq 5$ GeV^{-1}) is also often used.

I. PHYSICS MOTIVATION

A. QCD Vacuum and Symmetries

There exists a simple argument to explain why the QCD vacuum is more complicated than the Quantum ElectroDynamics (QED) one. In any quantum theory due to quantum fluctuations a pair of oppositely charged particles (i.e., in a singlet state) can pop up from a vacuum; charge means electric charge e in QED, and color charge g_s in QCD. The relative momentum p acquired by these two particles and their separation in space r are restricted by the uncertainty relation: $p \cdot r \gtrsim 1$. Therefore, if they are separated by a distance r, their minimal kinetic energy should be $E_{\mathrm{kin}} = p \simeq 1/r$, where we neglect their masses. The potential energy between the point-like charges is given by $E_{\mathrm{pot}} = -q^2/(4\pi r)$, where q is either electric charge $q = e$ or strong charge $q = g_s$. Then for the total energy of the pair we obtain

$$E_{\mathrm{pair}} = E_{\mathrm{kin}} + E_{\mathrm{pot}} = \frac{1}{r} \cdot \left(1 - \frac{q^2}{4\pi}\right). \qquad (1)$$

In QED this estimate is correct for any distance r, down to the Planck scale ($\simeq 10^{-20}$ fm), but in QCD the validity of this expression is restricted to small distances r (below a few fm).

Let us first look at what is happening in QED. The square of the electric charge $q^2 = e^2 = 4\pi\alpha_{\mathrm{em}}$ is at large distances determined by the well-known electromagnetic fine-structure constant $\alpha_{\mathrm{em}} \simeq 1/137$. From a large distance we, however, do not see the "true" electric charge of an electron because there are many other e^+e^- pairs in the vacuum around. These pairs tend to be in the configuration where the opposite charge of the pair is closer to the observed electron and the like-sign charge is further from it. The vacuum near the electron is polarized, which effectively lowers the observed charge of the electron. When going to shorter distances, the electromagnetic constant is therefore increasing, due to less efficient screening by vacuum polarization. This rise is, however, relatively slow. For example, at the electroweak scale, i.e., at 100 GeV or $r \simeq 2 \cdot 10^{-3}$ fm, the electromagnetic running constant rises to $\alpha_{\mathrm{em}} = 1/128$, but even at the Planck scale, i.e., at 10^{19} GeV or $r \sim 10^{-20}$ fm, its value will still be small, only about $\alpha_{\mathrm{em}} = 1/76$. So, in QED the numerical factor $1 - q^2/(4\pi) = 1 - \alpha_{\mathrm{em}}$ in Eq. (1) varies very little, between 0.987 and 0.993, when changing the pair separation between the Planck scale and infinity. As a consequence, when e^+e^- pairs pop up from a vacuum, they will be unstable because their energy is always positive (see Fig. 2a). The pair will then annihilate within the time scale

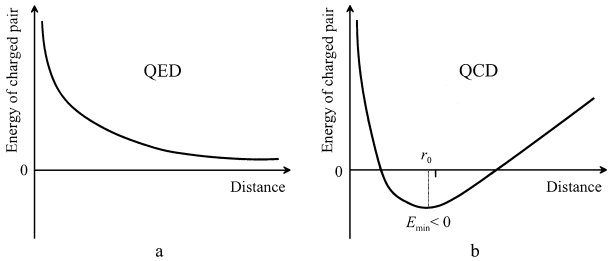


FIGURE 2 Qualitative dependence of the energy of a charge singlet pair, popped up from the vacuum, on the distance between the charges, in the case of QED (a) and in the case of QCD (b). (From CERN. Safarik, K. (2000). Heavy-ion physics. *In* "Proceedings of 1999 European School of High Energy," p. 267.)

$1/E_{\rm pair}$, which is again given by the uncertainty relation. The QED vacuum is filled with *virtual* charge pairs.

In QCD we get a qualitatively different behavior. The square of color charge $q^2 = g_{\rm s}^2 = 4\pi\alpha_{\rm s}$ at shorter distances decreases, i.e., $\alpha_{\rm s} \to 0$, which is known as asymptotic freedom. (Note, that there is a different numerical factor in this relation for the two singlet configurations: octet–antioctet—gg pair, and triplet–antitriples configuration—$\rm q\bar{q}$ pair; however, this does not change the qualitative conclusion of our discussion.) This is a consequence of the different structure of charges in QCD compared to QED. In fact the color charge in QCD is anti-screened (for the commonly assumed number of colors and flavors). The change of $\alpha_{\rm s}$ is the opposite to that of $\alpha_{\rm em}$, and is much faster. At the Planck scale it is expected to be $\alpha_{\rm s} \simeq 0.04$, at the eloctroweak scale the value $\alpha_{\rm s} = 0.118$ was measured, and eventually it rises to $\alpha_{\rm s} \simeq 1$ at the so-called $\Lambda_{\rm QCD} \simeq 0.2$ GeV scale, i.e., at distance $r \simeq 1$ fm. Therefore, the numerical factor $1 - q^2/(4\pi) = 1 - \alpha_{\rm s}$ in Eq. (1) decreases with distance, and at $r \simeq 1$ fm becomes negative. At even larger r the energy of a singlet pair in QCD is no longer given by Eq. (1), but is rather proportional to the distance r. This is because the field between separated color charges does not spread all over the space, like in QED, but is restricted to a string between them. The proportionality factor is the so-called string constant $\sigma \simeq 1$ GeV/fm (the value again depends on the color configuration of the singlet). The essential fact is that at large distances the energy of the pair rises linearly with the distance, $E_{\rm pair} = \sigma r$, and becomes positive again. As schematically shown in Fig. 2b, in QCD the energy of the pair first decreases, becomes negative, and then increases, as we separate the color charges. Therefore, the energy of the pair has a minimum at some distance $r_0 \sim 1$ fm, and, moreover, the value of this minimum is negative. As a consequence, an "empty" ($E = 0$) vacuum becomes unstable because there exists a configuration with lower energy. The pairs of color charges popped-up from the vacuum should stay there forever and become *real* pairs. In the QCD vacuum, one expects to have gg and $\rm q\bar{q}$ pairs with a typical separation $r_0 \sim 1$ fm, the gg pairs having larger probability, as the octet charge is numerically greater than the triplet one. These pairs will be in a singlet color and spin state.

In other words, when we try to create from a vacuum by quantum fluctuation a pair of charged particles, in QED the kinetic energy of the electron–positron pair always dominates over the energy stored in the electromagnetic field, because the field is relatively "weak." In QCD, on the other hand, the field is "strong," and the energy stored in the field overcomes at some distance the kinetic energy of the pair. The total energy of the pair of color charges then becomes negative. Therefore, the QCD vacuum is spontaneously filled by gg, and to a lesser extent, by $\rm q\bar{q}$ *real* pairs. This "vacuum condensate" behaves as a liquid, and a hadron can be imagined as a bubble in this liquid. Such a picture is a motivation for the bag model of hadrons.

The interaction between quarks and gluons is described by the QCD Lagrangian. The QCD Lagrangian has two approximate symmetries, which become exact in the two limiting cases for quark masses $m_{\rm q}$ that enter the Lagrangian (so-called "bare" masses):

- for $m_{\rm q} \to \infty$ we obtain a pure gauge SU(3) theory without dynamical quarks, which has Z_3 (center of SU(3) group) symmetry;
- for $m_{\rm q} \to 0$ we get QCD with massless dynamical quarks, which reveals chiral symmetry.

We shall give some arguments why these symmetries (or more precisely the way they are broken) are reflected in the transition between phases of QCD matter.

The center group Z_3 consists of elements, called gauge transformations, that commute with the QCD gauge group SU(3). Therefore, the Z_3 center transformations do not change the gauge (gluon) fields. Moreover, if we insert a static test colored quark in a purely gluonic world, at zero temperature, the detector will not feel the color charge because of destructive interference. To see this one has to calculate the expectation value for the trace of the quark propagator (Polyakov line, which is a quark observable) resulting in a three-valued path integral. The three components have equal absolute values and different phases $\exp(\mathrm{i}2\pi j/3)$, $j = 1, 2, 3$. As a consequence, we obtain zero due to the interference. This is similar to the well-known gedanken experiment where an electron is passing simultaneously through two slits. The detector will always see the test quark coming through three different paths with completely destructive interference, and therefore this quark will remain undetectable. Pure gauge theory (i.e., $m_{\rm q} \to \infty$) at zero temperature has the exact center Z_3 symmetry. This result remains true even at nonzero temperature T, up to some critical value. The expectation value of the Polyakov line (the quark propagator has to be continued over complex time $+\mathrm{i}/T$) will remain zero at low temperature, until the gluonic vacuum has enough time to rearrange coherently and to screen completely the test color charge.

When we raise the temperature T further, the complex time becomes shorter than the correlation length, $1/T < 1/\Lambda_{\rm QCD}$, and the coherence needed for destructive interference will be violated by suppression of some of the paths. The expectation value for the Polyakov line will become nonzero, which means that our test quark becomes detectable and hence deconfined. To summarize: at low temperature, the system of a gluonic vacuum and the

test charge has enough time to rearrange itself and it stays coherent, the color charge of the test quark is not visible because of destructive interference. However, when the temperature increases, i.e., the color charges shake faster. Above some critical value T_c the vacuum does not have sufficient time to follow with rearrangement, the coherence is destroyed, and the test color charge becomes visible. Therefore, we expect a phase transition at $T_c \simeq \Lambda_{QCD}$ between a low-temperature confined phase and a high-temperature deconfined phase. The order parameter of this transition is the expectation value of the Polyakov line mentioned previously, which is zero below T_c and finite above. The reason for this phase transition is the dynamical breaking of Z_3 symmetry. This symmetry is exact at low temperatures and breaks down at high temperatures, while usually dynamical symmetry breaking proceeds in an opposite way. Moreover, usually the symmetry is broken due to a degeneration of the potential energy minima, while Z_3 symmetry is broken as a consequence of the kinetic energy increase.

In the other limit ($m_q \to 0$) the quarks have to move in any system with the velocity of light because they are massless. As they are fermions with spin $1/2$ (internal angular momentum) they can have two possible spin projections, $-1/2$ and $+1/2$. At the velocity of the light, the helicity, i.e., the projection of the spin on the direction of flight, becomes a conserved quantity. This is a consequence of the fact that an observer cannot move faster than a massless quark and therefore he cannot see the quark's spin from the other side. The helicity of a quark does not flip if we change the reference system; we say it is Lorentz invariant. We call the quarks which have the helicity $-1/2$ left-handed and those with the helicity $+1/2$ right-handed. The gluons, which mediate the strong interactions between quarks and antiquarks, have spin 1 and they are massless too. Therefore they have only two helicity states, -1 and $+1$. This is similar to the case of a real photon which can have only two transverse polarizations, no longitudinal one, because of its zero mass. Due to helicity (angular momentum) conservation, a gluon with helicity -1 can decay only to left-handed quark left-handed antiquark pair and the one with helicity $+1$ only to right-handed quark right-handed antiquark pair. What happens in fact is that left-handed quarks interact only with left-handed antiquarks and right-handed quarks interact only with right-handed antiquarks. The QCD massless quark world decayed into two symmetric worlds, the left-handed one and the right-handed one, which do not communicate. This is called chiral symmetry. The QCD Lagrangian in the limit $m_q \to 0$ for light quarks (u, d, and s) reveal SU(3) flavor symmetry independently for left-handed and right-handed quarks, i.e., it has chiral symmetry $SU(3)_L \times SU(3)_R$.

As we discussed earlier, in the vacuum there exist $q\bar{q}$ pairs and they have to be in the singlet state in color and also to have zero net angular momentum. Already this means that the vacuum is broken. If we put inside such a vacuum a test massless quark, for example, with a left-handed helicity, it can annihilate on a left-handed antiquark thus liberating a right-handed quark. For an observer at some distance this will look like the test quark, being in a vacuum, changes its helicity spontaneously. Therefore, it cannot move with the speed of light, and hence it had to acquire some dynamical mass M_q. Chiral symmetry is dynamically broken due to the $q\bar{q}$ vacuum condensate.

As we raise the temperature, we increase the kinetic energy. At some critical value T_c of the order of the lightest meson mass m_π, we overcome the energy stored in the strong field. At this point the minimum of the total pair energy will become positive, and hence the *real* $q\bar{q}$ pairs would disappear from the vacuum. Above that temperature, chiral symmetry will be restored, and quarks will retain their zero mass in the chiral limit. The order parameter of this phase transition is the value of the vacuum quark condensate, $\langle 0|\bar{q}q|0\rangle$, i.e., a measure of the density of $q\bar{q}$ pairs in a vacuum. It has a nonzero value at zero temperature and drops to zero at critical temperature $T_c \simeq m_\pi$. In this case, chiral symmetry is broken at zero temperature (due to the potential energy), and restored at high temperature.

In addition to dynamical symmetry breaking, both Z_3 and chiral symmetries are also broken explicitly, by the finite mass term $-m_q\bar{q}q$ in the QCD Lagrangian. The bare masses m_q are only a few MeV for u and d quarks, and about 150 MeV for s quark; that means negligible compared to, or at most comparable with, the scale expected for T_c. Therefore, it seems reasonable that the chiral symmetry transition scenario will remain qualitatively valid: there is dynamical breaking of chiral symmetry at low temperature and its approximate restoration above T_c. The question is why the Z_3 symmetry at low temperature is not completely destroyed by such small values of m_q, which are far from infinity. We can argue that when we try to drop the quark mass from infinity down to its bare value at low temperature, the mass effectively stops decreasing at its dynamical value $M_q \simeq 350$ MeV (one third of the baryon mass, or one half of the ρ-meson mass), which is still well above any expectation for T_c. Therefore, the Z_3 symmetry remains an approximate symmetry at low temperature, even after this attempt of a severe explicit breaking. This is also an argument which suggests that the two phase transitions, confinement–deconfinement and chiral symmetry, occur at the same point. In other words, the chiral symmetry breaking, by effectively increasing the quark masses, drives the Z_3 symmetry restoration.

B. Phase Diagram of Matter

In order to have an estimate of the transition temperature at zero baryon density we compare the two simplest approximations: an ideal noninteracting Hadron Gas (HG) of massless pions, and an ideal gas (QGP) of massless gluons and of two-flavor quarks. These two situations differ (i) in the number of degrees of freedom, i.e., the degeneracy factor g; and (ii) by the presence of an external pressure from the QCD vacuum condensate in the QGP case. This external pressure is absent in HG, as the vacuum liquid is also present between the pions and the pressure acts on each pion separately. The energy density ϵ, the pressure p, and the entropy density s of an ideal gas are expressed by the following thermodynamic relations:

$$\begin{aligned} \epsilon &= \frac{g}{30}\pi^2 T^4, \\ p &= \frac{\epsilon}{3} = \frac{g}{90}\pi^2 T^4, \\ s &= \frac{\partial p}{\partial T} = \frac{2g}{45}\pi^2 T^3, \end{aligned} \tag{2}$$

where g is a degeneracy factor and T is the temperature. The degeneracy factor is different for boson degrees of freedom n_b and fermion degrees of freedom n_f, due to different quantum statistics, and is given by

$$g = n_\mathrm{b} + \left(1 - \frac{1}{2^3}\right) n_\mathrm{f}. \tag{3}$$

In the case of HG we have three boson degrees of freedom $n_\mathrm{b} = 3$, i.e., three isospin pion states, and no fermions. Therefore, the pressure as a function of temperature will be

$$p_\mathrm{HG} = \frac{1}{30}\pi^2 T^4. \tag{4}$$

In QGP, we have both the boson (gluon) degrees of freedom $n_\mathrm{b} = 16$ (eight color states times two spin states for gluons), and the fermion (quark) degrees of freedom $n_\mathrm{f} = 24$ (2 flavors $\times$ 3 colors $\times$ 2 spins $\times$ 2 for quark–antiquark). We estimate the external pressure using the bag model for hadrons where it is equal to the bag constant B. As a result, we have for the pressure in the QGP phase

$$p_\mathrm{QGP} = \frac{37}{90}\pi^2 T^4 - B. \tag{5}$$

Comparing Eqs. (4) and (5) we see that when the temperature T rises from zero we get at low temperatures larger pressure in the HG phase; then the pressures in the two phases become equal; and finally the pressure in the QGP phase will be above that of the HG. This is a consequence of the larger number of degrees of freedom in the QGP phase. According to the Gibbs criterion, the phase with the largest pressure is the stable one, so at low temperatures it will be the HG phase and at high temperatures the QGP phase. At the phase transition temperature T_c the two pressures are equal, and from Eqs. (4) and (5) we get

$$T_\mathrm{c} = \left(\frac{90B}{34\pi^2}\right)^{1/4}. \tag{6}$$

Assuming $B^{1/4} = 200$ MeV, i.e., the value used in the bag model to describe the hadron mass spectra, we finally get the critical temperature $T_\mathrm{c} = 144$ MeV.

Exact theoretical calculations are done in the lattice QCD. They have been performed, on different lattice sizes, both for pure gauge theory and for dynamical quarks at zero net baryon charge. These calculations confirm the existence of the two phases. However, the phase transition can be found to be of the first or of the second order, or even a smooth crossover, depending on the number of the quark flavors and their masses. The critical temperature T_c obtained at zero baryon density is about 260 MeV for pure gauge, and varies between 140 and 170 MeV for theories with dynamical quarks.

The QGP phase at nonzero baryon density can be studied at high temperatures in the framework of a thermal perturbative QCD. Assuming that interactions among quarks and gluons are small, the energy density as a function of temperature T was calculated in first-order perturbation theory for vanishing quark masses

$$\begin{aligned} \epsilon = &\left[16\left(1 - \frac{15}{4\pi}\alpha_\mathrm{s}\right) + \left(1 - \frac{1}{2^3}\right) 12 n_\mathrm{q}\left(1 - \frac{50}{21\pi}\alpha_\mathrm{s}\right)\right] \\ &\times \frac{1}{30}\pi^2 T^4 + \sum_\mathrm{q}\left(1 - \frac{15}{2\pi}\alpha_\mathrm{s}\right)\frac{3}{\pi^2}\mu_\mathrm{q}^2\left(\pi^2 T^2 + \frac{1}{2}\mu_\mathrm{q}^2\right), \end{aligned} \tag{7}$$

where n_q is the number of active quark flavors and μ_q is the quark chemical potential. Comparing with Eqs. (2) and (3) we see that Eq. (7) is an equation for an ideal gas with some QCD corrections. If we put $\alpha_\mathrm{s} = 0$ (no interaction), $\mu_\mathrm{q} = 0$ (zero baryon density), and $n_\mathrm{q} = 2$ (two flavors in system), in fact, we obtain exactly the equation of our toy QGP model described earlier. As in that simple estimate, we can calculate the critical temperature T_c, now as a function of the chemical potentials μ_q. This way we can estimate the boundary between the two phases in the baryon density–temperature plane, shown schematically in Fig. 3. Note, however, that Eq. (7) was obtained under the assumption of high temperature, i.e., small α_s, and it might be far from reality at low temperatures and high baryon densities. In a baryon-free regime, i.e., $\mu_\mathrm{q} = 0$, assuming $\alpha_\mathrm{s} = 0.4$, we get $T_\mathrm{c} = 0.82 B^{1/4} = 164$ MeV, for two quark flavors. This

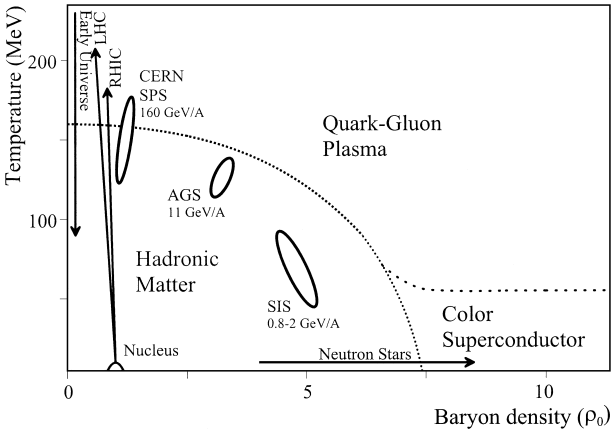


FIGURE 3 Phase diagram of matter in baryon density–temperature plane. Dotted lines indicate predicted phase boundaries. The values achieved in heavy-ion collisions at different accelerators are shown. Arrows indicate the transitions which presumably have occurred in nature or will be attempted at new colliders. (From CERN. Heavy-ion physics. *In* "Proceedings of 1999 European School of High Energy," p. 267.)

estimate will change to $T_c = 152$ MeV, if we include the strange quark.

Model calculations for very high baryon densities and low temperatures generally confirm the first-order phase transition for baryon densities 4–10 times higher than that of normal nuclear matter. Above these densities, and at temperatures lower then few tens of MeV, another phase, a superconducting quark condensate, can be expected. This phase is formed of quark Cooper pairs, as a consequence of an attractive quark–quark potential in an antitriplet color configuration. This is a topic which recently received a lot of attention; however, such densities are probably unreachable in accelerator experiments with heavy ions.

Figure 3 summarizes schematically the phase diagram of matter. The regions accessible to different accelerators are also shown. In addition, the phase transitions that presumably occurred in nature are indicated. At high baryon densities the phase transitions are expected to be of the first order. At low baryon densities and high temperatures the character of the phase transition probably changes to second order or to a smooth crossover. Therefore, somewhere at the phase boundary line a tricritical point may exist.

II. QGP SIGNATURES

Heavy-ion collisions proceed in a few sequential stages. When the two nuclei impact, the quarks and gluons (partons) start to interact between themselves destroying the color coherence of the pre-interaction states. At the beginning, partons undergo hard scatters (with large momentum exchange); however, the energy in such collisions descends rapidly to the thermal regime. The time needed to thermalize the early pre-equilibrium stage depends on the density and the parton–parton scattering cross section. Both are estimated to be relatively high, indicating that the initial thermalization will take place after a time of 1 fm or less. Then, if the initial energy density is sufficient, an equilibrated cluster of QGP is formed which will expand and cool down during the next few fm. When the temperature decreases to the critical temperature, hadron formation (hadronization) starts, leaving the system for some time in a mixed phase. The time spent in this stage depends on the nature of phase transition and details of the model. It is believed that the relative abundances of different hadron species are frozen at the end of this stage or shortly afterwards, when the temperature of the system is so low, that inelastic hadron–hadron cross sections are negligible. We call this time in the evolution of the system the chemical freeze-out. The produced hadrons (hadron gas) can still interact elastically among themselves, transferring the thermal energy into the radial flow until the system finally flies apart. The stage when hadrons cease to interact is called the thermal freeze-out. This evolution is schematically shown in Fig. 4 as the longitudinal coordinate–time diagram. There is an alternative scenario, in which such high pressure is built up quickly in the QGP phase that the QGP cluster explodes and a sudden hadronization follows. In this model, chemical freeze-out and thermal freeze-out occur at the same time.

Many different signatures have been proposed in order to confirm the formation of a QGP phase in heavy-ion collisions. We can divide them into two groups: (i) hard probes and (ii) soft probes, according to the stage at

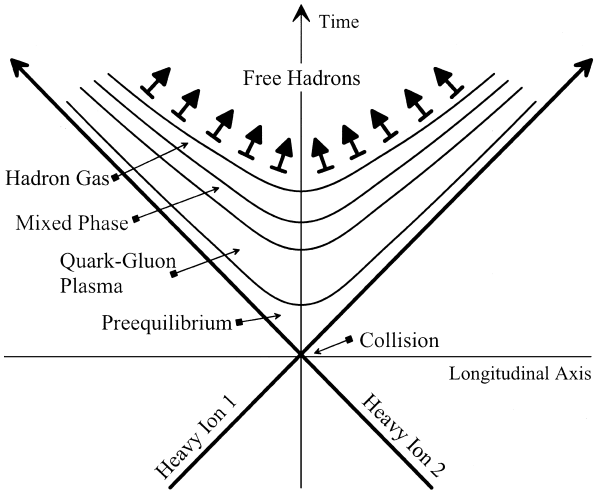


FIGURE 4 Space–time evolution of heavy-ion collision. Different phases are schematically indicated.

which they are produced. The hard probes are sensitive to the early stage of a collision, but they have generally a lower cross section and, in some cases, the produced particles can still be affected by later stages of a collision. The soft probes, on the other hand, come from the later stage of a collision and they do not directly witness the QGP phase, being the result of the hadronization process. Nevertheless, specific probes can retain information on the previous stages.

We can also divide QGP signatures according to another criterion, the type of final state particles, into: (i) hadronic probes and (ii) electromagnetic probes. Hadronic probes have large cross sections and are relatively easy to measure. However, they also have disadvantages: the hadrons undergo a substantial evolution through strong re-interactions in the period between their formation and their detection. The QGP has to first hadronize into resonances and particles, which then for some time will interact among themselves, both elastically and inelastically, until their spatial density decreases sufficiently and the final hadrons freeze-out. Therefore, both the momentum distributions and the final particle composition can be affected by later stages of the heavy-ion collision. Despite this we are able to access the properties of the first, very dense, stage of the collision observing specific final hadrons.

On the other hand, the electromagnetic probes are of course more direct tools for investigating the first stages of heavy-ion interactions because they have a negligible cross section for interacting with hadronic matter and hence, after being produced, they have little chance to be affected later. Here the problem is that generally their production cross sections are also very low, so their detection in the high-background environment of heavy-ion collisions is a very difficult task.

We shall concentrate on some of the QGP signatures that have been exploited in recent experiments with heavy ions.

A. Charmonium Suppression

Charmonium production is a hard hadronic probe since charm quarks in heavy-ion interactions are produced in hard parton collisions. In addition, they may be produced in the very early pre-equilibrium stage discussed previously, when the temperature is still high enough to overcome the charm production threshold. There is a large difference between charmonium production in ordinary hadron collisions and in deconfined matter. This effect is analogous to Debye screening in classical electrostatic theory. The potential between two charges in a dense medium of many other charges reduces its range to the screening radius r_s, due to the screening effect

$$E_{\text{pot}} = \sigma r \cdot \frac{1 - \exp(-r/r_s)}{r/r_s}. \tag{8}$$

The screening radius r_s is estimated from lattice QCD results to be 0.3–0.5 fm. In a QGP the $c\bar{c}$ pair produced in a hard parton collision cannot form a $c\bar{c}$ bound state if the size of this state is larger than the screening radius. In other words: if the density of the medium is large enough, before the c and $\bar{c}$ quarks reach the distance at which they would resonate, other quarks from the medium already appear inbetween them, and as a consequence the charm quarks fragment into D mesons rather than form a $c\bar{c}$ bound state.

This is a completely different situation than that for strange quarks, where we expect an enhancement of the ϕ ($s\bar{s}$ state) production. The $s\bar{s}$ pairs will be also screened; however, in this case, owing to the large strangeness density in a QGP (see further), the quarks which appear in between s and $\bar{s}$ quarks trying to prevent the ϕ formation can also be strange, with a reasonable probability. Contrary to charm quarks, strange quarks are easy to produce, e.g., in string fragmentation. Therefore, this screening mechanism would not alter the ϕ production in any significant way, as one might at first think.

It was predicted that the charmonium states production will be suppressed with a characteristic pattern. As the temperature of deconfined matter rises, i.e., the density of color charges increases. First the states with larger radii begin to disappear, while the smaller ones would be affected only at higher temperatures. Estimations show that ψ' should disappear right at the phase transition, i.e., at T_c, then the χ_c states will follow at about $1.1T_c$, and finally at about $1.3T_c$ the J/ψ itself will complete the disappearance of charmonium states. A fair fraction of J/ψ are normally produced in the decays of ψ' (5–8%) and even larger in those of χ_c states (32–40%). This is easy to understand: χ_c states can be produced via a two-gluon annihilation, unlike the J/ψ, which needs at least one additional gluon in order to get the correct parity. As a consequence a small suppression in J/ψ production has to happen when the conditions for a QGP are reached, then at higher-energy densities we have to observe a drop of about 40%, and, eventually, at even higher-energy densities J/ψ has to disappear completely.

On the other hand, the J/ψ, being a hadron, will be affected in subsequent hadron interactions. Even in a purely hadronic scenario, we expect some level of J/ψ suppression because when the J/ψ interacts with pions and nucleons, there is a large probability that it will disappear. The J/ψ cross section with ordinary hadrons is relatively small, about 2–3 mbarn. However, J/ψ absorption could be larger due to interactions with co-moving hadrons owing to the large density of these comovers. Another increase of the J/ψ absorption could be produced if, before

the J/ψ is formed, the $c\bar{c}$ pair is for a sufficiently long time in an octet color state, since such a state has a larger hadronic cross section due to its larger color charge (by a factor $9/4$, compared to a triplet charge). This effect would increase the effective J/ψ absorption cross section up to about 7 mbarn, and explains the J/ψ production measured in hadronic and h–A reactions.

In heavy-ion interactions we expect an additional suppression of the J/ψ production, if the hadronic matter is deconfined during the collision. The observation of such an "anomalous" suppression would be a strong indication for deconfinement.

A similar effect is predicted for heavier $b\bar{b}$ quarkonia, the Υ family. There, a clear hierarchy of suppression is predicted: complete disappearance of the heaviest (and largest) one, Υ'', and up to energy densities of about 10 Gev/fm^3 no effect at all for the lightest one, Υ, which is smaller then J/ψ.

B. Jet Quenching

Another hard probe proposed to search for QGP is jet quenching. High transverse momentum particles (a few GeV) are produced via hard parton–parton scattering, where first a high-p_T parton is produced which then fragments into a jet of hadrons. If the parton is produced in a heavy-ion collision it has to traverse a dense cluster of matter before it fragments and its cross section depends on the state in which that matter is found. The cross section of a parton with colored objects in QGP is higher than that with a neutral hadron gas. Therefore, inside a QGP a parton will lose part of its momentum which can be observed as a softening of transverse momentum particle spectra. This effect will be more pronounced at higher energies because high-p_T particle production will be dominated by scattering and fragmentation of gluons which have larger color charge, and therefore are more effectively slowed down by QGP. The estimates show that a gluon can lose from a few 100 MeV up to 1 GeV of momentum traversing 1 fm of QGP (depending on different assumptions about QGP).

Another possibility to look for jet quenching is to try to separate gluon and quark jets using the differences in their fragmentation. We have to observe significantly larger quenching for gluon jets than for quark ones.

C. Kinematic Probes

A large class of soft hadronic probe deals with the determination of the thermodynamical variables, like temperature T, energy density ϵ, entropy density s, and pressure p, of a dense hadronic matter produced in heavy-ion collisions. From Eq. (2) for an ideal gas we see that we can use these variables for the measurement of the number of degrees of freedom, or more precisely the degeneracy factor g, which increases rapidly when going from a hadron gas to a QGP. Observing a rapid rise of ϵ/T^4 or s/T^3 within a small temperature interval will signal a phase transition. The thermodynamical variables are connected to the measured observables: temperature to the inverse slope of m_T distribution or mean p_T, energy density to the transverse energy density dE_T/dy and entropy density to the particle density dN/dy.

A rapid increase of pressure p in the QGP phase has as a consequence of fast collective expansion of the system, and hence development of a large outward flow. This could be observed via modifications of particle spectra, especially those for heavier particles. If the heavy ions collide noncentrally, i.e., their centers in the transverse plane are displaced and their overlap region is not centrally symmetric. The spatial distribution of matter will influence the amount of flow in different directions. As a result we will observe elliptic flow. The size of the effect is sensitive to pressure build-up in the system, and hence to the stiffness of the equation of state.

Information about the geometry of the collision and about collective expansion can be obtained using interferometry measurements by studying identical particle correlations. Investigation of correlations in different directions with respect to particle momenta gives us information about the transverse and longitudinal size of the particle source, and its lifetime. This way we can reconstruct the space–time dynamics of heavy-ion reactions.

D. Strangeness Enhancement

Strangeness enhancement was among the first signature proposed for the observation of a QGP. It is a typical soft hadronic probe. There are two reasons why strangeness should be enhanced in a QGP; the first is due to a large temperature expected to be achieved in the heavy-ion collision at high energies, and the second is an additional enhancement at large baryon densities.

At low temperatures strange quark production is suppressed due to the large dynamical mass of the strange quark, $M_s \simeq 500$ MeV. The suppression factor with respect to u and d quark production (with dynamical masses $M_{u,d} = M_q \simeq 350$ MeV) can be approximated by the ratio of Boltzmann factors ($\simeq \exp -(E/T)$) with an additional factor $1/2$ taking into account the dilution of strangeness due to resonance decays, where mostly new u and d quarks are produced:

$$\gamma_s \simeq \frac{1}{2} \cdot \frac{\exp\left[-\left(M_s^2 + T^2\right)^{1/2}/T\right]}{\exp\left[-\left(M_q^2 + T^2\right)^{1/2}/T\right]}, \qquad (9)$$

which gives at typical hadronic temperature $T = 150$ MeV the value $\gamma_s \simeq 0.2$. After a chiral symmetry restoration

the quark masses suddenly drop, and we have to substitute the dynamical masses with bare masses: $M_s \to m_s \simeq 150$ MeV and $M_q \to m_q \simeq 0$ MeV. For the same temperature $T = 150$ MeV, γ_s will increase to about 0.4. Therefore, as a consequence of the chiral symmetry restoration, we would observe a global strangeness enhancement (or rather a reduction of suppression) by a factor of about 2.

A second reason for a strangeness enhancement arises when there is large baryon density. For example, in Pb–Pb collisions at CERN energies, the baryon chemical potential in the central region reaches $\mu_B \simeq 250$ MeV. Therefore, if the hadronic matter is deconfined during the collision, the volume of the central fireball is already occupied by many u and d quarks coming from the interacting nuclei. As a consequence the production of u and d quarks is suppressed due to the Pauli exclusion principle. In this way the production of the s quarks will be *relatively* enhanced. Taking into account the nonzero chemical potential of u and d quarks, $\mu_q = \mu_B/3$, the strangeness suppression factor γ_s given by Eq. (9) has to be multiplied by $\sim \exp(\mu_q/T)$, and it will further increase up to $\gamma_s \simeq 0.6$. Consequently, at CERN energies we expect a global increase of strangeness production by a factor of about $\eta_s \simeq 2.5$–3, using the very rough estimates presented here.

Under the same assumptions, the strangeness enhancement will be more pronounced for particles with more than one strange quark, i.e., the ϕ meson ($s\bar{s}$) and the cascade baryons, Ξ^- (ssd) and Ω^- (sss). In a first approximation, if they are produced by a recombination of quarks from a QGP, we would expect that ϕ and Ξ^- production will be enhanced by a factor of about $\eta_s^2 \simeq 6$–9 and the Ω^- production by a factor of about $\eta_s^3 \simeq 15$–25. These estimates, however, do not take into account many important details such as the hadronization process and different hadron wave functions.

Strangeness production can also be enhanced in a pure hadronic scenario, without QGP formation. If, during the heavy-ion collision, the gas of produced hadrons has enough time to interact, the inelastic collisions will drive the system toward chemical equilibrium. In this scenario, at the beginning strangeness production is suppressed and during the hadronic re-interactions the strangeness content will increase with time. A typical inelastic process of this type is $\pi^0 + p \to K^+ + \Lambda$. Once we produce a strange particle, the probability to destroy it is very low, because it interacts mostly with pions and nucleons (until strangeness will reach its equilibrium value). For multi-strange hadrons such re-interactions will be much less effective. In order to produce, for example, an Ω^- we have to go through the following reaction chain: (i) $\pi^0 + p \to K^+ + \Lambda$, (ii) $\pi^0 + \Lambda \to K^+ + \Xi^-$, (iii) $\pi^+ + \Xi^- \to K^+ + \Omega^-$. Such reaction chains have low probability, and therefore need a long time. In addition, multi-strange particles can easily be destroyed in subsequent interactions with pions or nucleons. Therefore, the approach to chemical equilibrium for the Ω^- will be very slow. For the $\bar{\Omega}^+$ we would have to go through another small cross-section process, the production of an antibaryon, which will further increase the equilibration time. The direct production of an $\Omega^-\,\bar{\Omega}^+$ pair is strongly suppressed because of the high threshold (above 3 GeV for $\pi^+\pi^-$ annihilation).

Estimates show that the chemical equilibration time in a hadronic gas for $\bar{\Omega}^+$ will be of the order of 100 fm, while the typical timescale for a Pb–Pb collision is given by the size of the Pb nucleus and is only of the order of 10 fm. On the other hand, in a QGP strangeness equilibration will proceed very fast, because after a chiral symmetry restoration strange quark production is at the same level as the other light species. Once a sufficient strangeness density has built up during the QGP phase, it is easy to fill the phase space during hadronization according to the maximum entropy principle, i.e., according to chemical equilibrium. If the system later spends a significant time in the interacting hadronic phase, the temperature will decrease, which, if anything, will lower the chemical equilibrium yields of heavy particles like Ω^- and $\bar{\Omega}^+$. Therefore, the observation of an order-of-magnitude enhancement of multi-strange baryon production would be a strong argument in favor of the creation of a QGP during heavy-ion collisions.

E. Electromagnetic Probes

There are predictions for the production of both direct thermal photons and thermal dileptons (e^+e^- or $\mu^+\mu^-$ pairs) in a QGP. At present energies the signals are, however, very small compared to expected background.

Near to the critical temperature the spectrum of direct photons from QGP, mainly from the reaction $gq \to \gamma q$, is very similar to that from hadron gas, produced in the reaction $\pi\rho \to \gamma\rho$, both in shape and intensity. A clear direct photon signal from a QGP is expected only at significantly higher temperatures and transverse momenta $p_T > 2$ GeV.

The thermal lepton pair yield from a QGP competes with other dilepton sources. At lower effective masses it will be dominated by dilepton decays of vector mesons, ρ and ω. The vector mesons themselves, especially those with very short lifetime, such as the ρ, can be modified in hadronic dense matter. At higher masses the dilepton yield critically depends on the QGP thermalization time, and this in turn determines up to which effective mass this yield would be higher than that of initial Drell–Yan production, i.e., quark–antiquark annihilation

into virtual photons. Depending on the scenario, the effective mass limit reaches 5–10 GeV. It is worth noting that in this mass region a substantial contribution to the total lepton pair spectra might be due to semi-leptonic decays of charm, and for higher masses and higher energies due to semi-leptonic decays of beauty particles.

The electromagnetic probes are hard to observe in fixed-target experiments because of very small yields and large backgrounds at these energies. The situation will change at higher energies, i.e., in collider heavy-ion experiments, where higher temperatures are expected.

III. EXPERIMENTAL SITUATION

Heavy ions were for the first time accelerated to relativistic energies of about 1 GeV per nucleon at Lawrence Berkeley Laboratory (LBL), where the existing linear ion accelerator and Betatron were joined by the transfer line in the early 1970s. At this new complex called the Bevalac, pioneering experiments were done which have proven compressibility of nuclear matter. Bevalac ceased to operate after decommissioning at the beginning of the 1990s. At present in this energy domain, the experimental research continues at SIS facility (heavy ion synchrotron) in Gesellschaft für SchwerIonenforschung (GSI) at Darmstadt, Germany. Accelerator studies with heavy-ion beams were also undertaken at the JINR Dubna complex in Russia, where maximum energies of 4 to 5 GeV per nucleon were achieved.

Experiments with heavy-ion beams of ultra-relativistic energies have started in 1986 at Brookhaven National Laboratory (BNL) Alternating Gradient Synchrotron (AGS) and at CERN Super Proton Synchrotron (SPS). At BNL AGS the maximum energy per nucleon varies between 11 and 14.6 GeV depending on the ion (i.e., proton to neutron ratio) and ions up to ^{197}Au were accelerated. At the CERN SPS the maximum energy achieved is between 200 GeV per nucleon for ^{16}O and ^{32}S ions and 158 GeV per nucleon for ^{208}Pb ions. We shall concentrate on the results obtained with the heaviest ion beams at maximum energies achieved.

The two nuclei approaching each other have some transverse distance between their centers which is called the impact parameter. In order to achieve maximum energy density in the collision, it is better to select collisions with small impact parameters. These central collisions are selected in experiments by (i) lack of (or small) signal in a calorimeter looking for beam remnants in the forward region, (ii) large signal in a calorimeter measuring transverse energy in central region, or (iii) large multiplicity of secondary particles produced in the collision.

A. J/ψ Suppression

The $\mu^+\mu^-$ production in heavy-ion collisions have been systematically investigated at SPS using a dimuon spectrometer with a toroidal magnet for muon measurements. A clear J/ψ signal decaying into $\mu^+\mu^-$ in different collision systems has been observed.

The use of different analysis methods give compatible results. The J/ψ yield normalized to another hard process, the Drell–Yan $\mu^+\mu^-$ production, is presented. As a measure of the centrality of collisions the observed transverse energy E_T in a calorimeter is used. It can be related to energy density ϵ, assuming an initial thermalization time of about 1 fm.

In Fig. 5 we see the measured J/ψ yield normalized to the Drell–Yan divided by the yield expected in the case of the "normal" J/ψ absorption in nuclear matter. The results for different methods are presented. In addition to the data for Pb–Pb collisions with different centralities, the results for S–U and p–A reactions, as well as the data for pp and pd interactions, are plotted. The normalized J/ψ yields follow the expected absorption up to an energy density of about 2.2 GeV/fm^3 where a first sharp drop is observed. The data indicate also a second drop at the energy density of about 3.2 GeV/fm^3. Such a pattern of J/ψ suppression was predicted for J/ψ production in deconfined

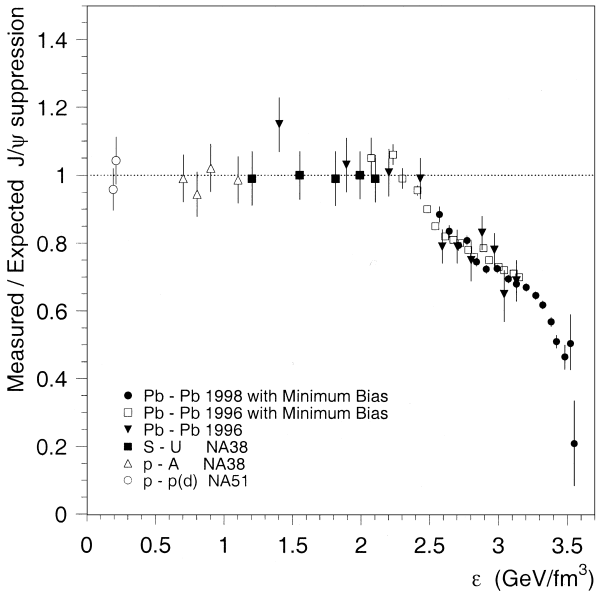


FIGURE 5 J/ψ yield normalized to that expected, after extrapolation of the measured absorption in normal nuclear matter, as a function of the energy density (NA50 collaboration). (From Abreu, M. C., *et al.* (2000). "Evidence for Deconfinement of Quarks and Gluons from the J/Psi Suppression Pattern Measured in Pb–Pb Collisions at the CERN-SPS," p. 28, Copyright 2000, with permission from Elsevier Science.)

matter. In addition, the previous results show that the ψ' is already suppressed in S–U central collisions.

B. Transverse Mass Spectra

Almost all experiments studying heavy-ion collisions measured transverse spectra for different particle species. They have been approximated by an exponential function $\propto \exp(-m_T/T)$ of transverse mass $m_T = \sqrt{m^2 + p_T^2}$ with an inverse slope T as a parameter. A very interesting pattern is obtained, when we look at how these inverse slopes depend on the particle mass m. The observed dependance is interpreted as the result of collective transverse flow, which comes about as a consequence of particle interactions before they thermally freeze out. It means that we observe the particle spectra "blue-shifted," i.e., with larger apparent temperature T. Their m_T distributions have a temperature component and a collective velocity component. When particles have the same mean velocity, more massive particles will have a larger momentum, i.e., larger collective flow component in the transverse momenta. Therefore, we expect to see an increase of the inverse slope parameter T with particle mass, if there is a substantial transverse expansion. The data collected from different experiments are compatible with a linear increase of the inverse slope with particle mass.

The freeze-out temperature T_0 and the average transverse flow velocity v_T could, in principle, be deduced from this dependence. However, these two parameters have a large correlation, when determined using only this information. Other experiments have measured the transverse radius parameter of the source, using two-pion correlations, as a function of the transverse momentum of the pion pair. The transverse radius decreases with p_T, which is again interpreted as a collective expansion because high p_T particles come, in general, from an earlier stage of the collisions. Measurement of the source size is more sensitive to the velocity than to the temperature. In a combined analysis of the inverse slope data and the p_T dependence of the two-pion correlation, the following values are obtained for Pb–Pb collisions at SPS: $T_0 = 110$–120 MeV and $v_T = (0.55$–$0.60)c$. This tells us that in central Pb–Pb collisions most of the hadrons move radially out with a collective transverse velocity of more than half the speed of light. A similar analysis for AGS energy gives the thermal freeze-out at $T_0 = 90$–100 MeV and $v_T \simeq 0.4c$.

The most marked deviation from the previously described behavior is the inverse slope for $\Omega^{\pm}$ which is well below the linear dependance. To understand this we recall that the transverse expansion is built up through interactions among the hadrons before their density drops to a point where they thermally freeze-out. These interactions are mostly elastic and, in the case of the Ω, only elastic collisions would contribute, as otherwise Ω particles would be destroyed. Usually, the elastic cross section near the threshold is governed by elastic resonances. But in the $\Omega\pi$ system there are no resonances (made up from three strange quarks) because of isospin conservation. Therefore Ωs, once they have been produced, interact little with the pion gas surrounding them, unlike most of the other particles. A quantitative estimate using microscopic model calculations agrees well with this behavior.

C. Multi-Strange Particle Production

As we have argued in the previous section, an enhanced multi-strange particle production is one of the prominent signatures of a QGP. The global strangeness enhancement in heavy-ion collisions was established both at AGS and SPS in many different experiments. In Pb–Pb collisions at SPS the results on kaon production indicated a global strangeness enhancement, since kaons carry about 75% of all strangeness produced in a collision. The measured enhancement is comparable to that observed in the Au-induced reaction at AGS and in S–S and S–Ag at SPS, i.e., a factor of about 2 when normalized to pion production.

One series of SPS experiments concentrated especially on measurements of yields of multi-strange particles in the central region. They have used a telescope made of silicon detectors for the reconstruction of strange particle decays. The number of wounded nucleons, i.e., the nucleons from colliding ions interacting between themselves, is used both as a centrality measure and for normalization. It is estimated from the measured charged particle multiplicity in the central region. The data were collected in p–Be, p–Pb and Pb–Pb collisions for negative particles (h^-), particles with 1 unit of strangeness: K_S^0, Λ and $\bar{\Lambda}$, with 2 units of strangeness: Ξ^- and $\bar{\Xi}^+$, and with 3 units of strangeness: $\Omega^{\pm}$. The yields for Ω^- and $\bar{\Omega}^+$ are presented as a sum owing to the small statistics, particularly in proton-induced collisions.

For each particle type the yields per wounded nucleon are compatible between p–Be and p–Pb collisions. In Pb–Pb collisions they are enhanced with respect to the p–A yields by different factors, depending on the particle type. An interesting enhancement pattern is revealed when plotting these enhancement factors as a function of particle strangeness content (see Fig. 6). The observed enhancement increases with the strangeness of the particle, and reaches a value of about 15 for $\Omega^{\pm}$.

Results on Ξ^- and $\bar{\Xi}^+$ production in central Pb–Pb interactions have been confirmed by another SPS

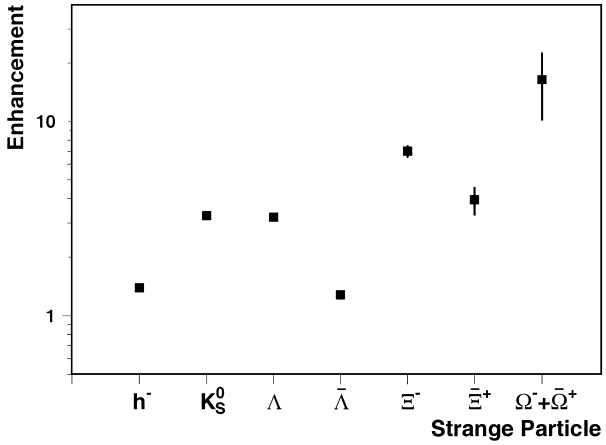


FIGURE 6 Enhancement for particles with different strangeness content produced in Pb–Pb collisions at 158 GeV per nucleon relative to p–A data (WA97 collaboration).

experiment using a large Time Projection Chamber (TPC). The global enhancement of Ξ^- with respect to pp interactions was estimated to be about an order of magnitude, in comparison to the interpolation of pp measurements at different energies.

The central yields per wounded nucleon are within experimental uncertainties constant as a function of the number of wounded nucleons in Pb–Pb collisions for N_{wound} above 100 (which is the region covered by the experiment). The yields per wounded nucleon observed in p–Be and p–Pb collisions are significantly lower than those in Pb–Pb collisions, suggesting that the enhancement sets in for N_{wound} somewhere below 100. A similar observation was made for the centrality dependence of the K^+ and K^- invariant cross section.

D. Electromagnetic Signals

Two experiments at the SPS were looking for electromagnetic signals, which at these energies is a very difficult task. They were trying to access direct photons. They succeeded in reconstructing π^0 and η mesons over a large p_T range, and found a small remaining photon signal, after subtraction of γ's from known sources.

In addition, one of these experiments has reported a low-mass dielectron (e^+e^-) enhancement in central Pb–Au collisions, which is most pronounced in the effective mass window $0.2 < m_{ee} < 0.7$ GeV and at low transverse momentum. In order to better understand this observation, the experimental set-up was modified by adding a more precise charge particle detector (a TPC) which significantly improved the dielectron mass resolution.

IV. INTERPRETATION

The results described in the previous section demonstrate two new experimental effects in central Pb–Pb collisions at SPS energies:

- Production of the J/ψ particle is anomalously suppressed. In addition to an absorption in nuclear matter, the data show two sudden drops, the first one at the energy density of about 2.2 GeV/fm^3 and the second at about 3.2 GeV/fm^3.
- Large enhancements are observed in the central region for the multi-strange baryon, reaching a factor of about 15 for Ω particles. The enhancement exhibits a pattern in which particle production is more, enhanced, the greater the strangeness content of the particle.

The observed behavior of J/ψ production in Pb–Pb interactions is naturally interpreted as a hierarchy of χ_c and J/ψ suppression, as predicted in QGP at temperatures well above T_c. Although, the second drop in J/ψ production, for very central collisions, can be due to the rapid change in transverse energy fluctuations when approaching zero impact parameter. In sulfur-induced interactions, neither χ_c nor J/ψ anomalous suppression is observed, though ψ' suppression is seen.

The observed inverse slopes of the m_T distributions generally follow a linear dependence of the inverse slope on the mass m of the particle species. The measured transverse radius parameter decreases with p_T. This behavior was predicted as a consequence of collective flow. The data indicate a mean transverse velocity $v_T = (0.55–0.60)c$, superimposed on the thermal motion with temperature $T = 110–120$ MeV at thermal freeze-out.

Thermal model analyses of strange particle yields have shown that the system approaches thermal and chemical equilibrium. The different features implemented in the models lead to a clustering of values around $T_c = 170–180$ MeV for chemical freeze-out temperature and μ_B about 250 MeV for baryon chemical potential in central Pb–Pb interactions. A fast strangeness equilibration was one of the fundamental predicted features in QGP formation, and this leads naturally to the thermal and chemical equilibration of strange and multi-strange particles at hadronization. The particle composition is essentially frozen at this point, because the inelastic hadronic cross sections at that temperature are too small to change particle abundances significantly within the rapidly expanding fireball.

Multi-strange baryons here play a key role since their abundance will not reach chemical equilibrium in a purely hadronic scenario. Attempts to describe the experimental observations within a hadronic scenario adding inverse

multiparticle reactions failed for Ω production. Moreover, in a QGP scenario, the enhancement is expected to increase with the strangeness content, as is observed. Such a pattern contradicts the expectations from rescattering in the hadronic fireball where multi-strange particle formation is hindered by high thresholds and low cross sections.

As previously described, strangeness is enhanced as soon as we approach energy densities corresponding to the critical temperature $T_c = 170$–180 MeV, which is still not sufficient to melt the χ_c and the J/ψ. As the energy density is proportional to the fourth power of temperature, further increase is needed to observe a χ_c suppression, and in order to melt the J/ψ directly one needs to raise the energy density by a factor of about 3.

Anomalous J/ψ suppression and the enhancement of multi-strange baryons are the two main results which lead us to claim that there is evidence for a new state of strongly interacting matter observed in Pb–Pb collisions at SPS energies. The observed effects coincide with those predicted for QGP and, when the quantitative predictions have been made, the measurements are in agreement with them. In fact particle abundance close to thermal and chemical equilibrium values and a hierarchy of enhancement increasing with the strangeness content of the particle were predicted about 20 years ago. Charmonium suppression was predicted a few years later. Other results obtained in the heavy-ion collision experiments at highest energies complete this interpretation.

V. FUTURE OUTLOOK

The heavy-ion program at the CERN SPS will continue at least for a few more runs. A systematic study of the onset of observed effects as a function of energy and size of the colliding system remains to be done. The continuation of the program later will depend on the physics results of ongoing analyses.

It is much more efficient to use heavy-ion beams in a colliding mode than in a fixed-target mode because the energy which can be used for particle production is the center-of-mass (cms) energy; the rest is just the energy used for the movement of the system as whole. For example, the maximum SPS energy, 200 GeV per nucleon, corresponds to a cms energy only of about 20 GeV per nucleon pair. On the other hand, in collider experiments the produced system does not move, and all the beam energy contributes to cms energy. Recently, the Relativistic Heavy-Ion Collider (RHIC), the first dedicated heavy-ion collider, has been commissioned at BNL. In a few years time, another collider capable of accelerating heavy ions, the Large Hadron Collider (LHC), will become operational at CERN. In this last section, we shall address the experimental programs foreseen for these new machines.

A. RHIC Program

RHIC is the first high-energy collider dedicated to heavy-ion physics. It will allow the study of pp, p–A and A–A interactions at a maximum collision energy of 200 GeV per nucleon pair for the Au–Au system. There are many advantages of the higher beam energies available at RHIC. The energy densities achieved will be higher than those at the SPS, and higher temperatures will be reached in large reaction volumes. RHIC will allow for precise measurements of QGP signatures with high statistical accuracy. At the same time, the importance of the hard component in the collisions at these energies opens up new channels and opportunities for new physics. In its first year, RHIC has reached the collision energy of 130 GeV per nucleon with the Au–Au system. Later, changes of beam energy and beam species will be made according to priorities.

There are two large experimental detectors (STAR and PHENIX) and two smaller ones (PHOBOS and BRAHMS) at RHIC. They each exploit a variety of detection techniques and they are designed to be complementary to one another in order to cover almost every proposed observable of QGP formation.

STAR, a large TPC-based detector, is designed to measure mainly hadronic signals. This includes the inclusive and single-event production of identified charged particles, as well as the inclusive production of neutral kaons, Λs, Ξs, and hadronic resonances. This program is accessible in its first year. The multi-strange particle detection capabilities will be improved with the upgrade of the silicon vertex detector, currently under construction.

PHENIX is primarily a detector designed to observe the leptonic and electromagnetic signals for QGP formation, i.e., electron and muon pairs, and photons. It comprises two muon arms on both sides, and central detectors which include tracking detectors, the Ring Imaging CHerenkov (RICH) detector and the electromagnetic calorimeter. In addition to leptons, the inclusive spectra of pions and kaons would be accessible.

PHOBOS is an experiment based on silicon-detector technology which is designed to look at RHIC collisions in an unbiased way (high-rate, minimum-bias data taking). It is going to measure charged hadrons, and possibly K^0_S and Λ, in the central region.

BRAHMS is a two-arm spectrometer capable of measuring identified charge particle spectra all the way from the central region to very close to the beam axis, depending on the position of its moveable arms. During the first year of RHIC, BRAHMS will be able to determine the K/π ratio over a large rapidity interval and measure high transverse momentum particles.

The first results from the RHIC experiments have been published recently. The basic event characteristics:

charged particle multiplicity, particle ratios, and momentum spectra have been measured. Very interesting results have been obtained concerning transverse momentum particle spectra. In order to describe them using proven models for heavy-ion collisions, one needs to assume jet quenching of high transverse momentum partons with a momentum loss of about 250 MeV/fm.

B. ALICE Experiment

At CERN a new accelerator, the Large Hadron Collider (LHC), is under construction. This will be a high-luminosity pp collider with a design energy of 7000 GeV per proton beam. From the beginning of the project it has been foreseen that this new collider will also run in heavy-ion mode, using the existing CERN infrastructure for injection. A center-of-mass energy up to 5500 GeV per nucleon pair will be achieved in Pb–Pb collisions. It will also be possible to accelerate smaller nuclei (at higher luminosity) and probably, in addition, to collide asymmetric systems, including p–A interactions.

In central Pb–Pb collisions at the LHC, the expected energy density is 5–10 times higher than that achieved at the CERN SPS, and about 4–5 times that for RHIC collisions. As a consequence the charged particle density will be extremely high; it is predicted that the number of charged particles in a central Pb–Pb collision can reach up to 60,000, which is the main challenge for the proposed detector.

There will be only one dedicated experiment for heavy-ion collisions at LHC: A Large Ion Collider Experiment (ALICE). Therefore, the ALICE collaboration made an effort to be able to detect as many as possible of the observables proposed to study a QGP formation. The proposed ALICE facility comprises central detectors covering a $\pm 45^\circ$ cone in the central region. This region is supposed to be completely baryon free (i.e., $\mu_B = 0$), which simplifies the physical interpretation with respect to the present experimental situation. The central detectors will be placed inside a 12-m-diameter solenoidal magnet.

Nearly all central ALICE detectors will be involved in hadron detection. The main tracking detector, a large cylindrical TPC will measure charged particle momenta and participate in particle identification via energy loss measurement with better than 10% precision. The Inner Tracking System (ITS), based on six layers of different silicon detectors, will reconstruct secondary vertices close to the interaction point, and hence detect the K^0_S, Λ, Ξ and Ω particles. Statistics for pions and kaons will be large enough to allow their study on an event-by-event basis. In addition, the ITS detector will improve both the momentum and energy loss measurements from the TPC.

Charged kaons with momentum below 0.5 GeV and protons below 1 GeV will be identified by energy loss measurement. The large Time-Of-Flight (TOF) detector at a radial distance of about 3.7 m will extend kaon identification up to 2 GeV and proton identification up to 3.5 GeV. For kaons these two measurements are complementary because slow kaons have a large probability of decaying before they reach the TOF detector. The ALICE experiment will also measure and identify electrons in the Transition Radiation Detector (TRD) placed between the TPC and TOF. In addition, two single-arm detectors, which cover only about 10% or less of the acceptance of the other central detectors, are foreseen in the central region. The High Momentum Particle Identification Detector (HMPID), which is based on a proximity focusing RICH detector, will be used for inclusive kaon and proton measurements up to a momentum of at least 3 and 5 GeV, respectively. The PHOton Spectrometer (PHOS) is an electromagnetic calorimeter made of lead–tungstate crystals and is dedicated to photon detection.

Dimuon signals will be measured in the forward muon spectrometer, covering angles between 2 and 10° from the beam on one side of the ALICE detector. It will use an additional dipole magnet for muon momentum measurements. The muon spectrometer has sufficient mass resolution to separate the three Υ states.

The ALICE experiment is at present entering the construction phase in order to be ready for LHC commissioning.

SEE ALSO THE FOLLOWING ARTICLES

ACCELERATOR PHYSICS AND ENGINEERING • ATOMIC AND MOLECULAR COLLISIONS • COLLIDER DETECTORS FOR MULTI-TEV PARTICLES • ELECTRODYNAMICS, QUANTUM • ELECTROMAGNETICS • PARTICLE PHYSICS, ELEMENTARY • QUANTUM CHROMODYNAMICS (QCD) • VACUUM TECHNOLOGY

BIBLIOGRAPHY

Greiner, W., and Schäfer, A. (1995). "Quantum Chromodynamics," 2nd edition, Springer-Verlag, Berlin.
Harris, J. W., and Müller, B. (1996). *Annu. Rev. Nucl. Part. Sci.* **46,** 71.
Heinz, U. (2001). *Nucl. Phys.* **A685,** 414.
Koch, P., Müller, B., and Rafelski, J. (1986). *Phys. Rep.* **142,** 167.
Margetis, S., Šafařík, K., and Villalobos Baillie, O. (2000). *Annu. Rev. Nucl. Part. Sci.* **50,** 299.
Müller, B. (1995). *Rep. Prog. Phys.* **58**(6), 611.
Shuryak, E. V. (1980). *Phys. Rep.* **61,** 71.
Stachel, J., and Young, G. R. (1992). *Annu. Rev. Nucl. Part. Sci.* **42,** 537.
Vogt, R. (1999). *Phys. Rep.* **310,** 197.



Neutrinos

S. M. Bilenky

Joint Institute for Nuclear Research
and Helsinki Institute of Physics

I. The History of the Neutrino. Pauli
II. The First Theory of β-Decay. Fermi
III. Lepton Number. Discovery of the Neutrino
IV. Nonconservation of Parity in β-Decay. The Two-Component Neutrino
V. Universal Current $\times$ Current Theory of Weak Interactions
VI. Discovery of the ν_μ. Electron and Muon Lepton Numbers
VII. Strange Particles in the Current $\times$ Current Interaction. The Cabibbo Angle
VIII. Glashow–Weinberg–Salam Theory of the Electroweak Interaction
IX. The Discovery of Neutral Currents
X. Deep Inelastic Neutrino–Nucleon Scattering and the Quark Structure of Nucleon
XI. Neutrino Masses. Introduction
XII. Discovery of the τ-Lepton, *b*- and *t*-Quarks. The Number of Flavor Neutrinos
XIII. Neutrino Mixing
XIV. Neutrino Oscillations
XV. Experiments on the Search for Neutrino Oscillations
XVI. The Neutrinoless Double β-Decay
XVII. Neutrino Masses from Experiments on the Measurement of the β-Spectrum of Tritium
XVIII. Conclusion

GLOSSARY

Charged current Provides transitions between neutrino and charged lepton $\nu_\mu \to \mu^-$ etc. between quarks with electric charges $2/3$ and $(-1/3)$ u $\to$ d, etc.

Dirac neutrino Particle that has electric charge equal to 0 and lepton number L equal to 1. The lepton number of Dirac antineutrino is equal to -1.

1 eV Energy gain of the electron that passes the potential difference 1 V; 1 MeV $= 10^6$ eV; 1 GeV $= 10^9$ eV.

Lepton number L_l of neutrino ν_l and charged lepton l^- is equal to 1 ($l = e, \mu, \tau$).

Lepton numbers of hadrons γ, $W^\pm$, and Z^0 are equal to 0.

Lepton numbers are conserved if neutrinos are massless particles.

Majorana neutrino Particle with equal to 0 electric charge and lepton number. Majorana neutrino is identical to its antineutrino.

Neutral current Provides transition $\nu_\mu \to \nu_\mu$, $u \to u$, etc.

Neutrino Elementary particle that has only weak interactions. Three types of neutrinos, ν_e, ν_μ, ν_τ, exist in nature.

Neutrino mixing Takes place if the states of flavor neutrinos ν_e, ν_μ, and ν_τ, are superpositions of states of massive neutrinos ν_1, ν_2, ν_3.

Neutrino oscillations Periodical transitions between different types of neutrinos in vacuum or in matter.

See-saw mechanism of neutrino mass generation A mechanism of generation of small Majorana neutrino masses.

The natural system $\hbar = c = 1$ All physical quantities in this system have a dimension of the mass M in some power. The energy, momentum, and mass have the same dimension, M. The angular momentum is a dimensionless quantity etc. Masses of particles are usually given in the unit of energy.

Two-component neutrino Neutrino with definite helicity, the projection of the spin on the detection of momentum. The helicity of neutrino (antineutrino) is equal to $-1(1)$.

NEUTRINOS are elementary particles. Three types of neutrinos exist in nature: electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ.

Neutrinos are members of the three lepton families. Other particles that are members of the families are, correspondingly, electron e^-, muon μ^-, and tau τ^-. There are also three families of quarks, particles with electric charges $2/3$ and $-1/3$ (in the units of electric charge of proton e): $(u,\ d)$, $(c,\ s)$, and $(t,\ b)$.

There are three fundamental interactions of elementary particles that are characterized by the strength of interaction: strong, electromagnetic, and weak. There is also the fourth gravitational interaction between particles. However, it is so weak that it can be neglected at all available energies.

The strong interaction is the interaction between quarks and gluons, neutral particles with spin 1. The interaction between quarks is the result of the exchange of gluons. Protons, neutrons, pions, and all other hadrons are bound states of quarks.

The electromagnetic interaction is the interaction between charged particles and γ-quanta. The Couloumb interaction between charged particles is due to the exchange of photons. Atoms of different elements are bound states of electrons and nuclei.

The weak interaction is the interaction between fundamental fermions (quarks, charged leptons, neutrinos) and charged $W^{\pm}$- and neutral Z^0-bosons, heavy particles with spin 1. For example, the β-decay of the neutron

$$n \to p + e^- + \bar{\nu} \qquad (1)$$

is due to the exchange of a charged W-boson between $e - \nu$ and $d - u$ pairs. Because of weak and electromagnetic interactions all particles , except the electron, proton, and neutrinos, are unstable. For example, the π^+-meson decays into μ^+ and ν_μ. The muon μ^+ decays into e^+, $\bar{\nu}_\mu$, and ν_e and so on. The weak and electromagnetic interactions are parts of a unified electroweak interaction.

Quarks take part in strong, electromagnetic and weak interactions: charged leptons—in electromagnetic and weak interactions. Neutrinos are exceptional particles: *neutrinos take part only in weak interactions*. The role of neutrinos in physics and astrophysics is determined mainly by this fact.

The investigation of neutrino processes allows one to obtain important information on the structure of the weak interaction. The detailed study of the scattering of high energy neutrinos on nucleons was very important for establishing the quark structure of nucleons. The detection of solar neutrinos allows one to investigate the internal invisible region of the sun, where solar energy is produced etc.

Neutrinos are also exceptional particles as for their internal properties. The masses of neutrinos are much smaller than the masses of the corresponding family partners. Because of small neutrino masses and so-called neutrino mixing, new neutrino processes, *neutrino oscillations*, become possible. Evidence in favor of neutrino oscillations was found in the Super-Kamiokande experiment in Japan. This discovery opened up a new field of research in neutrino physics: the physics of massive and mixed neutrinos.

The investigation of neutrino oscillations allows one to determine the value of very small neutrino mass-squared differences and neutrino mixing angles. It is a general belief that these internal neutrino properties are determined by physics at a new mass scale that is much larger than the scale of present-day physics (about 100 GeV).

We will list here the most important discoveries connected with neutrinos.

1. In 1954–1956 in the experiment of F. Reines and C. W. Cowan, the electron neutrino was discovered. In 1994, for this discovery, F. Reines was rewarded with the Nobel prize.
2. In 1956, in the experiment by C. S. Wu *et al*., the parity violation in β-decay was discovered.
3. In 1958, in the experiment by M. Goldhaber *et al*., the helicity of the neutrino was measured and evidence for the left-handed two-component neutrino was obtained.
4. In 1962, in the Brookhaven experiment, the second type of neutrino, the muon neutrino, was discovered. In 1988, for this discovery, L. Lederman, J. Steinberger, and M. Schwarz were rewarded with the Nobel prize.
5. In 1973, in experiments at the neutrino beam at CERN, a new type of weak interaction, Neutral Currents, was discovered.
6. In the 1980s, in experiments on the measurement of deep inelastic scattering on neutrinos on nucleons, the quark structure of nucleons was established.

7. In 1970, in the experiment of R. Davis *et al*., neutrinos from the sun were detected for the first time. In this experiment and also in GALLEX, SAGE, Kamiokande, and Super-Kamiokande solar neutrino experiments, the existence of a solar neutrino problem (deficit of solar ν_e's) was discovered.
8. In 1987, in the Kamiokande, IMB, and Baksan experiments, neutrinos from the explosion of the Supernova SN1987A in the Large Magellanien Cloud were detected.
9. In 1998, in the Super-Kamiokande experiment, compelling evidence in favor of oscillations of atmospheric muon neutrinos was found.

In this paper we present only the basics of neutrino. Those who are interested in the details must read the original papers and books. Some books and recent reviews are listed in the Reference Section.

I. THE HISTORY OF THE NEUTRINO. PAULI

The history of the neutrino is very interesting and instructive. It started in 1930 with the proposal by W. Pauli. At that time physicists knew very little about the world of elementary particles, and the electron e^- and proton p were considered as the only elementary particles. It was assumed that the nuclei of all elements heavier than hydrogen are bound states of electrons and protons.

In the framework of this assumption there were two fundamental problems. The first problem was connected with the spectrum of energies of electrons in β-decay, the process of the decay of a nucleus with emission of the electron. If a nucleus A is transferred into another nucleus A' with the emission of an electron then, according to the law of conservation of energy and momentum, the energy of the electron must be equal to M_A–$M_{A'}$ (M_A and $M_{A'}$ are masses of the initial and final nucleus). However, in experiments on the investigation of β-decay a continuous spectrum of energies, E up to $E_0 = M_A$–$M_{A'}$ was observed.

The second problem was the spin of the nitrogen $^7N_{14}$ and other nuclei. The atomic number of $^7N_{14}$ is equal to 14 and the charge of the nucleus is equal to $7e$. If we assume that nuclei are bound states of protons and electrons, the $^7N_{14}$ nucleus is a bound state of 14 protons and 7 electrons. The spins of the proton and electron are equal to $1/2$. Thus, for the spin of the $^7N_{14}$ nucleus we will obtain a half-integer value. However, from experiments on the investigation of the spectrum of $^7N_{14}$ molecules it was discovered that $^7N_{14}$ nuclei satisfy Bose statistics and, according to the theorem on the connection between spin and statistics, the spin of $^7N_{14}$ nucleus must be integer. This problem was known as the "nitrogen catastrophe."

In order to solve these problems, Pauli assumed that there exists in nature a neutral particle with spin $1/2$, with mass less than the electron mass, and with a mean free path much larger than the mean free path of a photon. Pauli called this particle "neutron" and he assumed that not only p's and e's but also "neutrons" are constituents of nuclei. This assumption allowed him to solve easily the problem of the spin of nitrogen and other nuclei. In fact, if in the $^7N_{14}$ nucleus there are an odd number of "neutrons" the spin of this nucleus will be integer.

In order to explain β-spectra, Pauli assumed that in β-decay the electron is emitted together with a "neutron" which is not detected in an experiment because of its large mean free path. The energy released in β-decay is shared beween the electron and the "neutron," and as a result the continuous spectrum of energies of electrons will be observed.

In 1932 the true neutron, a heavy particle with a mass approximately equal to the mass of the proton and the spin equal to $1/2$ was discovered by J. Chadwick in the nuclear reaction

$$^4\mathrm{He} + {}^9\mathrm{Be} \rightarrow {}^{12}\mathrm{C} + \mathrm{n}. \qquad (2)$$

This particle cannot be the light "neutron" proposed by Pauli.

Soon after the discovery of the neutron, it was assumed independently by W. Heisenberg, E. Majorana, and D. Ivanenko that the real constituents of nuclei are protons and neutrons. This hypothesis allowed one to explain all existing nuclear data. In particular, according to this hypothesis the nucleus $^7N_{14}$ is a bound state of seven protons and seven neutrons. Thus, the spin of this nucleus must be integer and the "nitrogen catastrophy" disappeared.

In 1933–1934, E. Fermi built the first theory of the β-decay assuming that nuclei are bound states of protons and neutrons.

II. THE FIRST THEORY OF β-DECAY. FERMI

The E. Fermi theory of β-decay was based on the Pauli hypothesis of the existence of a neutral, light, spin $1/2$ particle with a large mean free path. Fermi baptized this particle with the name neutrino (from Italian, neutral, small). Following Pauli, Fermi assumed that in β-decay the electron is emitted together with the neutrino. The problem was to understand how an electron–neutrino pair is emitted from a nucleus which is a bound state of protons and neutrons.

For Fermi it was important an analogy to electrodynamics. According to quantum electrodynamics a photon is emitted in the transition of an electron from an excited state of an atom into a lower state. In analogy with this process Fermi assumed that the electron–neutrino pair *is produced in the process of the quantum transition* of a neutron inside a nucleus into a proton

$$n \to p + e^- + \nu. \tag{3}$$

The first theory of β-decay was also built by Fermi in close analogy with quantum electrodynamics. The main quantity of the quantum field theory is the density of the energy of the interaction called the Hamiltonian of the interaction.

The Hamiltonian of the electromagnetic interaction has the form of the scalar product of the electromagnetic current $j_\alpha^{\rm em}(x)$ and electromagnetic field $A^\alpha(x)$

$$\mathcal{H}_I^{\rm em}(x) = e\, j_\alpha^{\rm em}(x) A^\alpha(x), \tag{4}$$

where e is the charge of the proton and the sum over $\alpha = 0, 1, 2, 3$ is assumed. The electric charge e characterizes the strength of the electromagnetic interaction.

The electromagnetic current $j_\alpha^{\rm em}$ is a 4-vector. The time component $j_0^{\rm em}$ is the density of charge, and the space components $j_i^{\rm em}$ $(i = 1, 2, 3)$ are components of the vector current. The electromagnetic field A^α is also a 4-vector: A^0 is a scalar potential and A^i is a component of a vector potential.

The electromagnetic current of protons is given by

$$j_\alpha^{\rm em}(x) = \bar{p}(x)\gamma_\alpha p(x). \tag{5}$$

Here γ_α are Dirac matrices and $p(x)$ is the proton field.

By analogy with Eq. (4) Fermi assumed that the Hamiltonian of β-decay had the form of the scalar product of the proton–neutron and electron–neutrino currents

$$\mathcal{H}_I^\beta = G_F(\bar{p}\gamma_\alpha n)(\bar{e}\gamma^\alpha \nu) + h.c. \tag{6}$$

Here G_F is the constant that characterizes the strength of the β-decay interaction (G_F is called a Fermi constant); $n(x)$ is the field of neutrons; $e(x)$ is the field of electrons; and $\nu(x)$ is the field of neutrino.

In the quantum field theory, $n(x)$ is the operator which annihilates the neutron in the initial state; the operator $\bar{p}(x)$ creates the proton in the final state; and operators $\bar{e}(x)$ and $\nu(x)$ create the final electron and neutrino.

The Fermi theory allows one to describe the β-decay of different nuclei. This theory, however, could not describe all β-decay data. In 1936 Gamov and Teller generalized the Fermi theory by including in the Hamiltonian additional scalar, tensor, pseudovector, and pseudoscalar terms with four additional interaction constants.

All β-decay data, which existed at that time, could be described by the Fermi–Gamov–Teller interaction. This was indirect evidence of the correctness of the Pauli–Fermi hypothesis of the neutrino. The direct proof of the existence of the neutrino was obtained only in the beginning of the 1950s in the F. Reines and C. L. Cowan experiment. We will discuss this experiment in the next section. Let us start with a discussion of the notion of *lepton number*.

III. LEPTON NUMBER. DISCOVERY OF THE NEUTRINO

As is well known, the total electric charge is conserved. This means that only such processes are allowed in which the sums of the electric charges of the initial and final particles are equal.

According to the quantum field theory every charged particle has its *antiparticle*, the particle with the same mass and spin but opposite charge. This general consequence of the quantum field theory is confirmed by all the existing experimental data. The antiparticle of the electron is the positron. The electron and positron have the same mass and the same spin, and the electric charges of the electron and positron are equal to $-e$ and e, respectively. The existence of the positron was predicted on the basis of the Dirac theory of the electron. The positron was discovered by C. D. Anderson in 1932. The antiparticle of the proton is the antiproton $\bar{p}$, the particle with electric charge equal to $-e$ and a mass equal to the proton mass. The antiproton was discovered in 1955 by O. Chamberlain, E. G. Segre *et al*. In 1956 the antineutron $\bar{n}$ was also discovered.

Except for the electric charge, there exist other conserved charges. One such charge is the *baryon number*. The baryon numbers of p and $\bar{p}$ are equal to 1 and -1, respectively. The baryon numbers of the $\pi^\pm$-mesons, $K^\pm$, γ-quantum, and other bosons are equal to 0. Due to the conservation of the baryon number the proton is a stable particle.

Let us now return to the neutrino. The fact that the neutrino is produced in β-decay together with the electron suggests that there exists some conserved quantum number that characterizes these particles. This number is called the lepton number. Let us assume that the lepton numbers of the electron and the neutrino are equal to 1 and the lepton numbers of the proton, neutron, photon, and other particles are equal to 0. According to the general theorem, as we mentioned earlier, the lepton number of the positron is equal to -1 and the antineutrino, the particle with the lepton number equal to -1, must exist. From the conservation of the lepton number it follows that in β-decay together with an electron an *antineutrino* is emitted. We will discuss later the experiment in which

evidence in favor of conservation of the lepton number was obtained.

Now we will discuss the experiment by F. Reines and C. L. Cowan in which the (anti)neutrino was discovered. In this experiment antineutrinos that are produced in β-decays of different nuclei, products of the fission mainly of ^{235}U in a reactor, were detected via the observation of the process

$$\bar{\nu} + p \to e^+ + n. \tag{7}$$

A reactor is a very intense source of antineutrinos: about 2×10^{14} antineutrinos per second are emitted per kilowatt, generated by the reactor. The power of a modern reactor is about 4 GW. Thus, about 10^{21} antineutrinos are emitted by a reactor per second. The experiment of Reines and Cowan was done at the Savannah River reactor in the United States. The detector in this experiment was a liquid scintillator loaded with cadmium. The positron, produced in the process (Eq. (7)), quickly slowed down to a rest and was annihilated with an electron into two γ-quanta with energies of about $E = m_e \simeq 0.51$ MeV, moving in opposite directions. These γ-quanta were detected by photomultipliers connected with scintillators.

The neutron produced in the process (Eq. (7)) was slowed down and was captured by the cadmium nucleus emitting γ-quanta with a total energy of about 9 MeV. These γ-quanta give several microseconds of a delayed signal in photomultipliers.

The probability of an interaction is characterized in physics by the *cross section* that has dimension of length (D. H. Perkins, 2000). In order to determine the cross section we will consider the flux of the particles that pass through the matter. Let us consider the element of the volume of the target presented in Fig. 1 (the axis x is oriented along the momentum $\vec{p}$ of the particles). The number of particles of the target in this volume is equal to $1 \cdot \Delta x \cdot \rho$ (1 is the unit square, ρ is the number density of the target). The cross section σ of a process of scattering, absorption, ... is the probability of the process per one particle in the target and per unit flux. For the change of the flux after passing through the element, drawn in Fig. 1, we have

$$\Delta I(x) = I(x + \Delta x) - I(x) = -\rho\sigma \Delta x I(x). \tag{8}$$

From Eq. (8) we obtain

$$I(x) = e^{-\rho\sigma x} I(0), \tag{9}$$

where x is the distance that the particles pass in the matter. We can rewrite Eq. (9) in the form

$$I(x) = e^{-\frac{x}{L}} I(0), \tag{10}$$

where $L = \frac{1}{\rho\sigma}$ is the mean free path.

For the cross section of the process (Eq. (7)) in the experiment of Reines and Cowan the following value was found

$$\sigma = (11 \pm 4) \times 10^{-44} \text{ cm}^2. \tag{11}$$

This is a very small cross section. Let us consider the propagation of reactor antineutrinos with an energy of a few megavolts in the earth. We have $\sigma \simeq 10^{-43}$ cm^2 and $\rho \simeq 10^{24}$ protons/cm^3. Thus, for the mean free path of a neutrino in the earth we have $L \simeq 10^{14}$ km. Remember that the earth's diameter is about 10^4 km. Thus, the probability for an antineutrino with an energy of a few megavolts to interact with the matter of the earth is about 10^{-10}!

The fact that the neutrino and antineutrino are different particles was established in the reactor experiment by R. Davis in 1955. As we discussed earlier, a reactor is a source of *antineutrinos*. If the lepton number is conserved, the reaction

$$\bar{\nu} + {}^{37}\text{Cl} \to \text{e}^- + {}^{37}\text{Ar} \tag{12}$$

is forbidden. In the Davis experiment a large tank with carbon tetrachloride (C_2Cl_4) liquid was irradiated over a long period of time by antineutrinos from the reactor. After every run, atoms of ^{37}Ar were extracted from the liquid by purging it with ^{4}He gas and they were put into a low-background Geiger counter. The γ-quanta produced in the e^- capture by ^{37}Ar were detected. No effect was observed. For the cross section of the process (Eq. (12)) it was found that

$$\sigma = (0.1 \pm 0.6) \times 10^{-45} \text{ cm}^2. \tag{13}$$

If the neutrino and antineutrino had been identical, for the cross section of the process in Eq. (12) the following value

$$\sigma = 2 \times 10^{-45} \text{ cm}^2 \tag{14}$$

would have been expected.

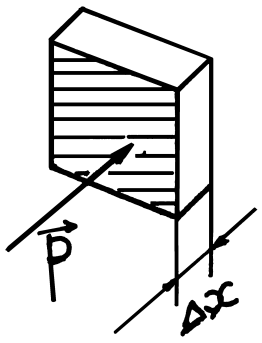


FIGURE 1 The element of the volume of the target.

IV. NONCONSERVATION OF PARITY IN β-DECAY. THE TWO-COMPONENT NEUTRINO

In 1956, in an experiment by C. S. Wu *et al*., parity nonconservation in β-decay was discovered. This was a very

important discovery in particle physics that drastically changed our understanding of the weak interaction and the neutrino.

In order to explain the phenomenon of parity violation we must remember that there are two types of vectors: (true) vectors and pseudovectors. The direction of the vector does not depend on the choice of the coordinate system. The direction of the pseudovector is changed if we change the handedness of the coordinate system. Typical vectors are momentum, coordinate, electric field etc. Angular momentum, spin, magnetic field etc. are pseudovectors.

Let us consider two coordinate systems: some right-handed systems and a system all axes of which are directed opposite to the direction of the axes of the first system. The second system is a left-handed one. If some vector $\vec{A}$ has coordinates A_i $(i = 1, 2, 3)$ in the first system, in the second system the coordinates of this vector will be $-A_i$. If $\vec{B}$ is a pseudovector with coordinates B_i in the first system, then in the second system its coordinates will be B_i (pseudovector changes direction). The transformation from the first system to the second one is called inversion or parity transformation.

In the Wu *et al*. experiment the β-decay of polarized nuclei ^{60}Co was investigated. The polarization (the average value of the spin) is a pseudovector. Let us consider in a right-handed system the emission of an electron at an angle θ between the direction of the polarization of the nucleus and the electron momentum. In the left-handed system the direction of the polarization is reversed and Fig. 2 corresponds to the emission of an electron at the angle $\pi - \theta$ (see Fig. 3). The emission of an electron at the angle $\pi - \theta$ in the right-handed system corresponds to the emission of an electron at the angle θ in the left-handed system. Thus, right-handed and left-handed systems are equivalent (the parity is conserved) if the number of electrons emitted at the angle θ and $\pi - \theta$ is equal.

In the experiment of Wu *et al*., a large asymmetry of the emission of the electrons with respect to the polarization of the nuclei was discovered. It was observed that electrons are emitted predominantly in the direction opposite to the direction of the polarization of the nuclei. Thus, it was proved that parity is not conserved in β-decay (the left-handed and right-handed systems are not equivalent). Later it was shown that parity is not conserved in other weak processes.

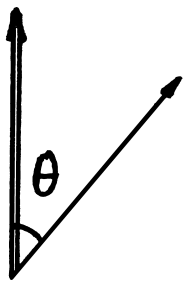


FIGURE 2 The emission of an electron at an angle θ in a right-handed system. The vector of the polarization of a nucleus is shown by the double line.

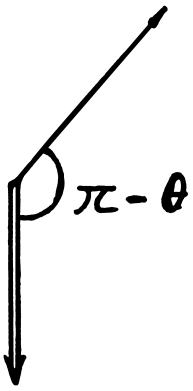


FIGURE 3 The emission of an electron at an angle $\pi - \theta$ in a left-handed system.

Let us consider now in a right-handed system the emission of a left-handed neutrino ν_L, a neutrino with the projection of the spin on the direction of momentum (helicity) equal to -1. In the left-handed system the projection of the spin on the vector of momentum of the neutrino will be equal to $+1$ (spin changes direction). Thus, if parity is conserved the probabilities of emission of the left-handed neutrino ν_L and the right-handed neutrino ν_R must be the same:

$$w(\nu_L) = w(\nu_R). \tag{15}$$

The discovery of the nonconservation of parity in weak interactions means that these probabilities are not equal.

In 1957 Landau, Lee, Yang, and Salam proposed the theory of the *two-component neutrino*. This theory is based on the assumption that the mass of neutrino is equal to 0. According to the theory of the two-component neutrino, the neutrino is a left-handed particle ν_L (or right-handed particle ν_R) and the anineutrino is a right-handed antiparticle $\bar{\nu}_R$ (or left-handed antiparticle $\bar{\nu}_L$), and the equality (Eq. (15)) is violated maximally.

The helicity of the neutrino was measured in 1957 in a spectacular experiment by Goldhaber *et al*. In this experiment neutrinos were produced in the K-capture

$$\begin{aligned} e^- + \mathrm{Gd} &\rightarrow \nu_\mathrm{e} + \mathrm{Sm}^* \\ &\quad\downarrow \\ &\quad\mathrm{Sm} + \gamma. \end{aligned} \tag{16}$$

The measurement of the circular polarization of a γ-quanta from the decay of Sm^* allowed one to determine

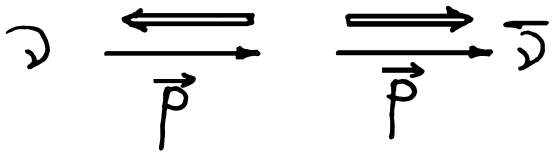


FIGURE 4 Helicities of the two-component neutrino and antineutrino. The vector of the spin (momentum) of neutrino is shown by double line (single line).

the helicity of the neutrino. The two-component neutrino theory was confirmed by this experiment. It was established that the neutrino is a particle with negative helicity. (see Fig. 4).

V. UNIVERSAL CURRENT × CURRENT THEORY OF WEAK INTERACTIONS

The discovery of parity nonconservation in the weak interaction and the confirmation of the theory of a two-component neutrino led to enormous progress in the development of the weak interaction theory (Feynman and Gell-Mann, Marshak and Sudarshan, 1958). At that time not only β-decay but also other weak processes were known. One such processes is μ-capture

$$\mu^- + p \to \nu + n. \tag{17}$$

The first idea on a possible interaction, responsible for the decay (Eq. (17)), was put forward by B. Pontecorvo. He compared the probabilities of μ-capture and K-capture of an electron by an nucleus and came to the conclusion that the corresponding interaction constants are of the same order. B. Pontecorvo assumed that there exists a *universal weak interaction* that includes $e-\nu$ and $\mu-\nu$ pairs. The idea of $\mu-e$ universality was also proposed by G. Puppi, O. Klein, and other authors.

Let us note that any fermion field $\psi(x)$ can be presented as a sum of a left-handed component $\psi_L(x)$ and a right-handed component $\psi_R(x)$

$$\psi(x) = \psi_L + \psi_R(x), \tag{18}$$

where

$$\psi_{L,R}(x) = \frac{1 \mp \gamma_5}{2}\psi(x) \tag{19}$$

and γ_5 is a Dirac matrix.

The fact that the neutrino is a particle with negative helicity means that the field of a neutrino is a left-handed field ν_L. Feynman and Gell-Mann, Marshak and Sudarshan assumed that in the Hamiltonian of the weak interactions *left-handed components of all fields* enter. If we make this assumption, the Hamiltonian of β-decay takes the very simple form

$$\mathcal{H}_I^{\beta} = \frac{G_F}{\sqrt{2}} 4\big(\bar{p}_L\gamma^{\alpha}\nu_L\big)(\bar{e}_L\gamma_{\alpha}\nu_L) + h.c. \tag{20}$$

This interaction, like the Fermi interaction, is characterized by only one interaction constant G_F. It contains, however, parity-conserving and parity-violating parts.

Assuming $\mu-e$ universality, Feynman and Gell-Mann proposed the theory that allowed one to describe all the weak processes known at that time and to predict new weak processes. They assumed that there exists a *weak current*

$$j^{\alpha} = 2\big[\bar{p}_L\gamma^{\alpha}n_L + \bar{\nu}_{eL}\gamma^{\alpha}e_L + \bar{\nu}_{\mu L}\gamma^{\alpha}\mu_L\big], \tag{21}$$

and that the Hamiltonian of the weak interaction has the simple current × current form

$$\mathcal{H}_I = \frac{G_F}{\sqrt{2}} j^{\alpha} j_{\alpha}^{+}, \tag{22}$$

where

$$j_{\alpha}^{+} = 2[\bar{n}_L\gamma_{\alpha}p_L + \bar{e}_L\gamma_{\alpha}\nu_{eL} + \bar{\mu}_L\gamma_{\alpha}\nu_{\mu L}] \tag{23}$$

is the conjugated current.

In Eq. (21) the neutrino field that enters into the current together with the electron field (muon field) is denoted by $\nu_e(\nu_\mu)$. We will call the corresponding particles the electron neutrino and the muon neutrino. It was proved in the famous 1962 Brookhaven neutrino experiment that ν_e and ν_μ are different particles. We will discuss this experiment in the next section. Now we will continue the discussion of the current × current Hamiltonian. There are two types of terms in the Hamiltonian (Eq. (22)): nondiagonal and diagonal. Nondiagonal terms are given by

$$\begin{aligned}\mathcal{H}_I^{\mathrm{nd}} = \frac{G_F}{\sqrt{2}} 4\big\{&\big[\big(\bar{p}_L\gamma^{\alpha}n\big)\,(\bar{e}_L\gamma_{\alpha}\nu_{eL}) + h.c.\big] \\ &+ \big[\big(\bar{p}_L\gamma^{\alpha}n_L\big)\,(\bar{\mu}_L\gamma_{\alpha}\nu_{\mu L}) + h.c.\big] \\ &+ \big[\big(\bar{e}_L\gamma^{\alpha}\nu_{eL}\big)(\bar{\nu}_{\mu L}\gamma_{\alpha}\mu_L) + h.c.\big]\big\}\end{aligned} \tag{24}$$

The first term of this expression is the Hamiltonian of β-decay of the neutron (Eq. (3)), of the process

$$\bar{\nu}_e + p \to e^+ + n \tag{25}$$

and other processes.

The second term of Eq. (24) is the Hamiltonian of μ-capture (Eq. (17)), of the process

$$\nu_\mu + n \to \mu^- + p \tag{26}$$

and other processes.

Finally the third term of Eq. (24) is the Hamiltonian of μ-decay

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu \tag{27}$$

and other processes.

Some processes that are described by nondiagonal terms of the Hamiltonian were observed in an experiment at the time when the current × current theory was

proposed. This theory *predicted also new weak processes* such as the process of elastic scattering of the electron antineutrino on the electron

$$\bar{\nu}_e + e \to \bar{\nu}_e + e \tag{28}$$

and others. The Hamiltonian of these new processes is given by the diagonal terms of Eq. (22):

$$\mathcal{H}^{\mathrm{d}} = \frac{G_{\mathrm{F}}}{\sqrt{2}} 4\left[\left(\bar{\nu}_{eL}\gamma^{\alpha} e_L\right)\left(\bar{e}_L \gamma_{\alpha} \nu_{eL}\right) + \cdots\right]. \tag{29}$$

The predicted cross section of the process in Eq. (28) is very small and its measurement was a difficult problem. After many years of efforts, F. Reines *et al*. observed the process (Eq. (28)) with reactor antineutrinos.

The detailed investigation of this and other similar processes showed that except for the diagonal terms in the Hamiltonian of such processes there are additional Neutral Current (NC) terms. We will discuss NC later.

There were two alternatives for the weak interaction theory: the current × current theory we described and the theory with an intermediate vector charged $W^{\pm}$-boson. We will discuss now this last theory. Let us assume that there exist heavy particles $W^{\pm}$ with spin equal to 1 and charges $\pm e$ and that the fundamental weak interaction has the form

$$\mathcal{H} = \frac{g}{2\sqrt{2}} j_{\alpha} W^{\alpha} + h.c., \tag{30}$$

where g is the interaction constant and current j^{α} is given by Eq. (21). It is possible to show that at energies much less than the mass of the W-boson m_W for the processes with a virtual (intermediate) W-boson, the current × current theory and the theory with the W-boson are equivalent.

In fact, let us consider μ-decay:

$$\mu^- \to e^- + \bar{\nu}_e + \nu_{\mu}. \tag{31}$$

In quantum field theory the processes are described by Feynman diagrams that are a convenient language of physicists. With the help of special rules Feynman's diagrams allow one to calculate the probabilities of decay, cross sections, and other measurable quantities.

In the current × current theory the decay (Eq. (31)) is the process of the first order in the perturbation theory in the constant G_F and its Feynman diagram is presented in Fig. 5. In the theory with the W-boson the decay (Eq. (31)) is the process of the second order in the perturbation theory in the constant g. The Feynman diagram of the process is presented in Fig. 6. Figure 6 describes the following chain of transitions: the initial muons emit the final ν_{μ} and a virtual W^-; the vector boson propagates in the virtual state; the virtual boson decays into the final e^- and $\bar{\nu}_e$. At every vertex the conservation of 4-momenta takes place. This ensures the conservation of energy and momentum

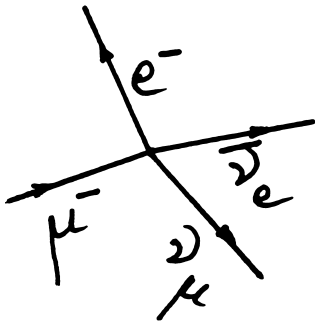


FIGURE 5 The Feynman diagram of the decay $\mu^- \to e^- \bar{\nu}_e \nu_{\mu}$ in the current × current theory.

for the process. For a free particle the square of the 4-momentum is equal to the square of its mass. This is not the case for a virtual particle. For the square of the 4-momentum of the W-boson we have $q^2 = (p - p')^2$ where p and p' are the momenta of muon and ν_{μ}. If the mass-squared of the W-boson m_W^2 is much larger than q^2 then in this case the propagator of the W-boson (dashed line in Fig. 6) gives a $1/m_W^2$ contribution and the diagrams in Fig. 5 and Fig. 6 are equivalent if the Fermi constant is connected to the constant g by the relation

$$\frac{G_{\mathrm{F}}}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \tag{32}$$

The universal current × current theory of the weak interactions as well as the theory with the intermediate W-boson allowed one to describe the data of many experiments. Nevertheless both theories could not be considered as final theories of the weak interactions. The main reason was that both theories were not renormalizable quantum field theories. The probability of transitions calculated in the lowest order of perturbation theory was in good agreement with experimental data. However, the corrections

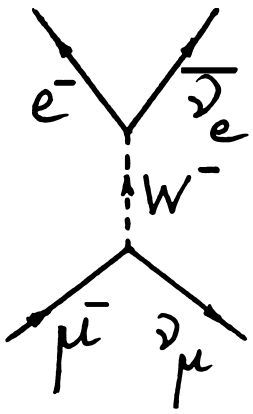


FIGURE 6 The Feynman diagram of the decay $\mu^- \to e^- \bar{\nu}_e \nu_{\mu}$ in the theory with intermediate W boson.

due to higher orders of perturbation theory cannot be calculated: they contained divergent integrals from which the finite corrections by the renormalization of masses and interaction constants cannot be found. At that time the only known renormalizable theory that allowed one to calculate the higher-order correction was quantum electrodynamics. This theory was in excellent agreement with the experiment.

The enormous progress in the understanding of weak interactions and the neutrino is connected with the development of the renormalizable unified theory of weak and electromagnetic interactions, so-called the Standard Model (SM). We will discuss this theory later.

VI. DISCOVERY OF THE ν_μ. ELECTRON AND MUON LEPTON NUMBERS

The mass of the muon is approximately 200 times larger than the electron mass ($m_\mu = 105.66$ MeV and $m_e = 0.51$ MeV). From the very beginning of the investigation of muons the possible decay channel

$$\mu \to e + \gamma \tag{33}$$

was searched for. No indications in favor of this decay were found. In the first experiments that were done at the end of the 1940s, for the upper bound of the ratio R of the probability of the decay $\mu^+ \to e^+ + \gamma$ to the probability of the decay $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$, which is the main decay channel of muon, it was found that $R < 10^{-2}$. At present the upper bound of R is found to be $R < 1.2 \times 10^{-11}$.

The fact that the decay $\mu^+ \to e^+ + \gamma$ was not found could be explained either by some dynamical suppression of the decay or by the existence of some conserved fundamental quantum numbers that disinguish muon and electron.

If the muon and electron neutrinos are the same particles the process in Eq. (33) is possible. At the end of 1950s the probability of the decay $\mu \to e + \gamma$ was calculated in a nonrenormalizable theory with a cutoff and the estimated value of the ratio R was larger than existed at that time the upper bound ($R < 2 \times 10^{-5}$). This was a possible indication that ν_e and ν_μ were different particles. It was necessary, however, to prove this in a direct experiment. Such an experiment was proposed by B. Pontecorvo in 1959 and M. Schwarz in 1960, and it was done by L. Lederman, M. Schwarz, J. Steinberger *et al*. in 1962, in Brookhaven, New York.

The Brookhaven experiment was the first experiment that had been done with neutrinos from an accelerator. The beam of pions in this experiment was produced by the bombardment of a Be target by 15-GeV protons. Neutrinos were produced in the decays of pions in a decay channel (about 20 m long). After the channel there was an iron shielding, 13.5 m thick, in which charged particles were absorbed. After the shielding there was a neutrino detector (about 10 tons).

There are two decay modes of the π^+:

$$\pi^+ \to \mu^+ + \nu_\mu \tag{34}$$

$$\pi^+ \to e^+ + \nu_e. \tag{35}$$

In the Feynman–Gell–Mann theory the decay (Eq. (35)) is strongly suppressed. In fact, let us consider this decay in the rest frame of the pion. In this frame the e^+ and neutrino are moving in opposite directions. The helicity of the neutrino is equal to -1. If we neglect the mass of the positron the helicity of the positron will be the same as the helicity of the antineutrino and it will be equal to $+1$. Thus, the projection of the total angular momentum on the direction of the momentum of the positron will be equal to 1. However, the spin of the pion is equal to 0 and the projection of the initial angular momentum on any direction is equal to 0. Thus, in the limit $m_e \to 0$, the decay (Eq. (35)) is forbidden. For $m_e \neq 0$ the decay in (Eq. (35)) is not forbidden but it is strongly suppressed with respect to the decay in Eq. (34). The ratio of the probabilities of the decays in Eqs. (35) and (34) is given by

$$R = \left(\frac{m_e}{m_\mu}\right)^2 \frac{\left(1 - \dfrac{m_e^2}{m_\pi^2}\right)^2}{\left(1 - \dfrac{m_\mu^2}{m_\pi^2}\right)^2} \simeq 1.2 \cdot 10^{-4}. \tag{36}$$

Thus, decays of pions are the source of mainly muon neutrinos.

In the neutrino detector the processes of the interaction of neutrinos with nucleons were observed. If ν_μ and ν_e are different particles, in this case muons produced in the process

$$\nu_\mu + N \to \mu^- + X \tag{37}$$

will be observed in the detector (X means any hadrons). If ν_μ and ν_e are the same particles, in this case the process

$$\nu_\mu + N \to e^- + X \tag{38}$$

is also possible and in the detector, muons *and electrons* will be observed. Due to the $\mu - e$ universality of weak interactions the number of muons and electrons that will be observed in this latter case will be practically the same.

In the Brookhaven experiment, 29 muons were detected. Only six electron events were observed. All electron events could be explained by background. Thus, it was proved that the process in Eq. (38) is forbidden, i.e., that *muon and electron neutrinos are different particles*.

To explain the results of the Brookhaven and other experiments, it is necessary to introduce two conserved

TABLE I Lepton Numbers of Particles

	ν_e, e^-	ν_μ, μ^-	*Hadrons*, $\gamma, \ldots$
L_e	1	0	0
L_μ	0	1	0

lepton numbers: the electron lepton number L_e and the muon lepton number L_μ. The electron and muon lepton numbers of the particles are given in Table I.

From the conservation of the total electron and total muon lepton number

$$\sum L_e = \text{const}, \qquad \sum \mathrm{L}_\mu = \text{const}, \tag{39}$$

it follows that the decays

$$\mu^+ \to e^+\gamma, \quad \mu^+ \to e^+e^-e^+, \tag{40}$$

and others are forbidden.

VII. STRANGE PARTICLES IN THE CURRENT × CURRENT INTERACTION. THE CABIBBO ANGLE

In the 1950s, a large family of new particles $K^\pm$, K^0, $\bar{K}^0$, Λ, $\Sigma^{\pm,0}$, $\Xi^{-,0}$ was discovered and investigated in detail. These particles were called strange particles.

Strange particles are produced in nucleon–nucleon and pion–nucleon collisions only in pairs. For example, the process

$$\pi^- + p \to \Lambda + K^0 \tag{41}$$

in which two strange particles are produced, was observed. On the other hand, it was shown that the process of production of one strange particle

$$n + p \to \Lambda + p \tag{42}$$

was forbidden.

In order to explain the fact of the production of strange particles in pairs in nucleon–nucleon and pion–nucleon collisions it was necessary to introduce a conserved quantum number that distinguished strange particles from nonstrange ones (nucleons, pions, and others). This quantum number was called *strangeness* S . If we assume that the nucleon and pion have $S=0$, K^0 has $S=1$, and Λ has $S=-1$, then the process in Eq. (41) is allowed and the process in Eq. (42) is forbidden.

Strange particles are unstable and in their decay the strangeness is not conserved. The investigation of such processes as

$$\begin{aligned} &K^+ \to \mu^+ + \nu_\mu, \qquad \Lambda \to n + e^- + \bar{\nu}_e, \\ &\Sigma^- \to n + e^- + \bar{\nu}_e \quad \Xi^- \to \Lambda + e^- + \bar{\nu}_e, \end{aligned} \tag{43}$$

and others allowed one to formulate two phenomenological rules that govern these decays.

1. In the decays of strange particles the strangeness is changed by 1 $|\Delta S| = 1$.
2. The rule $\Delta Q = \Delta S$ is satisfied ($\Delta Q = Q_f - Q_i$ and $\Delta S = S_f - S_i$, $Q_i(S_i)$ and $Q_f(S_f)$ are initial (final) total charge and strangeness of hadrons).

According to rule 1 the decay

$$\Xi^- \to \Lambda + e^- + \bar{\nu}_e \tag{44}$$

is allowed and the decay

$$\Xi^- \to n + e^- + \bar{\nu}_e \tag{45}$$

is forbidden (the strangeness of Ξ is equal to -2).

According to rule 2 the decay

$$\Sigma^+ \to n + e^+ + \bar{\nu}_e \tag{46}$$

is forbidden (the strangeness of $\Sigma^\pm$ is equal to -1). All these predictions are in perfect agreement with the experiment.

In 1964 Gell-Mann and Zweig made the crucial assumption that the proton, neutron, pions, strange particles, and all other hadrons are bound states of *quarks*. Quarks are particles with spin 1/2, electric charges 2/3 or $-1/3$ (in the units of the electric charge of the proton), and baryon number 1/3. Gell-Mann and Zweig introduced three quarks, constituents of nonstrange and strange hadrons: nonstrange quarks u and d with charges 2/3 and $-1/3$, respectively, and a strange quark s with charge $-1/3$ and strangeness -1. In the framework of the quark model the proton is a bound state of two u-quarks and a d-quark; the π^+-meson is a bound state of a u-quark and a $\bar{d}$-antiquark; the K^+-meson is a bound state of a u-quark and a $\bar{s}$-antiquark; the Λ-hyperon is a bound state of a u-quark; d-quark, and s-quark, etc. The correctness of the quark hypothesis was confirmed by numerous experiments. Later we will discuss the role of neutrinos in revealing the quark structure of the nucleon.

If nucleons, pions, strange particles, and other hadrons are not elementary particles and are bound states of quarks it is natural to assume that the fundamental weak interaction is the interaction of leptons, neutrinos, *and quarks*. For example, the Feynman diagram of the β-decay of the neutron has the form presented in Fig. 7.

Strange particles were included in the current × current interaction by N. Cabibbo. The current $2\bar{p}_L\gamma_\alpha n_L$ does not change strangeness and changes the charge by one. The only quark current that has such properties is $\bar{u}_L\gamma_\alpha d_L$. The only quark current that changes charge by 1 and changes strangeness is $\bar{u}_L\gamma_\alpha s_L$. It is easy to see that this current automatically satisfies rules 1 and 2.

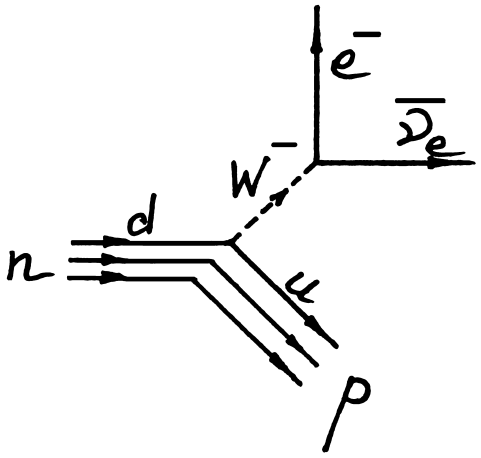


FIGURE 7 The Feynman diagram of the proceess $n \to pe^-\bar{\nu}_e$ in the quark model.

It was also known from experimental data that decays of strange particles are suppressed with respect to the decays of nonstrange particles. To take into account this suppression, N. Cabibbo introduced an additional parameter. This parameter is called the Cabibbo angle θ_C. For the quark weak current he proposed the following expression:

$$j_\alpha^{\mathrm{C}} = 2[\cos\theta_C \bar{u}_L \gamma_\alpha d_L + \sin\theta_C \bar{u}_L \gamma_\alpha s_L]. \qquad (47)$$

It was shown that the weak interaction Hamiltonian with such a current allows one to describe numerous experimental data. From the analysis of the data it was found that $\sin\theta_C \simeq 0.2$.

Let us write down the total weak current in the form

$$j_\alpha = 2[\bar{\nu}_{eL}\gamma_\alpha e_L + \bar{\nu}_{\mu L}\gamma_\alpha \mu_L + \bar{u}_L \gamma_\alpha d'_L], \qquad (48)$$

where

$$d'_L = \cos\theta_C d_L + \sin\theta_C s_L \qquad (49)$$

is the Cabibbo mixture of the fields of d and s quarks.

Notice that there are two lepton terms and one quark term in the expression (Eq. (48)). In 1970 it was shown by Glashow, Illiopulos, and Maiani that in the framework of the weak interaction with the current (Eq. (48)) the probability of the decays of the type

$$K^+ \to \pi^+ + \nu + \bar{\nu} \qquad (50)$$

in which $\Delta S = -1$ and $\Delta Q = 0$ is larger than the upper bound obtained in the experiment. In order to avoid this problem they assumed that there exists a fourth quark with charge $2/3$ and that there is an additional term in the weak current in which a field of the new quark enters. This new quark was called the charm quark (c). The weak currents took the form

$$j_\alpha = 2[\bar{\nu}_{eL}\gamma_\alpha e_L + \bar{\nu}_{\mu L}\gamma_\alpha \mu_L + \bar{u}_L \gamma_\alpha d'_L + \bar{c}_L \gamma_\alpha s'_L], \qquad (51)$$

where

$$\begin{aligned} d'_L &= \cos\theta_C d_L + \sin\theta_C s_L \\ s'_L &= -\sin\theta_C d_L + \cos\theta_C s_L. \end{aligned} \qquad (52)$$

The symmetry between leptons and quarks was restored.

In 1976 the first charmed mesons $D^{\pm,0}$ were discovered in experiments on $e^+ - e^-$ colliders. Later charmed baryons were also observed.

VIII. GLASHOW–WEINBERG–SALAM THEORY OF THE ELECTROWEAK INTERACTION

The current × current theory of the weak interaction and the theory with heavy charged vector W-boson in the lowest-order perturbation theory allowed one to describe all existing experimental data. However, both theories were only effective nonrenormalizable theories: in the framework of these theories it was not possible to calculate corrections due to higher orders of perturbation theory.

The modern renormalizable theory of the weak interaction (S. L. Glashow (1961), S. Weinberg (1967), and A. Salam (1968)) appeared as a result of *unification of the weak and electromagnetic interactions into an electroweak interaction*. This theory which is called the Standard Model (SM) is one of the greatest achievements of particle physics in the 20th century. This theory successfully predicted the existence of families of new hadrons (charmed, bottom, top), new interactions (Neutral Currents), the existence of $W^\pm$- and Z^0-bosons and their masses, etc. All predictions of the Standard Model are in perfect agreement with existing experimental data including very precise data that were obtained in experiments at $e^+ - e^-$ colliders at CERN (Geneva) and at SLAC (Stanford).

The Hamiltonian of the electromagnetic interaction has the form of the scalar product of the electromagnetic current and the electromagnetic field

$$\mathcal{H}_I^{\mathrm{em}} = e j_\alpha^{\mathrm{em}} A^\alpha. \qquad (53)$$

Here

$$j_\alpha^{\mathrm{em}} = \sum_{l=e,\mu} (-1)\bar{l}\gamma_\alpha l + \sum_{q=u,d,\ldots} e_q \bar{q}\gamma_\alpha q \qquad (54)$$

is the electromagnetic current of leptons and quarks ($e_u = 2/3$, $e_d = -1/3, \ldots$).

The electromagnetic field A_α is determined up to the derivative of an arbitrary function. The observable physical quantities are not changed if we make the following transformation

$$A_\alpha(x) \to A_\alpha(x) - \frac{1}{e}\frac{\partial \Lambda(x)}{\partial x^\alpha}, \qquad (55)$$

and change correspondingly the unobserved phases of the quark and lepton fields. In Eq. (55) $\Lambda(x)$ is an arbitrary function. This invariance is called gauge invariance and the electromagnetic field is called the gauge field. The gauge field must be a spin 1 vector field.

Weak and electromagnetic interactions are unified on the basis of gauge invariance. The gauge fields must include not only the electromagnetic field but also the fields of the charged vector particles. The generalized gauge invariance that includes gauge fields of neutral and charged particles is called the Yang–Mills gauge invariance.

The Standard Model is based on spontaneously broken SU(2) × U(1) gauge symmetry which assumes the existence of, in addition to the massless photon, three massive spin 1 particles: two charged and one neutral. The Hamiltonian of the Standard Model has the following form

$$\mathcal{H}_I = \left(\frac{g}{2\sqrt{2}} j_\alpha W^\alpha + h.c.\right) + \frac{g}{2\cos\theta_W} j_\alpha^0 Z^\alpha + e j_\alpha^{\rm em} A^\alpha \tag{56}$$

Here

$$j_\alpha^0 = 2j_\alpha^3 - 2\sin^2\theta_W j_\alpha^{\rm em} = \sum_l \bar{\nu}_{lL}\gamma_\alpha \nu_{lL} + \cdots \tag{57}$$

is the so-called neutral current and θ_W is a parameter (weak angle).

The first term of Eq. (56) is the charged current (CC) interaction, that we discussed earlier; the second term is a new neutral current (NC) interaction; and the third term is the well-known electromagnetic interaction.

Thus, the unified theory of the electroweak interaction *predicted* the existence of a new neutral vector boson Z^0 and a new NC interaction.

This new interaction means the existence of new weak interaction processes. The first such processes were discovered in 1973 at CERN. We will discuss this discovery in the next chapter. Charged $W^\pm$- and neutral Z^0-bosons were discovered at CERN in 1983.

IX. THE DISCOVERY OF NEUTRAL CURRENTS

Beams of neutrinos (antineutrinos) that can be obtained at accelerators are mainly the beams of muon neutrinos (antineutrinos) from the decays of pions with a small (few %) admixture of electron neutrinos and antineutrinos from the decays of other particles. We will discuss processes that were observed on the beam of high energy neutrinos at CERN in the beginning of the 1980s.

If the muon neutrino (antineutrino) interacts with a nucleon the following processes

$$\nu_\mu(\bar{\nu}_\mu) + N \to \mu^-(\mu^+) + X \tag{58}$$

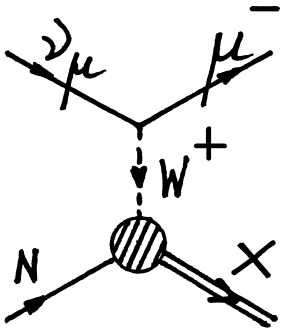


FIGURE 8 The Feynman diagram of the inclusive process $\nu_\mu + N \to \mu^- + X$.

are possible. The diagram of the neutrino process is presented in Fig. 8. In Fig. 8 due to the CC interaction in Eq. (56) the initial ν_μ produces the final μ^-- and virtual W^+-boson. The virtual W^+-boson propagates and is absorbed by a quark inside of the nucleon; the initial nucleon is transferred into final hadron state (nucleon or nucleon and pions and etc). If only the muon is observed and the effective mass of the final hadrons is much larger than the mass of the nucleon, the process is called a deep inelastic inclusive process.

The process in Eq. (58) is a typical weak interaction process: absorption of a neutrino is *accompanied* by the production of a corresponding charged lepton (like in β-decay of the neutron, the production of an electron is accompanied by the emission of a $\bar{\nu}_e$ and so on)

If there is NC interaction in Eq. (56) the deep inelastic NC processes

$$\nu_\mu(\bar{\nu}_\mu) + N \to \nu_\mu(\bar{\nu}_\mu) + X \tag{59}$$

with a neutrino (and not a muon) in the final state becomes possible (see diagram Fig. 9). In the diagram in Fig. 9, due to the NC interaction in Eq. (56), the initial ν_μ produces the final ν_μ and a virtual Z^0-boson. The virtual Z^0-boson propagates and is absorbed by a quark inside of nucleon. As a result of this absorption the initial nucleon is transferred in a final hadron state.

Such a new weak process was first observed in CERN in 1973 in the bubble chamber "Gargamelle." It was found that the ratio of the NC and CC cross sections is approximately equal to 0.3. Thus, the investigation of neutrino processes allowed one to *discover new weak processes*. The discovery of NC processes and their detailed investigation were crucial to the confirmation of the unified theory of the weak and electromagnetic interactions.

Another NC process is the process of elastic scattering of $\nu_\mu(\bar{\nu}_\mu)$ on the electron (see diagram Fig. 10)

$$\nu_\mu(\bar{\nu}_\mu) + e \to \nu_\mu(\bar{\nu}_\mu) + e. \tag{60}$$

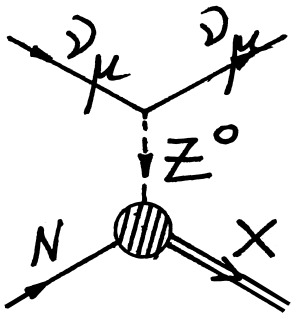


FIGURE 9 The Feynman diagram of the inclusive process $\nu_\mu + N \to \nu_\mu + X$.

The cross sections of these processes was measured at high energies by the CHARM collaboration at CERN. For the cross sections it was found

$$\sigma_{\nu_\mu e} = (1.9 \pm 0.4 \pm 0.4)10^{-42} \frac{E}{\text{GeV}} \text{cm}^2. \quad (61)$$

$$\sigma_{\bar{\nu}_\mu e} = (1.5 \pm 0.3 \pm 0.4)10^{-42} \frac{E}{\text{GeV}} \text{cm}^2. \quad (62)$$

From these measured cross sections the following value of the parameter $\sin^2\theta_W$ was found

$$\sin^2\theta_W = 0.215 \pm 0.032 \pm 0.012. \quad (63)$$

This value of $\sin^2\theta_W$ is in agreement with the values obtained from the measurement of other NC effects.

Only the NC interaction gives contribution to the cross sections of the processes in Eq. (60). The processes of elastic scattering of the ν_e and $\bar{\nu}_e$ on the electron

$$\nu_e(\bar{\nu}_e) + e \to \nu_e(\bar{\nu}_e) + e \quad (64)$$

are possible due to Z and W exchanges (see diagram Fig. 11). Cross sections of these processes were measured in the experiments at reactors and at Los Alamos Meson Factory. Notice that the CC part of elastic scattering of the ν_e on the electron (diagram Fig. 11) plays a crucial role in the propagation of neutrinos through matter (see later).

Effects of neutral currents were also measured in the inclusive deep inelastic scattering of electrons and muons

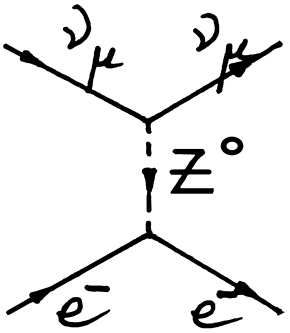


FIGURE 10 The Feynman diagram of the $\nu_\mu + e \to \nu_\mu + e$.

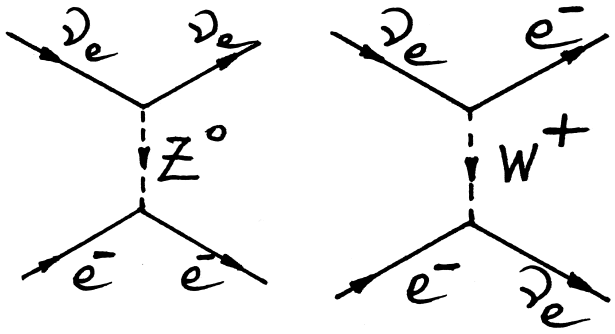


FIGURE 11 The Feynman diagram of the process $\nu_e + e \to \nu_e + e$.

on nucleons (SLAC and CERN) and in atomic transitions. All NC data confirm perfectly the Standard Model of electroweak interactions. For the value of the parameter $\sin^2\theta_W$, which characterizes NC, it was found

$$\sin^2\theta_W = 0.23155 \pm 0.00019. \quad (65)$$

X. DEEP INELASTIC NEUTRINO–NUCLEON SCATTERING AND THE QUARK STRUCTURE OF NUCLEON

Experiments on the investigation of the deep inelastic CC neutrino processes

$$\nu_\mu + N \to \mu^- + X \quad (66)$$

$$\bar{\nu}_\mu + N \to \mu^+ + X \quad (67)$$

that have been done in Fermilab and CERN in the 1970s and the 1980s were very important for establishing the quark structure of nucleon. In particle physics these experiments and also the experiment on deep inelastic scattering of the electron (muon) on nuclei played a role of the famous Retherford experiments in atomic physics. Like the Reserford experiments, which allowed one to establish the existence of heavy nuclei in atoms, these experiments allowed one to establish the existence of quarks and antiquarks in nucleons.

Let us first introduce the variables that are usually used to describe deep inelastic scattering:

$$x = \frac{Q^2}{2pq}, \qquad y = \frac{pq}{pk}, \qquad E = \frac{pk}{M}, \quad (68)$$

where $q = k - k'$ is the 4-momentum transfer (4-momentum of the W-boson), $Q^2 = -q^2$ and M is the mass of the nucleon. (p, k, and k' are 4-momenta of the initial nucleon, neutrino, and final muon).

From the conservation of energy and momentum it follows that the variable x takes values in the interval

$0 \le x \le 1$. In the lab system (the system where the initial nucleon is at rest) the variable y becomes

$$y = \frac{E - E'}{E}, \tag{69}$$

where E and E' are the energies of the initial neutrino and final muon. Thus, y is the relative energy that is transferred to the hadrons. At high energies $0 \le y \le 1$. Let us introduce also the variable $\nu = pq/M$. In the region of deep inelastic scattering $\nu \gg M$ and $Q^2 \gg M^2$.

Let us consider the processes of interaction of the neutrino with the u and d quarks and antiquarks,

$$\nu_\mu + d \to \mu^- + u \tag{70}$$

$$\nu_\mu + \bar{u} \to \mu^- + \bar{d}. \tag{71}$$

At high energies the masses of quarks can be neglected, and from the conservation of energy and momentum it follows that virtual W-boson interacts only with those quarks the momentum of which is a fraction x of the nucleon momentum p. The contributions to the differential cross section of the process $\nu_\mu + p \to \mu^- + X$ of the subprocesses in Eqs. (70) and (71) are given by the following equation:

$$\frac{d\sigma_{\nu p}}{dxdy} = 2\sigma_0 x[d(x) + (1 - y)^2 \bar{u}(x)]. \tag{72}$$

Here

$$\sigma_0 = \frac{G_F^2}{\pi} ME \simeq 1.5 \cdot 10^{-38} \frac{E}{\text{GeV}} \text{ cm}^2 \tag{73}$$

is the total cross section of the interaction of the neutrino with a point-like particle with mass M, $d(x)$ and $\bar{u}(x)$ are number-densities of the d-quarks and u-antiquarks with momentum xp in the proton.

The dependence of the cross sections on the variable y is determined by the helicities of the initial particles. Let us consider the process in Eq. (71) in the center of the mass system. In this system the total momentum of the initial (final) particles is equal to 0. The helicity of the neutrino is equal to -1 and the helicity of the antiquark $\bar{u}$ is equal to 1 (we neglect quark masses). Thus, the projection of the total angular momentum on the direction of the momentum of the neutrino is equal to $2 \times (-1/2) = -1$. Let us consider the emission of a μ^- in the backward direction. This case corresponds to $y = 1$ (the energy, which is transferred to the hadrons, is maximal). The helicity of the μ^- is equal to -1, and the projection of the total angular momentum on the direction of the momentum of the neutrino is equal to $+1$ in this case. Thus, the emission of the μ^- in the backward direction is forbidden by the conservation of the total angular momentum. This corresponds to the $(1 - y)^2$ dependence of the contribution of the antiquarks to the cross section.

In the case of the process in Eq. (70) the projections of the total angular momentum on the direction of momentum of the neutrinos are equal to 0 for the initial and final particles. Thus, emission of μ^- in a backward direction is allowed. This corresponds to the absence of y-dependence in the contribution of quark to the cross section in Eq. (73).

In neutrino experiments the target nuclei are usually nuclei with approximately equal numbers of protons and neutrons. For the averaged cross sections over the p and n we have

$$\frac{d\sigma_{\nu N}}{dxdy} = \sigma_0 x[q(x) + (1 - y)^2 \bar{q}(x)]. \tag{74}$$

Here

$$\begin{aligned} q(x) &= u(x) + d(x) \\ \bar{q}(x) &= \bar{u}(x) + \bar{d}(x). \end{aligned} \tag{75}$$

For the averaged cross section of the process

$$\bar{\nu}_\mu + N \to \mu^+ + X, \tag{76}$$

we have

$$\frac{d\sigma_{\bar{\nu}N}}{dxdy} = \sigma_0 x[(1 - y)^2 q(x) + \bar{q}(x)]. \tag{77}$$

The expressions in Eqs. (73), (74), and (77) were obtained in the so-called naive quark–parton model in which interactions between quarks were neglected. If we take into account the interaction of quarks with gluons in this case the expressions for the cross sections have the same form, but the quark and antiquark distribution functions q and $\bar{q}$ will depend not only on the variable x but also on the $\ln Q^2$.

Expressions (74) and (77) allow one to describe all existing experimental data. From these expressions it is possible to obtain information on the distribution of quarks and antiquarks in the nucleon.

For y-distributions from Eqs. (74) and (77) we have

$$\begin{aligned} \frac{d\sigma_{\nu N}}{dy} &= \sigma_0[Q + (1 - y)^2 \bar{Q}] \\ \frac{d\sigma_{\bar{\nu}N}}{dy} &= \sigma_0[(1 - y)^2 Q + \bar{Q}], \end{aligned} \tag{78}$$

where

$$Q = \int_0^1 xq(x)\,dx, \quad \bar{Q} = \int_0^1 x\bar{q}(x)\,dx \tag{79}$$

are the fractions of the momentum of nucleon carried by quarks and antiquarks, respectively.

From the relations in Eq. (78) it follows that at $y = 0$ the cross sections of the processes in Eqs. (66) and (67) must be equal. This is confirmed by the data of the

neutrino experiments. From the data of the CDHS experiment in CERN with neutrino energies in the range $30 < E < 200$ GeV it was found

$$\left(\frac{d\sigma_{\bar{\nu}N}}{dy}\right)_{y=0} \Big/ \left(\frac{d\sigma_{\nu N}}{dy}\right)_{y=0} = 1.01 \pm 0.07. \tag{80}$$

If the contribution of antiquarks into the cross sections is much less than the contribution of quarks, in this case we must expect weak dependence of the cross section $d\sigma_{\nu N}/dy$ on the y and practically $(1-y)^2$-dependence of the cross section $d\sigma_{\bar{\nu}N}/dy$. This behavior corresponds to the experimental data. From the analysis of the CDHS data it follows

$$\frac{\bar{Q}}{Q+\bar{Q}} = 0.15 \pm 0.01. \tag{81}$$

Thus, the contribution of antiquarks to the nucleon momentum is about 15% of the total contribution of the quarks and antiquarks.

For the fraction of nucleon momentum that is carried by the quarks and antiquarks the following value was found:

$$Q + \bar{Q} = 0.492 \pm 0.006 \pm 0.019. \tag{82}$$

Thus, neutrino experiments proved that not all nucleon momentum is carried by the quarks and antiquarks. The other part of the nucleon momentum is carried by the gluons, vector particles that interact with quarks.

Finally, from the quark–parton model it follows that the total neutrino and antineutrino cross sections must depend linearly on neutrino energy E.

$$\begin{aligned} \sigma_{\nu N} &= \frac{G^2}{\pi} M \left(Q + \frac{1}{3}\bar{Q}\right) E \\ \sigma_{\bar{\nu}N} &= \frac{G^2}{\pi} M \left(\frac{1}{3}Q + \bar{Q}\right) E. \end{aligned} \tag{83}$$

The data perfectly confirm this prediction of the theory:

$$\begin{aligned} \sigma_{\nu N} &= (0.686 \pm 0.019) \times 10^{-38} \frac{E}{\text{GeV}} \text{ cm}^2 \\ \sigma_{\bar{\nu}N} &= (0.339 \pm 0.010) \times 10^{-38} \frac{E}{\text{GeV}} \text{ cm}^2. \end{aligned} \tag{84}$$

Thus, the investigation of the neutrino processes at high energies allowed one to resolve the elementary constituents inside of the nucleon and obtained the important information on the distribution of quarks and antiquarks in the nucleon.

XI. NEUTRINO MASSES. INTRODUCTION

The interaction of neutrinos with a matter is given by the Standard Model and is well known. On the other hand, the neutrino masses, neutrino magnetic moments, and other fundamental neutrino properties are basically unknown. We come now to the problem of *the neutrino masses and neutrino mixing*.

A brief history of neutrino masses is the following. Pauli introduced the neutrino as a particle with a mass (as a constituent of nuclei). He thought that the mass of neutrino was less than the electron mass. Fermi and Perren proposed the first method of measuring the neutrino mass. This method was based on the precise measurement of the shape of the high energy part of the β-decay spectrum. This part of the spectrum corresponds to the emission of the neutrino with small energy, and the effects of the neutrino mass in that part of the spectrum are the most pronounced. In experiments on the determination of the neutrino mass by the Fermi–Perren method the spectrum of electrons from the decay of tritium

$$^3\text{H} \to {}^3\text{He} + e^- + \bar{\nu}_e \tag{85}$$

is usually measured.

In the first experiments that were done in the 1940s no effects of neutrino mass were seen. From these experiments it was found that the upper bound of the neutrino mass is much less than the electron mass:

$$m_\nu < 500 \text{ eV}. \tag{86}$$

With the improvement of experimental technique this upper bound became smaller and smaller, and at the time, when the parity violation in β-decay was discovered, the upper bound of the neutrino mass was about 100 eV.

The theory of the two-component neutrino was based on the *assumption* that the neutrino mass is equal to 0. After the success of this theory for many years there was a general belief that the neutrinos are massless particles. The Standard Model of Glashow–Weinberg–Salam was also based on this assumption.

In 1957–1958 B. Pontecorvo considered the possibility of *small but nonzero neutrino masses*. The only known massless particle is the photon. There is a symmetry reason for the photon to be massless—the gauge invariance of the quantum electrodynamics. B. Pontecorvo asserted that there was no such principle in the case of neutrino. He demonstrated that, if the neutrino masses are not equal to 0, and the neutrino states are superpositions of the states of the massive neutrinos, in this case in the beams of the neutrinos *in vacuum*, it will be observed that the neutrino oscillations would be similar to the well-known $K^0 \to \bar{K}^0$ oscillations. B. Pontecorvo showed that the search for neutrino oscillations is a very sensitive method to search for the effects of very small neutrino masses.

In 1962, at the time of the Brookhaven experiment, Maki, Nakagawa, and Sakata discussed some model in which the nucleon was considered as a bound state of some vector particle and massive neutrinos. They assumed that

the fields of ν_e and ν_μ are linear orthogonal combinations of the fields of the massive neutrinos and pointed out that in such a case transition of the muon neutrinos into the electron neutrinos becomes possible.

In the 1970s in Dubna (Russia) and other places in the framework of the Standard Model, the neutrino masses and mixing were considered as phenomena analogous to the Cabibbo–GIM quark mixing. The neutrino oscillations between two types of neutrinos were discussed, and the different experiments on the search for the neutrino oscillations were proposed.

At that time majority of physicists still believed in the massless neutrinos. The opinion about the neutrino masses drastically changed at the end of the 1970s with the appearance and development of models beyond the Standard Model such as models of the Grand Unification. These models are based on the large symmetry groups and fields of neutrinos enter into the same multiplets of the groups as the fields of leptons and quarks. A mechanism of the generation of the masses of quarks and leptons generally also provides masses to the neutrinos. The neutrino masses and mixing started to be considered as phenomena connected with the physics beyond the Standard Model.

In the 1980s special experiments were employed on the search for neutrino oscillation. The problem of the neutrino masses and neutrino oscillations became the most important problem of neutrino physics.

XII. DISCOVERY OF THE τ-LEPTON, *b*- AND *t*-QUARKS. THE NUMBER OF FLAVOR NEUTRINOS

Up to now we have considered four leptons: the two charged leptons, e and μ; and the two neutrinos, ν_e, ν_μ. In 1975, the third heavy charged lepton τ with a mass of about 1.8 GeV was discovered by M. Perl *et al.*, at the $e^+ - e^-$ collider, in Stanford, California.

In the framework of the Standard Model this was a discovery of the third family of leptons and quarks. It meant that a new type of neutrino ν_τ and two new quarks with charges $2/3$ and $-1/3$ must exist. These quarks were called the top and bottom. The real triumph of the Standard Model was the discovery of the bottom particles in the 1980s and top particles in the 1990s.

After these discoveries the charged current of leptons and quarks took the form

$$j_\alpha^{CC} = 2\left(\sum_{l=e,\mu,\tau} \bar{\nu}_{lL}\gamma_\alpha l_L + \bar{u}_L\gamma_\alpha d'_L + \bar{c}_L\gamma_\alpha s'_L + \bar{t}_L\gamma_\alpha b'_L\right), \tag{87}$$

where

$$d'_L = \sum_{q=d,s,b} V_{uq}q_L, \qquad s'_L = \sum_{q=d,s,b} V_{cq}q_L, \qquad b'_L = \sum_{q=d,s,b} V_{tq}q_L. \tag{88}$$

Here V is the unitary matrix that is called the Cabibbo–Kobayashi–Mascawa matrix. The modulas of elements of this matrix are well known from the data of the numerous experiments.

How many families of quarks and leptons exist in nature? As we have seen, the number of families is equal to the number of the neutrino types. The number of the neutrino types was measured in experiments at the $e^+ - e^-$ colliders at SLC (Stanford) and LEP (CERN). In these experiments the probability (width) of the decay

$$Z \to \nu_l + \bar{\nu}_l \qquad l = e, \mu, \tau, \ldots \tag{89}$$

was measured. The width of the decay in Eq. (89) is proportional to the number of the neutrino types n_ν. From the data of the recent LEP experiments it follows

$$n_\nu = 2.994 \pm 0.012. \tag{90}$$

Thus, only three types of neutrinos and three of families of quarks and leptons exist in nature.

XIII. NEUTRINO MIXING

If the neutrinos are massless, the weak interaction conserves three lepton numbers L_e, L_μ, and L_τ:

$$\sum L_e = \text{const}, \quad \sum L_\mu = \text{const}, \quad \sum L_\tau = \text{const}. \tag{91}$$

The values of the lepton numbers of the charged leptons and neutrinos are given in the Table II.

We will assume now that the neutrinos are massive and the lepton numbers are violated by *a neutrino mass term*. In this case fields of neutrinos ν_e, ν_μ, and ν_τ will be linear combinations of the fields of neutrinos with definite masses

$$\nu_{lL} = \sum_{i=1,2,3} U_{li}\,\nu_{iL} \qquad (l = e, \mu, \tau). \tag{92}$$

TABLE II Lepton Numbers of Neutrinos and Charged Leptons

	ν_e, e^-	ν_μ, μ^-	ν_τ, τ^-	*Hadrons*, $\gamma, \ldots$
L_e	1	0	0	0
L_μ	0	1	0	0
L_τ	0	0	1	0

Here U is the unitary matrix ($UU^+ = 1$) and ν_i are the fields of neutrinos with mass m_i.

The relation in Eq. (92) is called *the neutrino mixing* relation. Before we discuss neutrino mixing, let us note that there are two types of particles with the spin 1/2: Dirac particles and Majorana particles.

Dirac particles possess some conserved charges. Every Dirac particle has an antiparticle, the particle with the same mass and opposite charges. The electron, proton, and neutron are the Dirac particles.

Other types of particles with the spin 1/2 are *Majorana particles*. All charges of the Majorana particles are equal to 0 and the Majorana particles and Majorana antiparticles are identical. Up to now the Majorana particles were not observed. The massive neutrinos and neutralinos, particles predicted by models of supersymmetry, are possible candidates. Neutral bosons such as the photon, π^0, and others are well-known neutral particles that are identical to their antiparticles.

There are two possibilities of the violation of the law of conservation of the lepton number.

1. The lepton numbers L_e, L_μ, and L_τ are violated separately, but the total lepton number $L = L_e + L_\mu + L_\tau$ is conserved.

$$\sum L = \text{ const.} \tag{93}$$

In this case the neutrinos ν_i are the Dirac particles that possess the lepton number $L = 1$. The lepton number of the antineutrinos $\bar{\nu}_i$ is equal to -1. The Dirac neutrino masses and neutrino mixing can be generated in the framework of the Standard Model by the same mechanism that is responsible for the generation of the masses and mixing of quarks.

2. There are no conserved lepton numbers. In this case the massive neutrinos ν_i are the Majorana particles. The Majorana neutrino masses and mixing can be generated only in the framework of the models beyond the Standard Model.

If massive neutrinos are Majorana particles there exists a very plausible mechanism of the generation of neutrino masses that connects the smallness of neutrino masses with the violation of lepton numbers at a mass scale M that is much larger than the masses of leptons and quarks. This is the so-called see-saw mechanism. The masses of neutrinos are given in the see-saw case by the relation

$$m_i \simeq \frac{\left(m_{\rm f}^i\right)^2}{M} \ll m_{\rm f}^i \qquad (i = 1, 2, 3), \tag{94}$$

where $m_{\rm f}^i$ is the mass of the lepton or quark in i family ($i = 1, 2, 3$). Let us note that in the see-saw case the neutrino masses satisfy the hierarchy relation

$$m_1 \ll m_2 \ll m_3 \tag{95}$$

that follows from the hierarchy of masses of the leptons (quarks) of the different families.

XIV. NEUTRINO OSCILLATIONS

If there is neutrino mixing,

$$\nu_{lL} = \sum_{i=1}^{3} U_{li}\, \nu_{iL}, \tag{96}$$

where ν_i is the field of neutrino (Dirac or Majorana) with the mass m_i, for the vector of the state of the flavor neutrino ν_l with the momentum $\vec{p}$ we have

$$|\nu_l\rangle = \sum_{i=1}^{3} U_{li}^* |i\rangle, \tag{97}$$

where $|i\rangle$ is the vector of the state of neutrino with the mass m_i and energy $E_i = \sqrt{m_i^2 + \vec{p}^2} \simeq p + \frac{m_i^2}{2p}$, $(m_i^2 \ll p^2)$. Thus, in the case of the neutrino mixing the state of the flavor neutrino is *the superposition of the states of neutrinos with different masses*.

The relation in Eq. (97) is based on the assumption that the mass differences of neutrinos are so small that they cannot be revealed in the processes of neutrino interaction (neutrino production and detection). The neutrino mass differences can be revealed in the neutrino oscillation experiments, special experiments with a large *macroscopic distance* between the neutrino source and the neutrino detector.

Let us assume that at $t = 0$ the neutrino ν_l was produced. At the moment t for the neutrino state we have

$$|\nu_l\rangle_t = \sum_{i=1}^{3} U_{li}^*\, e^{-iE_i t}\, |i\rangle. \tag{98}$$

The state $|\nu_l >_t$ is the superposition of the states of all neutrinos ν_e, ν_μ, and ν_τ

$$|\nu_l\rangle_t = \sum_{l'=e,\mu,\tau} |\nu_{l'}\rangle \mathcal{A}(\nu_l \to \nu_{l'}), \tag{99}$$

where

$$\mathcal{A}(\nu_l \to \nu_{l'}) = \sum_{i=1}^{3} U_{l'i} e^{-iE_i t} U_{li}^* \tag{100}$$

is the amplitude of the transition $\nu_l \to \nu_{l'}$ for the time t. For the transition probability we have

$$P_{\nu_l \to \nu_{l'}} = \left| \delta_{l'l} + \sum_i U_{l'i} \left(e^{-i\Delta m_{i1}^2 \frac{L}{2p}} - 1 \right) U_{li}^* \right|^2. \tag{101}$$

Here $L \simeq t$ is the distance between the neutrino source and the detector, and $\Delta m^2_{i1} = m^2_i - m^2_1$ (we have assumed that $m_1 < m_2 < m_3$).

Thus, the transition probabilities depend on the ratio L/p. If all the neutrino mass-squared differences are so small, that

$$\Delta m^2_{i1} \frac{L}{2p} \ll 1, \tag{102}$$

in this case there will be no transitions between different neutrinos: $P(\nu_l \to \nu_{l'}) = \delta_{l'l}$.

In the simplest case of the transitions between two types of neutrinos the mixing matrix has the form

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{103}$$

where the θ is the mixing angle (if $\theta = 0$ there is no mixing). For the transition probability we have in this case

$$\mathrm{P}(\nu_l \to \nu_{l'}) = \mathrm{P}(\nu_{l'} \to \nu_l) = \frac{1}{2}\sin^2 2\theta\left(1 - \cos\frac{\Delta m^2 L}{2p}\right), \tag{104}$$

where $l' \neq l$ and l, l' take values (μ, τ) or (μ, e) or (e, τ) and $\Delta m^2 = m^2_2 - m^2_1$. For the survival probability we have

$$\mathrm{P}(\nu_l \to \nu_l) = \mathrm{P}(\nu_{l'} \to \nu_{l'}) = 1 - \frac{1}{2}\sin^2 2\theta\left(1 - \cos\frac{\Delta m^2 L}{2p}\right). \tag{105}$$

The expression in Eqs. (104) and (105) can be rewritten in the form

$$\mathrm{P}(\nu_l \to \nu_{l'}) = \frac{1}{2}\sin^2 2\theta\left(1 - \cos 2.53\Delta m^2 \frac{L}{E}\right) \tag{106}$$

$$\mathrm{P}(\nu_l \to \nu_l) = 1 - \frac{1}{2}\sin^2 2\theta\left(1 - \cos 2.53\Delta m^2 \frac{L}{E}\right), \tag{107}$$

where L is the distance in m; E is the neutrino energy in MeV; and Δm^2 is the neutrino mass-squared difference in eV^2. Thus, the transition probability is the periodical function of the parameter L/E.

Let us consider the $\nu_\mu \to \nu_\tau$ transitions and assume that $\sin^2 2\theta = 1$. The $\nu_\mu \to \nu_\mu$ survival probability is equal to 1 at the points $(\frac{L}{E})_1 = \frac{\pi}{2.53\Delta m^2} 2n$ $(n = 0, 1, 2\ldots)$, and we will find at these points only ν_μ. At the values $(\frac{L}{E})_2 = \frac{\pi}{2.53\Delta m^2}(2n+1)$ the survival probability is equal to 0, and only the ν_τ will be found at these points. At all other values of L/E we can find the ν_μ and ν_τ. It is obvious that the sum of the probabilities to find ν_μ and ν_τ is equal to 1.

The phenomena we described is called *neutrino oscillations*. In order to observe the neutrino oscillations it is necessary that the mixing angle be large and the parameter Δm^2 satisfies the following condition:

$$\Delta m^2 \geq \frac{E}{L}. \tag{108}$$

The sensitivities to the parameter Δm^2 of neutrino experiments at different facilities are quite different and cover a very broad range of values of Δm^2. The experiments with the accelerator neutrinos have sensitivity to the parameter Δm^2 in the range 10–10^{-3} eV^2. The experiments with the reactor neutrinos in the range 10^{-2}–10^{-3} eV^2, the experiments with the atmospheric neutrinos in the range 10^{-1}–10^{-4} eV^2, and finally experiments with the solar neutrinos have sensitivity to the parameter Δm^2 in the range 10^{-10}–10^{-11} eV^2.

It is convenient to introduce the neutrino oscillation length

$$L_0 = 4\pi\frac{E}{\Delta m^2}. \tag{109}$$

The transition probability takes the form

$$\mathrm{P}(\nu_l \to \nu_{l'}) = \frac{1}{2}\sin^2 2\theta\left(1 - \cos 2\pi\frac{L}{L_0}\right) \quad (l \neq l'). \tag{110}$$

The expression for the oscillation length can be written in the form

$$L_0 = 2.48\frac{E(\mathrm{MeV})}{\Delta m^2(\mathrm{eV}^2)}\,\mathrm{m}. \tag{111}$$

Neutrino oscillations can be observed if the oscillation length is smaller or of the order of the distance between the neutrino source and the neutrino detector.

In order to compare the neutrino oscillation theory with an experiment, it is necessary to average the corresponding theoretical expressions for transition probabilities over the neutrino energy spectrum, the region where neutrinos were produced and so on. As a result of such averaging, the cosine term in Eq. (107) usually gives 0 contribution to the averaged probability.

XV. EXPERIMENTS ON THE SEARCH FOR NEUTRINO OSCILLATIONS

There are at present, data from numerous experiments on the search for neutrino oscillations. The important indications in favor of the neutrino masses and mixing were found in all the solar neutrino experiments. The compelling evidence in favor of the neutrino oscillations was obtained by the Super-Kamiokande atmospheric neutrino experiment. Some indications in favor of $\nu_\mu \to \nu_e$ oscillations were also found in the Los Alamos accelerator neutrino experiment. In many experiments with the accelerator and reactor neutrinos, no indications in favor of

neutrino oscillations were found. We will discuss first the solar neutrino experiments.

A. The Solar Neutrino Experiments

The energy of the sun is generated in the reactions of the thermonuclear *pp* and *CNO* cycles. From the thermodynamic point of view the energy of the sun is produced in the transition of four protons and two electrons into the ^{4}He and two neutrinos

$$4p + 2e^- \rightarrow {}^4\mathrm{He} + 2\nu_e + 26.7\ \mathrm{MeV}, \qquad (112)$$

Thus, the generation of the energy of the sun is *accompanied by the emission of the electron neutrinos*.

The main sources of the solar neutrinos are the reactions that are listed in Table III. In this table the maximal neutrino energies and neutrino fluxes, predicted by the Standard Solar Model (SSM), are also given.

As seen from Table III, the flux of solar neutrinos constitute mainly the low energy *pp* neutrinos. According to SSM the flux of the medium energy monochromatic ^7Be neutrinos is about 10% of the total flux. The flux of the high energy ^8B neutrinos is only about 10^{-2}% of the total flux. The ^8B neutrinos give, however, main contribution to the event rates of experiments with high energy thresholds.

The results of the five underground solar neutrino experiments are available at present. In the pioneering radiochemical experiment by R. Davis *et al*. (Homestake Mine, South Dakota, USA), a tank filled with 615 tons of C_2Cl_4 liquid was used as the target. The solar neutrinos are detected in this experiment by the radiochemical method through the observation of the reaction

$$\nu_e + {}^{37}\mathrm{Cl} \rightarrow e^- + {}^{37}\mathrm{Ar}. \qquad (113)$$

The radioactive atoms of ^{37}Ar are extracted from the tank by purging it with ^{4}He gas. The atoms of ^{37}Ar are placed in a low background proportional counter in which the electron capture process

$$e^- + {}^{37}\mathrm{Ar} \rightarrow \nu_e + {}^{37}\mathrm{Cl} \qquad (114)$$

is observed by the detection of the Auger electrons (the electrons of conversion).

TABLE III Main Sources of Solar ν_e' s

Reaction	Maximal energy (MeV)	Standard Solar Model flux ($\mathrm{cm^{-2}s^{-1}}$)
$pp \rightarrow d e^+ \nu_e$	$\leq$0.42	6.0×10^{10}
$e^- {}^7\mathrm{Be} \rightarrow \nu_e {}^7\mathrm{Li}$	0.86	4.9×10^9
${}^8\mathrm{B} \rightarrow {}^8\mathrm{Be}\, e^+ \nu_e$	$\leq$15	5.0×10^6

After 2 months of exposition about 16 atoms of the ^{37}Ar are extracted from the volume that contains 2.2×10^{30} atoms of the ^{37}Cl!

The solar neutrinos have been observed in the Davis experiment for about 30 years. For the observed event rate, Q_{Cl} averaged over 108 runs, and the following value was obtained

$$Q_{\mathrm{Cl}} = 2.56 \pm 0.16 \pm 0.16\ \mathrm{SNU}, \qquad (115)$$

where 1 SNU $= 10^{-36}$ events/atom s. The observed event rate is about three times lower than the predicted rate

$$(Q_{Cl})_{SSM} = 7.7 \pm 1.2\ \mathrm{SNU}. \qquad (116)$$

The minimal neutrino energy at which the process in Eq. (113) is possible (the threshold of the process) is equal to $E_{th} = 0.81$ MeV. Thus, the low energy *pp* neutrinos are not detected in the Davis experiment. The most important contribution to the event rate comes from the high energy ^8B neutrinos.

In the radiochemical GALLEX (Italy) and SAGE (Russia) experiments, the solar neutrinos are detected through the observation of the reaction

$$\nu_e + {}^{71}\mathrm{Ga} \rightarrow e^- + {}^{71}\mathrm{Ge}. \qquad (117)$$

In the GALLEX experiment the target is a tank with 30.3 tons of the ^{71}Ga in the gallium–chloride solution. In the SAGE experiment a metallic ^{71}Ga target is used (57 tons of ^{71}Ga).

The threshold of the process in Eq. (117) is $E_{th} =$ 0.23 MeV. Thus, the neutrinos from all the solar neutrino reactions are detected in these experiments (according to the SSM the contributions of the *pp*, ^7Be, and ^8B neutrinos to the event rate in the gallium experiments are 54, 27, and 10%, respectively). The event rates obtained in the GALLEX and SAGE experiments are equal:

$$\begin{aligned} Q_{\mathrm{Ga}} &= 77.5 \pm 6.2^{+4.3}_{-4.7}\ \mathrm{SNU\ (GALLEX)} \\ Q_{\mathrm{Ga}} &= 66.6 \pm {}^{+6.8+3.8}_{-7.1-4.0}\ \mathrm{SNU\ (SAGE)}. \end{aligned} \qquad (118)$$

The predicted rate

$$(Q_{\mathrm{Ga}})_{\mathrm{SSM}} = 129 \pm 8\ \mathrm{SNU} \qquad (119)$$

is about two times higher than the observed rates.

In the underground Kamiokande and the Super-Kamiokande experiments (Japan) the solar neutrinos are detected through the direct observation of the process

$$\nu + e \rightarrow \nu + e. \qquad (120)$$

In the Super-Kamiokande experiment the large 50-kton water-Cerenkov-detector is used. The inner surface of the detector is covered with 11,146 large photomultipliers in which the Cerenkov light from the recoil electrons is detected. About 14 neutrino events per day are observed

in the Super-Kamiokande experiment (in the previous Kamiokande experiment only one neutrino event per day was detected). If the energy of the electron is much higher than its mass, the direction of the momentum of the electron is practically the same as the direction of the momentum of the neutrino. Thus, the measurement of the direction of the momentum of the electrons allows one to choose only events induced by the neutrinos coming from the sun. The recoil electron energy thresholds in the Kamiokande and in the Super-Kamiokande experiments are rather high: 7 and 5.5 MeV, respectively. Thus, only the ^8B neutrinos are detected in these experiments. The following solar neutrino fluxes were obtained from the results of the Kamiokande and the Super-Kamiokande experiments, respectively,

$$\begin{aligned}\Phi &= (2.80 \pm 0.19 \pm 0.33)10^6 \text{ cm}^{-2}\,\text{s}^{-1} \\ \Phi &= \left(2.44 \pm 0.05^{+0.09}_{-0.07}\right)10^6 \text{ cm}^{-2}\,\text{s}^{-1}.\end{aligned} \tag{121}$$

The measured fluxes are about $1/2$ of the predicted one

$$\Phi_{\text{SSM}} = \left(5.15^{+1.00}_{-0.72}\right)10^6 \text{ cm}^{-2}\,\text{s}^{-1}. \tag{122}$$

Thus, from the results of all solar neutrino experiments it follows that in different ranges of energies the flux of the solar ν_e on the earth is significantly smaller than the predicted flux. This deficit constitutes *the solar neutrino problem*.

Neutrino oscillation is the most plausible explanation of the solar neutrino problem. If neutrinos are massive and mixed, the solar ν_e's on the way to the earth can transfer into other neutrinos (ν_μ and/or ν_τ). A few megavolts $\nu_\mu(\nu_\tau)$ cannot produce $\mu(\tau)$ in the CC reaction. Thus, ν_μ and ν_τ cannot be observed in the chlorine and gallium experiments. The muon and/or tau neutrinos give some contribution to the event rate of the Kamiokande and the Super-Kamiokande experiments. However, cross section of $\nu_\mu(\nu_\tau) - e^-$ scattering is about $1/6$ of the cross section of $\nu_e - e-$ scattering, and, therefore, the main contribution to the event rate of these experiments also comes from ν_e's. Thus, if there are neutrino oscillations detected in the solar neutrino experiments flux of the solar neutrinos will be less than expected.

The solar neutrinos produced in the central zone of the sun on their way to the earth pass through a large amount of matter of the sun. At some values of the mixing parameters, effects of the coherent interactions of neutrino with the matter can significantly enhance the probability of the transition of solar ν_e's into other states.

The refraction index of the neutrinos in the matter depends on the amplitude of the elastic scattering of neutrinos in the forward direction. Both CC and NC interactions contribute to the amplitude of elastic $\nu_e - e$ scattering. The amplitude of the elastic $\nu_\mu(\nu_\tau) - e$ scattering is determined only by the NC interaction. Thus, the refraction indexes of the ν_e and $\nu_\mu(\nu_\tau)$ are different. Hence, when neutrino wave propagates through the matter, the flavor content of the neutrino state is changes. Under the condition

$$\Delta m^2 \cos 2\theta = 2\sqrt{2}\text{G}_\text{F}\rho_\text{e}\text{E}, \tag{123}$$

where ρ_e is the electron number-density, the combined effect of neutrino masses and mixing and coherent neutrino interaction in matter can enhance significantly the probability of the transition of ν_e's into other states. This is the so-called Mikheev–Smirnov–Wolfenstein (MSW) effect. This effect can take place in the sun if $10^{-4} < \Delta m^2 < 10^{-7}$ eV^2.

All the existing solar neutrino data can be described, if we assume that there is the mixing of two neutrinos and the values of the solar neutrino fluxes are given by the SSM. In such a case there are only two free parameters: Δm^2 and $\sin^2 2\theta$. From the fit of the all-existing solar neutrino data the following best-fit values of the parameters were obtained:

$$\begin{aligned}\Delta m^2 &= 4.8 \cdot 10^{-5} \text{ eV}^2 \quad \tan^2\theta = 0.35 \\ \Delta m^2 &= 8.0 \cdot 10^{-6} \text{ eV}^2 \quad \tan^2\theta = 6.0 \cdot 10^{-4} \\ \Delta m^2 &= 1.0 \cdot 10^{-9} \text{ eV}^2 \quad \tan^2\theta = 0.70.\end{aligned} \tag{124}$$

We have presented here the best-fit values of the parameters. In reality there are three regions of the allowed values of the parameters Δm^2 and $\tan^2\theta$. In the first two regions matter effects are important. In the large part of the third region only vacuum oscillations takes place.

From the model-independent analysis of all existing data, made under the assumption of the absence of neutrino oscillations, it follows that the flux of ^8Be neutrinos is strongly suppressed. This general conclusion that can be drawn from the existing solar neutrino data will be checked in the BOREXINO experiment that was planned to begin in 2001 in the underground Laboratory Gran Sasso (Italy). In this experiment, mainly the medium energy ^8Be neutrinos will be detected through the observation of the $\nu - e$ scattering.

In another experiment SNO (the Sudbury, Canada), that started in 1999, the solar ν_e's were detected through observation of the electrons in the CC reaction

$$\nu_e + d \to e^- + p + p. \tag{125}$$

In the SNO experiment the solar neutrinos will also be detected also through the observation of the neutrons from the NC process

$$\nu + d \to \nu + n + p. \tag{126}$$

Not only ν_e's but also ν_μ's and ν_τ's will be detected by this method. The comparison of the NC and the CC data

will allow one to obtain model-independent information on the transitions of the solar ν_e's into other states.

B. The Atmospheric Neutrino Experiments

The most compelling evidence in favor of the neutrino oscillations was obtained in the atmospheric neutrino experiments. The main source of the atmospheric neutrinos is the chain of the decays

$$\pi \to \mu + \nu_\mu, \qquad \mu \to e + \nu_e + \nu_\mu, \tag{127}$$

the pions being produced in the interaction of cosmic rays with nuclei in the earth's atmosphere. At relatively low energies ($\leq$1 GeV) the ratio of the muon and electron neutrinos is equal to 2. At higher energies this ratio becames larger than 2 (not all muons have enough time to decay in the atmosphere). The ratio can be predicted, however, with accuracy of better than 5%. The absolute fluxes of the electron and muon neutrinos are predicted at present with accuracy of 20–25%. The results of the measurements of the total neutrino fluxes of the electron and muon neutrinos are usually presented in the form of the double ratio R of the ratio of the observed muon and electron events to the ratio of the muon and electron events calculated by the Monte Carlo method under the assumption that there are no neutrino oscillations. In all the latest atmospheric neutrino experiments it was found that the ratio R is significantly lower than 1.

$$\begin{aligned} R &= 0.65 \pm 0.05 \pm 0.08 \quad \text{(Kamiokande)} \\ R &= 0.54 \pm 0.05 \pm 0.11 \quad \text{(IMB)} \\ R &= 0.61 \pm 0.15 \pm 0.05 \quad \text{(Soudan2)} \\ R &= 0.680^{+0.023}_{-0.022} \pm 0.053 \quad \text{(Super}-\text{Kamiokande)}. \end{aligned} \tag{128}$$

The fact that the double ratio R is less than 1 is a model-independent indication in favor of the disappearence of ν_μ (or appearence of ν_e). Compelling evidence in favor of the disappearence of ν_μ was obtained in the Super-Kamiokande experiment. In this experiment the zenith angle dependence of the number of electron and muon events was measured (the zenith angle θ is the angle between the vertical direction and the neutrino momentum). The angle θ is connected with the distance from the neutrino production region to the neutrino detector. Down-going neutrinos ($\cos\theta = 1$) pass the distance about 20 km. The distance that up-going neutrinos travel ($\cos\theta = -1$) is about 13,000 km.

The only possible source of the zenith angle dependence of the numbers of atmospheric neutrino events is the magnetic field of the earth. At energies higher than 1 GeV the effect of the magnetic field of the earth is low and the numbers of down-going and up-going ν_μ (ν_e) (if there are no neutrino oscillations) must be equal.

The Super-Kamiokande collaboration observed the significant up–down asymmetry of the muon events

$$A_\mu = \frac{U - D}{U + D} = -0.311 \pm 0.043 \pm 0.010. \tag{129}$$

Here U is the total number of the up-going muon events and D is the total number of the down-going muon events.

For the up–down asymmetry of the electron events the value compatible with 0 was found

$$A_e = -0.036 \pm 0.067 \pm 0.02. \tag{130}$$

The data obtained by the Super-Kamiokande collaboration can be explained by the $\nu_\mu \to \nu_\tau$ neutrino oscillations. From the analysis of the data for the parameters Δm^2 and $\sin^2 2\theta$ the following best-fit values were obtained

$$\Delta m^2 \simeq 3 \cdot 10^{-3}\ \text{eV}^2, \quad \sin^2 2\theta \simeq 1. \tag{131}$$

The disappearence of the up-going muon neutrinos is due to the fact that these neutrinos travel a longer distance than the down-going muon neutrinos and have more time to transfer into ν_τ. Because of the high energy threshold ν_τ's practically do not interact in the Super-Kamiokande detector. Thus, disappearence of the ν_μ's is observed.

The ν_μ survival probability depends on the ratio L/E and is given by the expression

$$\text{P}(\nu_\mu \to \nu_\mu) = 1 - \frac{1}{2}\sin^2 2\theta \left(1 - \cos 2.54 \Delta m^2 \frac{L}{E}\right). \tag{132}$$

In Fig. 12 the ratio of the observed and predicted muon (electron) events as a function of L/E is presented. The

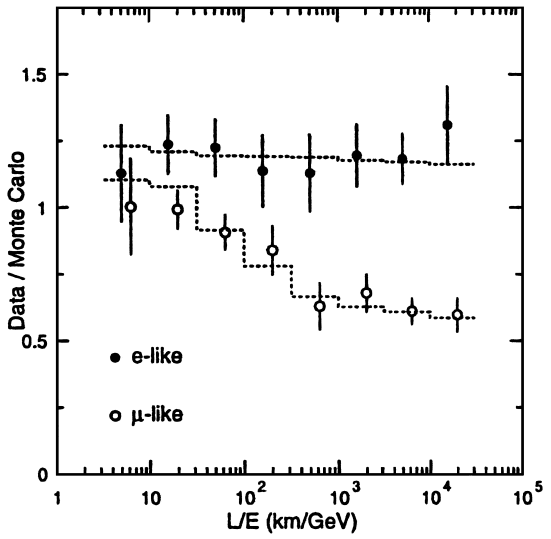


FIGURE 12 The ratio of the number of the Super-Kamiokande events to the number of events, calculated by the Monte Carlo method under the assumption of the absence of oscillations, as a function of L/E. The dashed lines are expected ratios for $\nu_\mu \to \nu_\tau$ oscillations with $\Delta m^2 = 2.2 \cdot 10^{-3}\ \text{eV}^2$ and $\sin 2\theta = 1$.

ratio practically does not depend on L/E for the electron events, but strongly depends on L/E for the muon events. At $L/E \geq 10^3$ km/GeV the argument of the cosine in Eq. (132) is large and the cosine in this equation disappears due to averaging over the neutrino energies and distances. As a result at $L/E \geq 10^3$ km/GeV for the averaged survival probability we have $\bar{P}(\nu_\mu \to \nu_\mu) = 1 - \frac{1}{2}\sin^2 2\theta \simeq \frac{1}{2}$ (see Fig. 12).

The atmospheric neutrino range $\Delta m^2 \simeq 10^{-3}$ eV^2 will be probed in the long-baseline (LBL) accelerator neutrino experiments. The first LBL experiments started in Japan in 1999. The distance between the source (accelerator) and the detector is about 250 km. Two other LBL experiments are under preparation. In the MINOS experiment neutrinos produced from the accelerator at Fermilab (USA) will be detected by the detector in Soudan mine (the distance is about 730 km). In another LBL experiment OPERA neutrinos produced from the accelerator at CERN (Geneva) will be detected by the detector at the underground laboratory, Gran Sasso (Italy) (the distance is also about 730 km). In the accelerator experiments initial neutrinos are mainly ν_μ with a small admixture of ν_e. In the CERN-Gran Sasso OPERA experiment appearence of the ν_τ will be searched for.

C. The LSND Experiment

Some indications in favor of $\nu_\mu \to \nu_e$ oscillations were also obtained in the short-baseline experiment that was done at the Los Alamos linear accelerator (USA). In this experiment a beam of pions produced by 800 MeV protons hits a copper target. In the target the π^+-mesons come to a rest and decay ($\pi^+ \to \mu^+ + \nu_\mu$). The produced muons also come to a rest in the target and decay ($\mu^+ \to e^+ + \bar{\nu}_\mu + \nu_e$). Thus, in decays of the π^+'s and μ^+'s ν_μ's, $\bar{\nu}_\mu$'s and ν_e's are produced (there are no $\bar{\nu}_e$'s from these decays). Let us note that practically all π^-'s are absorbed in the target before decay.

In the neutrino detector LSND at the distance of about 30 m from the target, the electron antineutrinos $\bar{\nu}_e$'s were searched for through the observation of the classical process

$$\bar{\nu}_e + p \to e^+ + n. \tag{133}$$

In the interval of the positron energies $30 < E < 60$ MeV in the LSND experiment it was observed 33.9 ± 8.0 events.

The observed events can be explained by the $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillations. If we take into account the results of the different short-baseline experiments in which neutrino oscillations were not found, in this case from the LSND experiment the following ranges of the oscillation parameters can be found

$$0.2 \leq \Delta m^2 \leq 2\ \text{eV}^2 \qquad 2 \cdot 10^{-3} \leq \sin^2 2\theta \leq 4 \cdot 10^{-2}. \tag{134}$$

The indications in favor of the $\nu_\mu \to \nu_e$ oscillations obtained in the LSND experiment will be checked by the BOONE (Fermilab, USA) experiment.

XVI. THE NEUTRINOLESS DOUBLE β-DECAY

We have discussed in the previous sections neutrino oscillation experiments that allow one to obtain information on very small neutrino mass-squared differences. Important information on the neutrino masses and the nature of massive neutrinos can be obtained from experiments on the investigation of the neutrinoless double β-decay

$$(A, Z) \to (A, Z+2) + e^- + e^-. \tag{135}$$

Here (A, Z) is even–even nucleus. In the experiments the neutrinoless double β-decay of ^{76}Ge, ^{136}Xe, ^{130}Te, ^{100}Mo, and other nuclei are searched for. The process in Eq. (135) is allowed, if the total lepton number L is not conserved, i.e., if massive neutrinos are the Majorana particles.

In the framework of the standard CC weak interaction with the Majorana neutrino mixing the neutrinoless double β-decay is the second order in the Fermi constant G_F process with a virtual neutrino. The matrix element of the process is proportional to the effective Majorana mass

$$\langle m \rangle = \sum_i U_{ei}^2 m_i, \tag{136}$$

where m_i are the neutrino masses.

There are many experiments in which neutrinoless double β-decay are searched for. No positive indications in favor of such decay have been found up to now. The best lower bound of the life-time was obtained in the Heidelberg–Moscow experiment in wich the neutrinoless double β-decay of the ^{76}Ge was searched for:

$$T_{1/2} > 1.6 \times 10^{25}\ \text{years} \tag{137}$$

From this and other results the following upper mass of the effective Majorana mass was found

$$|\langle m \rangle| > (0.5 - 1)\ \text{eV} \tag{138}$$

In future experiments on the search for the neutrinoless double β-decay the sensitivity $|\langle m \rangle| \leq 10^{-1}$ eV will be achieved.

XVII. NEUTRINO MASSES FROM EXPERIMENTS ON THE MEASUREMENT OF THE β-SPECTRUM OF TRITIUM

The first method of measuring neutrino mass was proposed in the classical paper by Fermi on β-decay. The method consists of the precise measurement of the end point part of β-spectrum. This part of the spectrum corresponds to the emission of low energy neutrinos.

Usually, for the determination of neutrino mass the β-spectrum of the decay of the tritium

$$^3\mathrm{H} \rightarrow {}^3\mathrm{He} + \mathrm{e}^- + \bar{\nu}_\mathrm{e} \tag{139}$$

is investigated. The β-spectrum of this decay is determined by the phase-space factor

$$\frac{dN}{dT} = CpE\,(Q-T)\sqrt{(Q-T)^2 - m_\nu^2}F(E). \tag{140}$$

Here p and E are electron momentum and energy; $T = E - m_e$ is the electron kinetic energy; $Q = 18.6$ keV is the energy release; $C = \mathrm{const}$; $F(E)$ is the known function which describes the Coloumb interaction of the final particles; and m_ν is the mass of neutrino. If the neutrino mass is equal to 0, $T_{max} = Q$. For nonzero neutrino mass $T_{max} = Q - m_\nu$. Thus, for nonzero neutrino mass at the end-point part of the electron spectrum, the deficit of the events (with respect to the number of the events expected for $m_\nu = 0$) must be observed.

At the moment no positive indications in favor of nonzero neutrino mass have been obtained from the ^{3}H experiments. For the upper-bound of the neutrino mass it was found that

$$\begin{aligned} m_\nu &\leq 2.5 \text{ eV} \quad \text{(Troitsk)} \\ m_\nu &\leq 2.8 \text{ eV} \quad \text{(Mainz)}. \end{aligned} \tag{141}$$

XVIII. CONCLUSION

The neutrinos play a special role in the fields of particle physics and astrophysics. They have enormous penetration properties and give us the unique possibility to investigate the internal structure of the nucleon, the internal invisible region of the sun where solar energy is produced.

The neutrinos are also exceptional particles with respect to their internal properties. The neutrino masses are many orders of magnitude smaller than the masses of their family partners (electron, muon, tau). Because of the smallness of neutrino masses, new physical phenomenon, *neutrino oscillations*, the periodical transitions between different types of neutrinos *in vacuum* or in matter, become possible. The evidence for this phenomenon, predicted many years ago, was obtained recently. The investigation of the neutrino oscillations that is going on all over the world is a new field of research.

The investigation of the neutrino oscillations, the neutrinoless double β-decay and β-spectrum of ^3H-decay, will allow one to obtain important information on the neutrino masses, neutrino mixture, and neutrino nature (Dirac or Majorana?).

The exceptional smallness of the neutrino masses requires special explanation. The most plausible explanation of this smallness is connected with the physics beyond the Standard Model, namely, with a violation of the lepton number at a mass scale M that is much larger than the mass scale of the violation of electroweak symmetry $\sim 10^2$ GeV, which determines the values of masses of the leptons, quarks, $W^\pm$- and Z^0-bosons. If the explanation is correct in this case, the massive neutrinos must be truly neutral Majorana particles. All other fundamental fermions (leptons, quarks) are charged Dirac particles.

SEE ALSO THE FOLLOWING ARTICLES

DENSE MATTER PHYSICS • PROTON DECAY • NEUTRINO ASTRONOMY • NEUTRON STARS • PARTICLE PHYSICS, ELEMENTARY

BIBLIOGRAPHY

Bahcall, J. N. (1989). "Neutrino Physics and Astrophysics," Cambridge University Press, United Kingdom.

Bilenky, S. M., and Pontecorvo, B. (1978). *Phys. Rep*. **41,** 225, North-Holland.

Bilenky, S. M., Giunti, C., and Grimus, W. (1999). *Progr. Part. Nucl. Phys*. **43,** 1.

Boehm, F., and Vogel, P. (1994). "Physics of Massive Neutrinos," Cambridge University Press, Cambridge, United Kingdom.

Brunner, J. (1997). Fortsh. Phys. **45,** 343.

Feynman, R. P., and Gell-Marr, M. (1958). *Phys. Rev.* **109,** 193, American Institute of physics, NY, USA.

Haxton, W., and Holdstein, B. (2000). "Neutrino physics," *Am. J. Phys.* **68,** 15.

Kim, C. W., and Pevsner, A. (1993). "Neutrinos in Physics and Astrophysics," Contemporary Concepts in Physics, Vol. 8, Harwood Academic Press, Chur, Switzerland.

Mohapatra, R. N., and Pal, P. B. (1998). "Massive Neutrinos in Physics and Astrophysics," World Scientific, Singapore.

Oberauer, L., and von Feilitzsch, F. (1992). *Rep. Prog. Phys*. **55,** 1093.

Pais, A. (1986). "Inward Bound of Matter and Forces in the Physical World," Oxford, Clarendon Press, United Kingdom.

Perkins, D. H. (2000). "Introduction to High Energy Physics," 4th Ed. Cambridge University Press, Cambridge, United Kingdom.

Sudarshan, E. C. G., and Marshak, R. E. (1958). *Phys. Rev.* **109,** 1860, American Institute of Physics, NY, USA.

Winter, K. (ed.) (2000). "Neutrino Physics," 2th Ed., Cambridge University Press, Cambridge, United Kingdom.

Zuber, K. (1998). *Phys. Rep*. **305,** 295, North-Holland.



Nuclear Chemistry

Walter Loveland
Oregon State University

I. Introduction
II. Radioactive Decay Kinetics
III. Fundamental Properties of Nuclei
IV. Nuclear Structure Models
V. Alpha Decay
VI. Beta Decay
VII. Gamma Decay
VIII. Nuclear Reactions
IX. Production of New Elements
X. Recent Developments

GLOSSARY

Collective model Model of the nucleus that emphasizes the collective or cooperative actions of the nucleons, leading to nuclear oscillations and rotations.

Compound nucleus Intermediate state in nuclear reactions in which the projectile and target nucleus have completely amalgamated and shared their energy, and the system has amnesia about its method of formation.

Deep inelastic scattering Type of heavy ion reaction mechanism in which the colliding nuclei exchange large amounts of energy with relatively small changes in overall identity of projectile and target nuclei.

Direct reaction Type of nuclear reaction mechanism in which the incident projectile collides with one, or at most a few, target nucleons, some of which may be ejected.

Internal conversion Type of electromagnetic decay process in which the nuclear electromagnetic field may interact with an orbital electron, transferring energy to it and causing its ejection from the atom.

Nuclear binding energy Energy required to break up a nucleus into its constituent neutrons and protons.

Q value Overall energy liberated or taken up in a nuclear process.

Quantum chromodynamics Theory of the nuclear force in terms of quarks and their interactions.

Radiochemistry Use of chemical techniques to study radioactivity and its properties.

Shell model Model of nuclear structure that emphasizes the independent behavior of single nucleons in the average potential created by the other nucleons.

Spallation High-energy nuclear reaction mechanism, in which small pieces of the target nucleus are chipped away by the interaction of the projectile with the target nucleus.

NUCLEAR CHEMISTRY is the study by chemists of (1) the structure, decay, and reactions of nuclei; (2) microscopic phenomena using nuclear techniques, such as activation analysis or radiotracer methods; and (3) macroscopic phenomena in which nuclear processes play an important role, such as cosmochemistry and geochronology. In addition, many nuclear chemists are employed in the use of nuclear technology, as in the development of nuclear energy, nuclear weapons, nuclear medicine, and so forth.

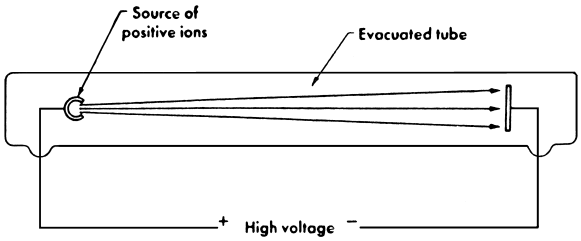


FIGURE 1 Simplest type of particle accelerator. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 88. Copyright 1965 by Prentice-Hall, Inc.]

I. INTRODUCTION

A. Nature of Nuclear Chemistry

Many years ago there was a clear-cut difference between nuclear chemistry and nuclear physics. Nuclear chemists used chemical techniques to study complex, many-body problems with phenomenological and statistical approaches. Nuclear physicists used physical measurement techniques to study simpler, few-body systems with exact, rigorous methods. Today the distinction between the two disciplines, nuclear physics and nuclear chemistry, has become blurred. Both disciplines use physical measurement techniques, and there is considerable overlap in the type of problems studied. It is fashionable to refer to the two disciplines collectively as nuclear science and the practioners as nuclear scientists.

B. Tools of the Nuclear Chemist

Nuclear chemists have a wide variety of sophisticated and expensive tools to help in their studies. To study nuclear reactions and to produce radioactive nuclei for a variety of purposes, some means must be available to produce energetic particles to interact with nuclei. Two types of devices are in general use, the nuclear reactor (which serves as a neutron source) and the particle accelerator. In the latter category are simple electrostatic accelerators, such as the Cockcroft–Walton machines (Fig. 1), which accelerate ions to typical energies of a few megaelectron-volts ($1\ \text{MeV} = 1.602 \times 10^{-13}$ J) or less. These accelerators depend on simple acceleration of a charged particle that passes through an electrostatic potential V. The Van de Graaff accelerator is similar in principle to the Cockcroft–Walton accelerator; however, an ingenious scheme of achieving high voltages by transferring charge to a moving belt, where it is transported to the inside of a hollow conductor and deposited (Fig. 2), makes the applied potential V much larger.

Magnetic resonance accelerators, such as the cyclotron, produce high-energy particles without the use of high voltages. As shown in Figure 3, charged particles are produced in an ion source, accelerated in hollow, charged metal boxes called dees (in reference to their shape), and bent in spiral paths by an applied magnetic field. The crucial point of the cyclotron design is that the voltage applied to the dees oscillates in such a manner that each particle is given a potential energy kick as it passes between the dees. This is possible because as the particle energy becomes larger (higher velocity), the distance traveled by the ion becomes larger, and thus the time for the traversal of the dees remains constant. Finally, as the particle energy exceeds ~10 MeV/A, where A is the particle mass number, relativistic effects such as the increase in projectile

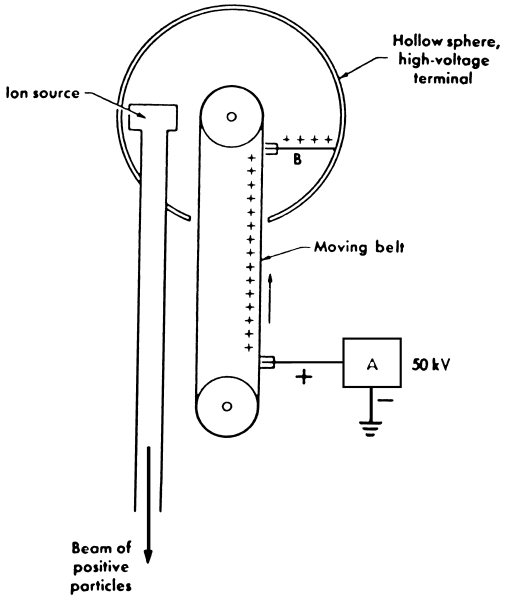


FIGURE 2 Van de Graaff accelerator, schematic. The high-voltage supply A sprays positive charge onto the moving belt. The charge is carried into the high-voltage terminal and is there removed by the conductor B. It then flows to the outside surface of the high-voltage electrode. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 89. Copyright 1965 by Prentice-Hall, Inc.]

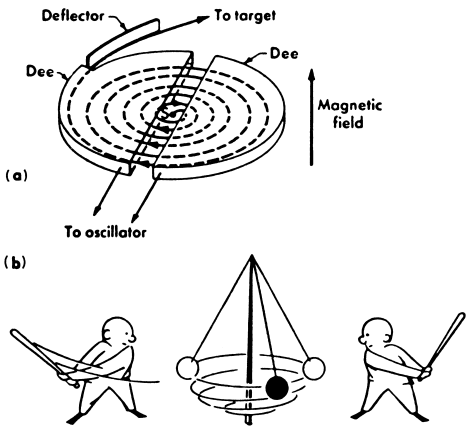


FIGURE 3 (a) Schematic view of the dees of a cyclotron. The ion source is at S. (b) An analogy to the way particles are accelerated in a cyclotron. The ball represents the particles, the string and pole represent the magnetic field that keeps the particles in a spiral path, and the two batters represent the dees and radio-frequency energy that actually cause the acceleration. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 90. Copyright 1965 by Prentice-Hall, Inc.]

mass become important. Because of this increase in mass, the dee traversal time changes, and further acceleration is not possible. A sector-focused cyclotron avoids these difficulties by changing the magnetic field to compensate for the increased particle mass. The frequency modulated (FM) cyclotron, or synchroeyelotron, varies the frequency of the cyclotron to compensate for this increase in particle mass.

The synchrotron, which is widely used to produce high-energy (GeV) ions, works on the principle of accelerating the ions by constraining them to move in a fixed orbit while changing the electric and magnetic fields to accelerate a single bunch of ions at one time (Fig. 4). The linear accelerator accelerates charged particles (usually electrons or heavy ions) by having the particles move in a straight line down an evacuated tube. The particles are given repeated kicks as they pass through a series of hollow electrodes (Fig. 5). Of special interest to nuclear chemists are the heavy-ion linear accelerators at Argonne (Illinois) in the United States and Darmstadt in the Federal Republic of Germany, which have been used extensively in studies of heavy nuclear projectiles and their nuclear reactions. These accelerators are of the Alvarez type, in which the accelerator tank is made to resonate at the desired radio frequency to achieve acceleration. For lower-energy heavy ions ($v/c \simeq 0.005$), a different design—the Wideröe design—is employed; here the rf field is applied directly to the drift tubes.

The most common neutron source is the nuclear reactor. Neutrons are emitted during the fission of the uranium reactor fuel and have energies from zero to a few mega-electron-volts, with typical particle fluxes of

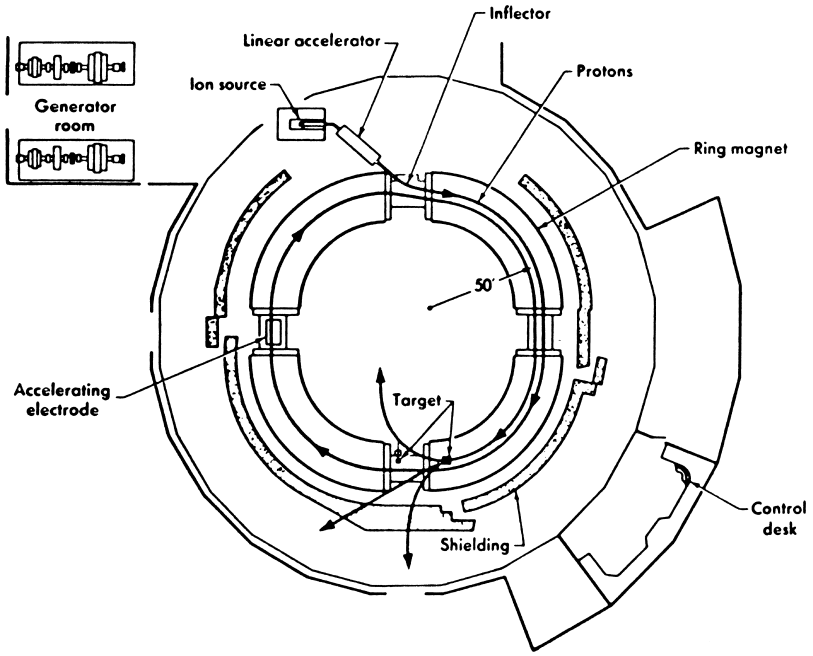


FIGURE 4 Typical proton synchrotron. The ring magnet is divided into four sections. Equipment, such as the target and accelerating electrode, is in the straight parts where there is no magnetic field. The generations supply dc current to the ring magnet. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 93. Copyright 1965 by Prentice-Hall, Inc.]

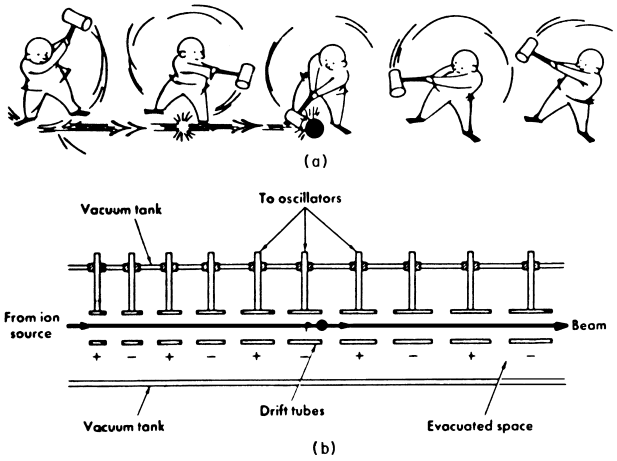


FIGURE 5 (a) Principle of acceleration by repeated impulses, as in a linear accelerator. (b) Principle of the linear accelerator. A particle at P is shown leaving a drift tube and receiving acceleration toward the positively charged drift tube to the right. While the particle is in that tube, the potentials on the tubes all change sign. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 95. Copyright 1965 by Prentice-Hall, Inc.]

10^{12}–10^{14} n/cm^2 sec in the center of the reactor. Samples are irradiated by placing them in or near the reactor core or by building tubes through the reactor shielding to extract neutron beams from the reactor.

In addition to having a means of inducing nuclear reactions and of producing radioactivity, one needs a means of detecting it. Nuclear radiation detectors are designed to use some feature of the interaction of radiation with matter to provide a signal (usually electrical) indicating some feature of the radiation. Gas ionization detectors, such as the Geiger—Müller counter or the proportional counter, operate on the principle of collecting and increasing the number of ions produced in a gas when ionizing radiation strikes it. Semiconductor detectors collect the ionization produced when radiation interacts with Si or Ge semiconductors. One type of detector, the surface barrier detector, is used extensively to detect charged particles and fission fragments. Another type, the intrinsic germanium detector, and its silicon equivalent Si(Li) are used primarily to detect photons. A third class, scintillation detectors, record the passage of radiation through a liquid or solid by means of the tiny flashes of light created by the excitation by radiation of the fluorescent materials in the detector. Depending on the device, one can determine the number of particles striking a detector, their energy, their velocity and/or mass (from time-of-flight techniques) as their atomic numbers (from a measurement of the energy loss dE in traversing a material or fixed thickness dx).

Chemical manipulations of radioactivity are called radiochemistry. Radiochemistry is generally thought of as a pseudoanalytical chemistry applied to radioactive materials. It generally involves the separation of a specific radioactive substance from a mixture. The chemistry of a radioactive species is, in general, the same as that of an inactive species except for effects related to the very small amounts of material that are usually present. In general, special techniques need to be employed in handling such small amounts of material. Chemical investigations frequently have to be performed with remote-handling apparatus and massive amounts of shielding between the radiochemists and their samples. Because of the short half-lives of many radionuclides, speed is frequently of the essence in radiochemistry. In recent times chemical and instrumental techniques are frequently combined to isolate a radioactive substance and measure its radiations. A number of individual elements are separated chemically from a complex mixture, and then instrumental techniques are used to detect the radiation specifically from the nuclide of interest.

C. Introductory Concepts about Nuclei

We now know that the atom consists of a tiny nucleus of radius $\sim 10^{-13}$ cm, in which most of the mass of the atom is concentrated. Moving about the nucleus in certain well-defined regions of space are the electrons, with neutral atoms having as many electrons as there are protons in the nucleus. The radius of the atom is $\sim 10^{-8}$ cm. The electrons are held in their orbitals by the attractive force between the negatively charged electron and positively charged nucleus. The nucleus itself is made of Z protons (where Z is the atomic number) and N neutrons (where N is the neutron number). The total number of particles in the nucleus, which is called the mass number, is designated by $A = (N + Z)$. The nomenclature for a nucleus containing Z protons, N neutrons, and A particles is ${}^{A}_{Z}\mathrm{X}_{N}$, where X is the chemical symbol for an element of atomic number Z. The nuclear force holds the neutrons and protons together in the nucleus; it is quite strong ($10^2 \times$ the strength of the electromagnetic force), as evidenced by the nuclear density of $>$200,000 tonnes/mm^3.

II. RADIOACTIVE DECAY KINETICS

A. Basic Nature of Radioactivity

Many nuclei are radioactive, that is, they spontaneously disintegrate with the emission of radiation. There are three well-known types of nuclear decay: α-, β-, and γ-ray decay. In addition, there are other decay modes that are important for specific classes of nuclei, such as spontaneous

fission for the heavy elements. In α decay, the nucleus decays by spontaneously emitting a doubly charged $^4_2$He nucleus, leading to a product that is two units lower in Z and four units lower in A. The emitted ^{4}He nuclei, called α particles, are monoenergetic and have energies ranging from 1.5 to 12 MeV, with typical energies being 4–6 MeV. The emitted α particles are highly ionizing when they interact with matter and can be stopped by a sheet of paper.

Nuclear β decay can occur in three different ways depending on the N/Z ratio of the decaying nucleus. (1) In β^- emission, a nuclear neutron changes into a proton with the emission of an energetic electron (β^-) and an electron antineutrino ($\bar{\nu}_e$). The A of the product nucleus is the same as that of the original nucleus; the Z increases by one unit. The electron energies range from zero to some maximum energy characteristic of each nucleus. The emitted electrons are not as ionizing as the α-particle but do not penetrate a few cm of metal. (2) In β^+ emission a nuclear proton changes into a neutron with the emission of a position (β^+) and an electron neutrion (ν_e). The corresponding product, called the daughter nucleus, has the same A as the original nucleus, but its N is increased by one, and its Z is decreased by one. The emitted positrons (β^+) have continuous energy distributions, and they annihilate when contacting ordinary matter, releasing two 0.511-MeV photons per annihilation. (3) In electron capture (EC) decay, the nucleus captures an orbital electron and emits an electron neutrino (ν_e). This mode of decay gives rise to a daughter product with the same Z, A as would occur in β^+ decay and competes favorably with β^+ decay in the heavy elements. An experimental signature of this mode of decay is the emission of X-rays following the electronic rearrangements that occur after capture of an orbital electron.

The third major type of nuclear decay process is electromagnetic decay, which occurs in two ways: γ-ray decay and internal conversion. In γ-ray decay the nucleus emits an energetic photon, decreasing its internal excitation energy. No changes in the Z or A of the initial nucleus are observed. A typical time scale for the emission of these photons is 10^{-11} sec. When γ-ray decay is inhibited so that the lifetime of the emitting state is measurable ($\lesssim 10^{-9}$ sec), the transition is referred to as an isomeric transition. In all types of γ-ray decay, the emitted photons are monoenergetic. In internal conversion decay the electromagnetic field of the nucleus interacts with an orbital electron, ejecting it and giving it the excitation energy that would have been emitted as a γ-ray photon. The internal conversion process is a radiationless transition, with the ejected electron being monoenergetic and with no change taking place in the nuclear Z or A. The relative number of decays taking place by internal conversion, compared with γ-ray emission, is referred to as the internal conversion coefficient.

B. Decay, Growth and Decay, Naturally Occurring Radionuclides

Radioactive decay is a first-order reaction, that is, the number of decays per unit time is directly proportional to the number of nuclei present. Thus if $-dN/dt$ is the decay rate, we can say that

$$-dN/dt = \lambda N, \tag{1}$$

where λ, the decay constant, is $\ln 2/t_{1/2}$, where $t_{1/2}$ and N are the half-life and number of radioactive nuclei present, respectively. It is easy to show that the rate of disappearance of radioactive nuclei is governed by the equation

$$N = N_0 e^{-\lambda t}, \tag{2}$$

where N is the number of radioactive nuclei present at time t and N_0 the number of nuclei present at $t = 0$. For the general situation, where species 1 decays to 2, which in turn decays to 3, and so on, one can show that

$$N_n = N_1^0 \lambda_1 \lambda_2 \cdots \lambda_{n-1} \sum_{j=1}^{n} \frac{e^{-\lambda_j t}}{\prod_{k=1,k\neq j}^{n} (\lambda_k - \lambda_j)}, (n > 1), \tag{3}$$

where N_n is the number of nuclei of the nth member of the chain at time t, and $N_2^0 = N_3^0 = \cdots = N_n^0 = 0$.

Certain special cases are important for the behavior of the daughter nucleus in a parent–daughter decay chain. Simple application of Eq. (3) leads to the general equation for the number of daughter nuclei present at time t,

$$N_2 = [\lambda_1/(\lambda_2 - \lambda_1)] N_1^0 (e^{-\lambda_1 t} - e^{-\lambda_2 t}), \tag{4}$$

which for the case where $\lambda_1 < \lambda_2$ (transient equilibrium) simplifies to

$$N_1/N_2 = (\lambda_2 - \lambda_1)/\lambda_1, \tag{5}$$

which in turn simplifies when $\lambda_1 \ll \lambda_2$ (secular equilibrium) and becomes

$$N_1/N_2 = \lambda_2/\lambda_1 \quad \text{or} \quad A_1 = A_2, \tag{6}$$

where $A_i = \lambda_i N_i$ is the radioactivity dN/dt of species i. This latter case applied for the natural radioactive decay chains is shown in Figure 6. The activity of each member of a decay chain is the same.

In addition to the four heavy element, naturally occurring radioactive series, there are several other naturally occurring radioactive nuclei. Prominent among these are ^{40}K, a 1.3×10^9-yr β^- emitter, which is present in all potassium, and ^{14}C, a 5730-yr β^- emitter, which is present

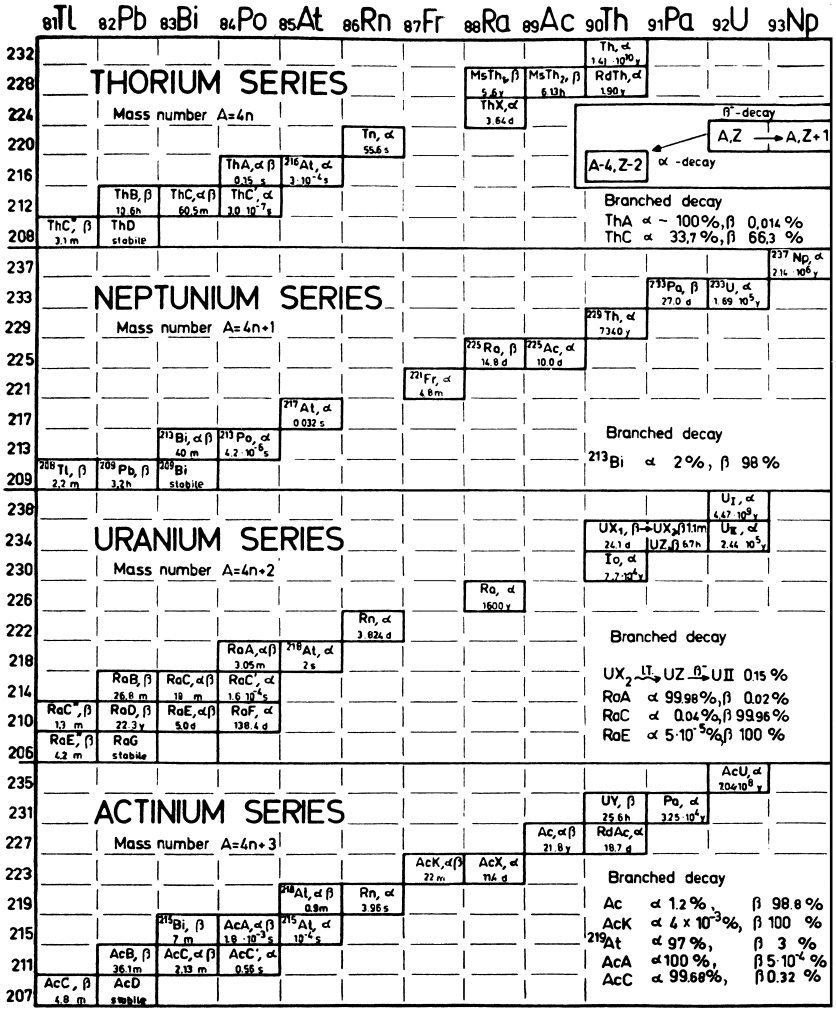


FIGURE 6 The three naturally occurring radioactive decay series and the synthetic neptunium series. Although ^{239}Pu (which is parent to the actinium series) and ^{244}Pu (which is parent to the thorium series) have been discovered in nature, the decay series shown here begins with the most abundant long-lived nuclides.

in the carbon of all living things. The latter radionuclide is being continuously generated in the upper atmosphere by the reactions of cosmic rays with the atmosphere. Another radionuclide that has become an important component of the natural environment is ^{3}H, tritium. The presence of this 12-yr β^- emitter in nature is primarily due to the effects of atmospheric tests of nuclear weapons.

III. FUNDAMENTAL PROPERTIES OF NUCLEI

A. Basic Constituents of Nuclear Matter

We have said that the nucleus is made up of neutrons and protons (collectively termed nucleons) held together by the nuclear or strong interaction force. All strongly interacting particles, such as the neutron and proton, are referred to as hadrons. We now believe that all hadrons are composed of smaller particles known as quarks. There are six kinds of quarks (*u*p, *d*own, *s*trange, *c*harm, *t*op, and *b*ottom), and each quark has a fractional electrical charge. For example, the proton is thought to be composed of two u quarks and one d quark with electrical charges $+\frac{2}{3}e$ and $-\frac{1}{3}e$, respectively. Similarly, the neutron is thought to be composed of one u quark and two d quarks. It is believed also that there is another internal quantum mechanical property of quarks called color, and, that each kind of quark actually represents a family of three particles that are identical in all respects except that they have different colors. It is possible to calculate many of the fundamental properties of nuclei using these ideas about quarks and their interactions (quantum chromodynamics).

It is also possible, without loss of insight, to describe nuclei in terms of groups of quarks (i.e., the neutron and proton) and their interactions.

B. Nuclear Masses and Binding Energies

During radioactive decay, nuclei undergo spontaneous changes that result in the formation of species that are more stable. One way of understanding these processes is in terms of the neutron-to-proton ratio in the nuclei. For each element there is a specific neutron-to-proton ratio that makes for the greatest stability. In the elements of lowest atomic weight, this ratio approximates one neutron to one proton; but as we move up the scale to elements of higher atomic weight, the ratio approaches 1.5 neutrons to one proton for maximum stability. Unstable nuclei that are neutron-rich relative to the stable species of the same A value attempt to reduce that neutron excess by β^- decay, changing nuclear neutrons into protons. Similarly, proton-rich unstable nuclei decay by β^+ or EC.

Another, more quantitative way of explaining why radioactive decay takes place is in terms of the energy changes involved. The mass of a nucleus is very small, being of the order of 10^{-23} g. Instead of using such small numbers, we can define two convenient units to discuss nuclear masses. The first is the atomic mass unit (amu), which equals 1.66053×10^{-24} g. Thus, if the mass of a ^{12}C atom is 1.992×10^{-23} g, the mass expressed in atomic mass units is

$$1.992 \times 10^{-23}\ \text{g}/1.660 \times 10^{-24}\ \text{g/amu},$$

or 12.00 amu. The second convenient unit of mass used is the mega-electron-volt (MeV), which is an energy unit. To express mass in terms of energy we use Einstein's equation

$$E = mc^2, \tag{7}$$

where m is the mass, E the energy, and c the speed of light.

Consider the concept of nuclear binding energy (BE). The binding energy of any nucleus is the energy liberated when a group of nucleons combine to form a nucleus. Thus the binding energy of ^{4}He is given by

$$\text{BE}(^4\text{He}) = 2M_{^1\text{H}} + 2M_\text{n} - M_{^4\text{He}}, \tag{8}$$

where $M_{^1H}$ is the atomic mass of ^{1}H, M_n the neutron mass, and $M_{^4He}$ the atomic mass of ^{4}He. Using values from a table of atomic masses, we see that $\text{BE}(^4\text{He}) = 2(1.00813) + 2(1.00896) - 4.00398 = 0.03030$ amu. Since 1 amu equals 931.5 MeV,

$$\text{BE}(^4\text{He}) = 28.22\ \text{MeV}. \tag{9}$$

A convenient quantity is the binding energy per nucleon or, in other words, the total nuclear binding energy divided by the number of nucleons in the nucleus. Since there are four nucleons in ^{4}He, the binding energy per nucleon is 28.21 MeV/4 or 7.1 MeV. Calculating this quantity for each nucleus in the periodic table and plotting it versus the nuclear mass number A, we arrive at Figure 7. Clearly, the nucleus with the highest binding energy per nucleon is most tightly held together, that is, it is most stable. As the figure shows, the maximum binding occurs near $A = 60$ in the vicinity of Fe and Ni.

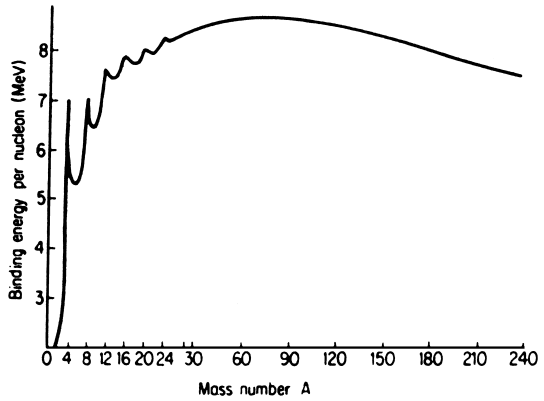


FIGURE 7 Binding energy per nucleon versus the nuclear mass number A.

C. New Elements

A problem of importance to nuclear chemists that is closely related to nuclear masses and binding energies is that of making new chemical elements. As one adds more protons to the nucleus, the Coulomb repulsion between these protons will eventually cause the nucleus to fission spontaneously. One model of the nucleus predicts that a nucleus will fission almost instantaneously when $E_c = 2E_s$, where E_c and E_s are the repulsive and attractive surface energy of the nucleus, respectively. The quantities E_c and E_s are given by

$$E_c = (3/5)(Ze)^2/R = k_c Z^2 \big/ A^{1/3}$$
$$E_s = 4\pi R^2 \gamma = k_s A^{2/3}, \tag{10}$$

where γ is the nuclear surface tension ($\sim$1 MeV/fm^2), Z the atomic number, and R the nuclear radius (proportional to $A^{1/3}$ where A is the nuclear mass number). The limiting value of the atomic number Z_{limit} is then

$$Z^2_{\text{limit}} = 2k_s/k_c A_{\text{limit}}. \tag{11}$$

The neutron/proton ratio in heavy nuclei is $\sim$1.5/1 ($A_{\text{limit}} \sim 2.5\, Z_{\text{limit}}$), thus

$$Z_{\text{limit}} = 5(k_s/k_c). \tag{12}$$

Thus the upper bound to the periodic table is proportional to the ratio of two fundamental constants related to the strength of the nuclear (surface) and electromagnetic forces. The ratio k_s/k_c is about 20–25; and thus, on the basis of this estimate, we might expect 100–125 chemical elements.

IV. NUCLEAR STRUCTURE MODELS

The nuclear force, or strong interaction, is the strongest of the four forces acting in nature (the strong, electromagnetic, weak, and gravitational) and has a short range ($\sim 1.4 \times 10^{-13}$ cm). The attractive portion of the force between two nucleons is thought to be due to a virtual exchange of π mesons between the interacting particles. A component of the nuclear force is due to the interaction of the spin and orbital angular momenta of the interacting nucleons. This component associates a different potential energy with the two configurations shown in Figure 8. While these details and other parts of the nuclear force are known, a complete description is not yet available. This aspect, coupled with the many-body nature of the problem of describing the structures of nuclei, has forced scientists to use models—oversimplifications emphasizing one feature of the phenomenon, that allow calculations and predictions about the future behavior of the system—to describe the structure of nuclei (and other phenomena).

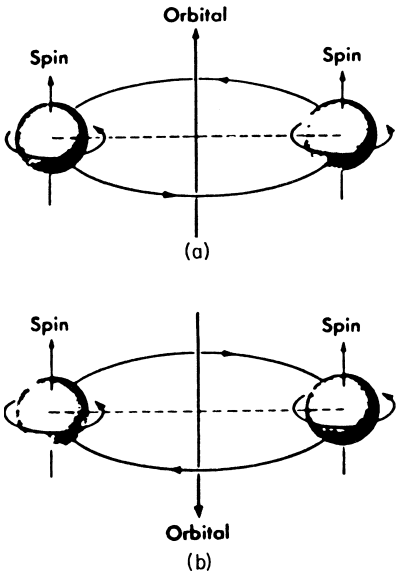


FIGURE 8 (a) Attractive spin–orbit force; orbital angular momentum parallel to the spins. (b) Repulsive spin–orbit force; orbital angular momentum antiparallel to the spins. [Reprinted by permission from Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ, p. 27. Copyright 1965 by Prentice-Hall, Inc.]

A. Nuclear Shell Model

In the nuclear shell model, each nucleon in the nucleus is assumed to move under the influence of an average force created by the action of all the other nucleons. Using quantum mechanics we can calculate the energy levels of the nucleons in the nucleus (see Fig. 9).

Associated with each level is a set of quantum numbers (n, l, j, s), and the nucleon configurations are built up by filling in these levels, much as the atomic electron configurations are determined using the Aufbau principle. Between certain sets of energy levels, there are large gaps in the energy level spacings (shaded areas, Fig. 9). When one set of levels (lying between the shaded areas) is completely full, as determined by the Pauli principle, then we say that a closed-shell configuration has been achieved. These closed nucleon shells have special stability, similar to the inert-gas structures corresponding to

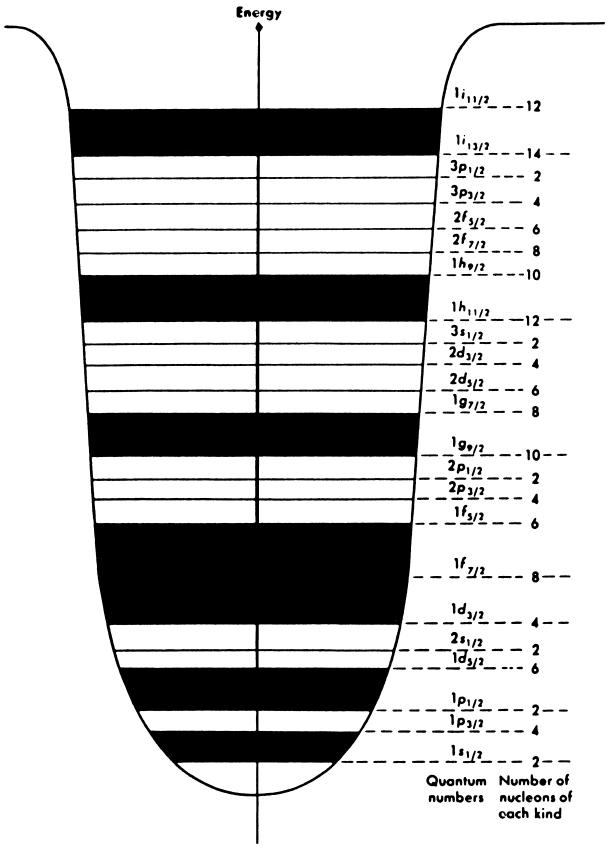


FIGURE 9 Schematic diagram of a nuclear potential well, showing energy levels of nucleons. Shaded areas show energy gaps corresponding to the filling of shells.

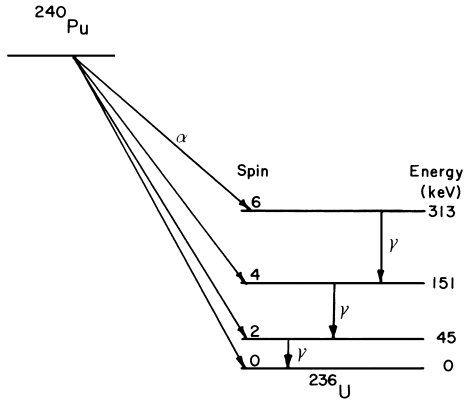


FIGURE 10 ^{240}Pu decay scheme. The excited states of ^{236}U decay by γ-ray emission.

closed electronic shells in the atom. The proton and neutron numbers corresponding to closed-shell configurations are called magic numbers and are 2, 8, 20, 28, 50, 82, and 126. The shell model predicts many of the regularities of nuclear structure and the special stability of nuclei with magic numbers of neutrons and protons.

B. Collective Model

The nuclear collective model assumes that certain features of nuclear behavior can be explained as resulting from the collective or coop erative motions of several nucleons acting together. Examples of such collective behavior are the occurrences of rotational and vibrational levels in nuclei. Figure 10 shows the excited states of ^{236}U. We deduce that this group of states represents configurations in which rotational waves are moving about the nuclear surface. The energies of the states should go as $I(I+1)$, where I is the nuclear spin quantum number. Note that the ratio of the spin-4 to spin-2 state is 151/45, which is very close to $4(4+1)/2(2+1)$.

V. ALPHA DECAY

All nuclei that are heavier than bismuth show some probability for α decay. In addition, some rare-earth nuclei also decay by α emission. There is a very close relationship between α-particle energy and the half-life of the decaying nucleus. The higher the particle energy, the shorter the half-life. This latter observation and the entire phenomenon of α decay find a natural explanation in quantum mechanics. The positive charge of the nucleus creates a zone of potential energy (a potential barrier) about the nucleus. This barrier acts to repel positively charged particles approaching the nucleus and to prevent charged particles within the nucleus, such as the α particle, from leaving the nucleus. Quantum mechanics tells us, however, that a finite probability exists that the α particle may "tunnel" through this barrier and emerge from the nucleus. The rate of barrier penetration is dependent on the energy of the α particle and accounts for the great sensitivity of the α half-life to the emitted α-particle energy. Up to now, we have assumed that the α particle already exists within the decaying nucleus, but this is not true. The α particle must be assembled from two neutrons and two protons in the nucleus. This α-particle assembly is easiest if the decaying nucleus has an even number of neutrons and protons. When the nucleus contains an odd number of nucleons, this assembly is more difficult, and the α decay proceeds to an excited state of a daughter nucleus (and thus the emitted α-particle energy is low and the half-life longer).

In 1984, Rose and Jones reported the first example of a new mode of radioactive decay, decay by spontaneous emission of heavy particles ($Z \geq 6$), in the decay $^{223}\text{Ra} \rightarrow {}^{14}\text{C} + {}^{209}\text{Pb}$. The probability for this decay was 10^{-9} of the α-decay probability. Since that pioneering experiment, a number of other examples of heavy particle radioactivity have been observed, such as ^{24}Ne and ^{28}Mg emission by nuclei with $Z \geq 88$ (allowing the daughter nucleus to be stabilized by the $Z = 82$ or $N = 126$ shell closures).

VI. BETA DECAY

Nuclear beta (β) decay is a relatively slow process involving the emission of electrons and neutrinos by a nucleus. Since neither of these species exists in the nucleus, they must be created at the moment of decay. The interaction responsible for β decay is the weak interaction, a force that is different from the gravitational, electromagnetic, or nuclear forces. The quantum statistical mechanical theory of β decay accounts for the fact that the emitted electron and neutrino share the available decay energy between them, thus leading to a continuous energy spectrum of the emitted β particles. The greater the overall decay energy in β decay, the shorter the half-life. Another factor influencing β-decay half-lives is the similarity between parent and daughter nuclei. Since the weak interaction cannot bring about great alterations in nuclear structure, β decay that necessitates such changes is unlikely. The most likely β decays are those between mirror nuclei, nuclei in which one nucleus has the same number of protons as the other has neutrons and vice versa. In addition to decay energy and similarity between parent and daughter nuclei, the rate of EC decay is also influenced by the ability of the nucleus to capture an orbital electron. Since K electrons have the greatest density near the nucleus, their capture is favored.

VII. GAMMA DECAY

The emission of gamma (γ) rays from nuclei in many respects resembles the emission of radio waves from an antenna. Both processes are due to the movement of charged particles: nucleons in the nuclear case, electrons in an antenna. In both cases, the radiation carries away energy and angular momentum. A photon can carry off any number of units of angular momentum, equal to or greater than $1\hbar$. (This minimum is due to the spin of the photon, which is equal to $1\hbar$.) Therefore, the initial and final nuclei must not both have zero spin, for then there would be no way to conserve angular momentum.

The greater the energy of the emitted photon, the greater the probability of γ-ray decay. The greater the angular momentum removed by the photon, the less probable γ-ray decay is. Gamma-ray transitions are classified by transition type. When the emission of the photon is caused by a disturbance in the electric field, the event is said to be an electric transition. When the disturbance of the magnetic field is responsible, the event is said to be a magnetic transition. A γ ray emitted as a result of an electric transition is exactly the same as one emitted as a result of a magnetic transitio, but the probabilities of the two types of events are different. The letters E and M are used to stand for electric and magnetic transitions; they are followed by a number equal to the number of units of angular momentum removed by the γ ray. Thus an E3 transition involves the loss of 3 units of angular momentum as a result of an electric transition. These rules are summarized in Table I. The half-lives for electric transitions are about 100 times shorter than for magnetic transitions of the same energy.

Another form of nuclear electromagnetic decay is internal conversion. In internal conversion, the electric and magnetic disturbances (which cause the emission of a photon in γ-ray decay) interact directly with an orbital atomic electron, causing that electron to be ejected from the atom. The kinetic energy of the conversion electron is $E_\gamma - E_B$, where E_γ is the transition energy and E_B the electron binding energy. Another important quantity that characterizes internal conversion is the internal conversion coefficient α, which is defined as the (number of decays proceeding by internal conversion)/(number of γ-ray decays). Internal conversion (large values of α) is favored by low-transition energies and large spin changes in the decay.

TABLE I γ-Ray Emission Rules

Radiation type	Name	Angular momentum carried away	Does nuclear parity change?
E1	Electric dipole	1	Yes
M1	Magnetic dipole	1	No
E2	Electric quadrupole	2	No
M2	Magnetic quadrupole	2	Yes
E3	Electric octupole	3	Yes
M3	Magnetic octupole	3	No
E4	Electric 2^4 pole	4	No
M4	Magnetic 2^4 pole	4	Yes

VIII. NUCLEAR REACTIONS

A. Basic Features

A nuclear reaction involves the interaction of one nucleus with another to alter the reacting partners in some way. One of the reacting nuclei is usually at rest (the target), while the other nucleus (the projectile) approaches it with sufficient kinetic energy to cause a reaction. If we consider a projectile nucleus P interacting with a target nucleus T to produce an emitted particle E and a residual nucleus R, we designate this reaction by the shorthand notation $T(P, E)R$. Thus the reaction of neutrons with ^{27}Al to make ^{24}Na is written as a ^{27}Al(n, α)^{24}Na reaction. As one notes from studying these examples, in ordinary nuclear reactions there is a strict conservation of the number of neutrons and protons in the reaction. The energy release in a nuclear reaction is called the Q value of the reaction and is given by the equation

$$Q = \sum(\text{masses of reactants})C^2 - \sum(\text{masses of products})C^2, \qquad (13)$$

where C is the speed of light. Exoergic reactions have positive Q values.

The probability of a nuclear reaction occurring is specified by giving the cross section for that reaction. The cross section represents the classical area a target nucleus would present to an incoming projectile. Large cross sections imply large-area target nuclei and/or high reaction probabilities. The units of cross section are area, usually expressed in square centimeters. A nucleus with a cross-sectional area of 10^{-24} cm^2 is said to be "as easy to hit as the broad side of a barn." Thus nuclear cross sections are traditionally reported in units of 10^{-24} cm^2, or barns.

Experimental studies of nuclear reactions are usually aimed at measuring the following:

1. Total probability of occurrence of a reaction, that is, the total reaction cross section $\sigma_{R'}$.
2. Probability of occurrence of the reaction as a function of the energy of the incident projectile, the excitation function.
3. Energy and angular distribution of the emitted particles, the differential cross section $d^2\sigma/dEd\Omega$, where

$$\sigma_R = \int_0^E \int^{\max} \frac{d^2\sigma}{dE\, d\Omega}\, d\Omega\, dE. \qquad (14)$$

4. Energy and angular distribution of the residual nuclei.

5. Yields of various emitted particles or residual nuclei.

B. Reaction Mechanisms

1. Nucleus-Induced Reactions

Nuclear reactions, like chemical reactions, take place by a variety of reaction mechanisms. In Figure 11, a simple conceptual model for illustrating the relationship between various reaction mechanisms is shown.

As projectile a moves near target nucleus A, a has a certain probability of interacting with the nuclear force field of A, causing a to suffer a change in direction but no loss of energy (Fig. 12a). This reaction mechanism is called shape elastic scattering. If shape elastic scattering does not occur, then a may interact with A via a two-body collision between a and some nucleon of A, raising the A nucleon to an unfilled level (Fig. 12b). If the struck nucleon leaves the nucleus, a direct reaction is said to have occurred. If the struck nucleon does not leave the nucleus, further two-body collisions may occur, and eventually the entire energy of the nucleus may be distributed among the nucleons of the $a + A$ combination, leading to the formation of a compound nucleus C. Loosely speaking, the compound nucleus "forgets" its mode of formation, so its subsequent breakup depends only on the excitation energy, angular momentum, and so on of C and not the nature of a or A. Sometimes the compound nucleus may emit a particle of the same kind as a (or maybe even a itself) with the same energy that a had when the original reaction process

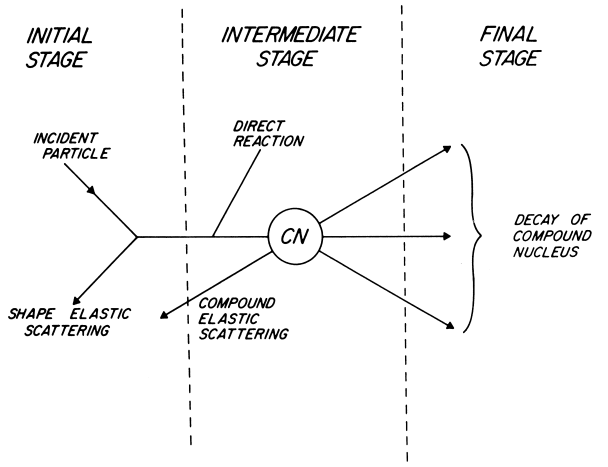


FIGURE 11 Schematic view of the time evolution of a nuclear reaction.

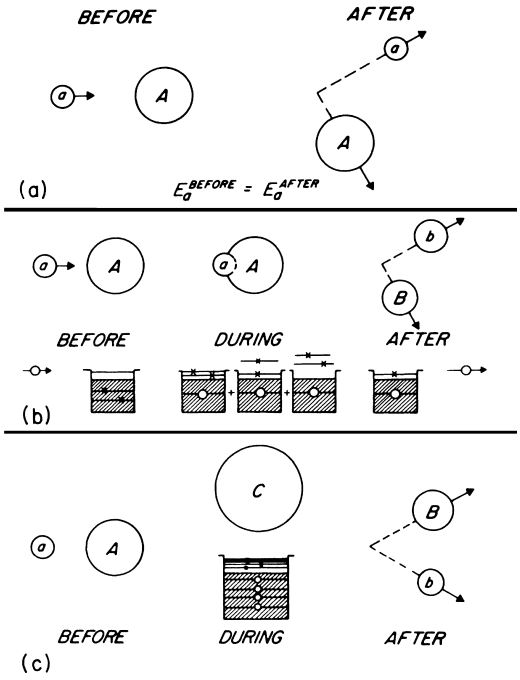


FIGURE 12 Schematic view of the various types of nuclear reaction mechanisms. (a) Shape elastic scattering (the energies are given in the center-of-mass system). (b) Direct reactions. (c) Compound nuclear reaction.

started. If this happens, we say that compound elastic scattering has occurred. Also, C may decay into other reaction products, such as $B + b$ or $D + d$, which are unlike either a or A.

Many reactions take place partly by compound nucleus and partly by direct reaction mechanisms. The relative contributions of the two processes can be determined by measuring the angular distribution of the reaction products. For compound nucleus reactions, this distribution is symmetric with respect to the plane perpendicular to the incident beam axis, while the angular distributions for direct reactions are forward-peaked. The probability of compound nucleus reactions can be calculated. In these calculations the probability of a given reaction is calculated as the product of two factors: the probability of forming the compound nucleus and the probability that the compound system will break up in a particular manner. Direct reactions, such as inelastic scattering or nucleon transfer reactions, are described using quantum mechanics.

2. Heavy-Ion-Induced Reactions

In the foregoing discussion of reaction mechanisms, the character of the projectile nucleus was not considered [i.e., whether it was a nucleon, a small nucleonic cluster, or a heavy ion ($A > 4$)]. Although heavy ions undergo many

interactions in a manner similar to lighter nuclei, the use of heavy ions as projectiles in nuclear reactions does introduce certain new features into the study of reaction mechanisms. Because the associated de Broglie wavelength of a heavy ion is small compared with its size, the interactions of heavy ions with nuclei show semiclassical features, allowing a qualitative and semiquantitative discussion of these reactions without the use of quantum mechanics. In addition, the large nuclear charge of the projectile and its large associated orbital angular momentum strongly influence the reaction mechanisms.

It is possible to give a description of heavy-ion reactions in an alternate manner, emphasizing a quantity called the impact parameter, which is the distance of closest approach if the incident particle continued without deflection. Four impact parameter regions corresponding to different reaction mechanisms can be defined: (1) the Coulomb region (distant collisions), (2) the peripheral region (grazing collisions), (3) the deep inelastic region (close collisions), and (4) the fusion region leading to compound nucleus formation. The orbits corresponding to these regions are shown in Figure 13.

In the peripheral region, the ions brush past each other in a grazing collision. The elastic cross section departs from the Rutherford cross section, nuclear inelastic excitation occurs, and one or more nucleons can be transferred from one nucleus to another. These interactions can be analyzed in the same way as the corresponding reactions initiated by nucleons, deuterons, and α particles.

In the deep inelastic region, the ions interact very strongly, and much kinetic energy becomes internal excitation energy, without the ions essentially losing their identities, in the sense that only a few nucleons are transferred from one to the other.

In the fusion region, the ions interact so strongly that they fuse together to form a compound nucleus that subsequently decays.

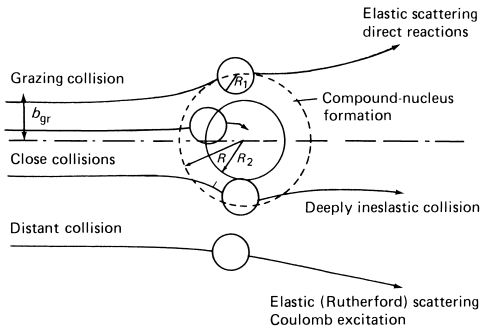


FIGURE 13 Classical picture of heavy-ion interactions showing the trajectories corresponding to distant, grazing, and close collisions.

The same formalism is used to describe compound nucleus reactions induced by light and heavy ions, with the restriction that there is an upper limit J_{crit} to the angular momentum of the compound system. Heavy-ion reactions could easily produce compound systems of very high angular momentum; and if such nuclei fission instantly, then no compound nucleus is formed.

Early in the study of heavy-ion reactions, it was found that the projectile could lose as much as 200 MeV of its energy without substantially altering its identity. The angular distributions are strongly forward-peaked and nonsymmetric, indicating that the compound system separates in a time that is shorter than its period of rotation. It is also found in some cases that the energy of the nuclei after the interaction is very similar to the electrostatic repulsion of the two nuclei when just touching, showing that the transfer of kinetic energy from the projectile to excite the target nucleus is essentially complete. This process is generally known as deep inelastic scattering.

To account for these observations it has been suggested that the projectile traverses a classical orbit through the surface region of the target nucleus, losing energy continually. There is then a relation between the path length inside the nucleus and the scattering angle. Theories of the interaction have been developed using the classical concepts of viscosity and friction.

3. High-Energy Reactions

When the projectile energy exceeds ~200 MeV/nucleon, one is in the realm of high-energy nuclear reactions. At these energies, the interactions of the colliding species are dominated by the effect of individual nucleon–nucleon collisions, whereas at low energies (<10 MeV/nucleon), the colliding species move in the mean or average force field created by the presence of the other nucleus. Three reaction mechanisms dominate in high-energy p–nucleus collisions:

1. Spallation, in which a small portion of the target nucleus is spalled or chipped off by the action of the incident proton.
2. Fission, which is generally similar to fission induced at lower energies with the exception that the excitation energy of the fissioning system is quite high, and thus the product distributions are quite broad.
3. Fragmentation, a process in which the target nucleus is shattered or fragmented into several smaller pieces.

Many aspects of p–nucleus collisions can be understood in terms of a cascade–evaporation model. In this model, the incident proton is assumed to collide with

the target nucleons one at a time, setting up a shower or cascade of struck nucleons that propagate through the nucleus. Following this fast process (which takes place in $<10^{-22}$ sec), the excited product nucleus then deexcites by particle emission or fission.

When the projectile is a heavy ion, one predicts that in addition to spallation, fission, and fragmentation reactions, some interactions will lead to the formation of nuclear matter at temperatures and densities that never have existed before, except perhaps in the early history of the universe or in neutron stars. At high enough energies and nuclear densities, it is predicted that a plasma or quarks and gluons will be observed. Current research efforts in nuclear chemistry are directed toward finding evidence for the existence of the quark–gluon plasma.

C. Fission

Of all nuclear reactions, fission has a special place because of its technological significance. Figure 14 shows the time sequence of a fission reaction. A target nucleus is excited by means of a nuclear reaction and begins to

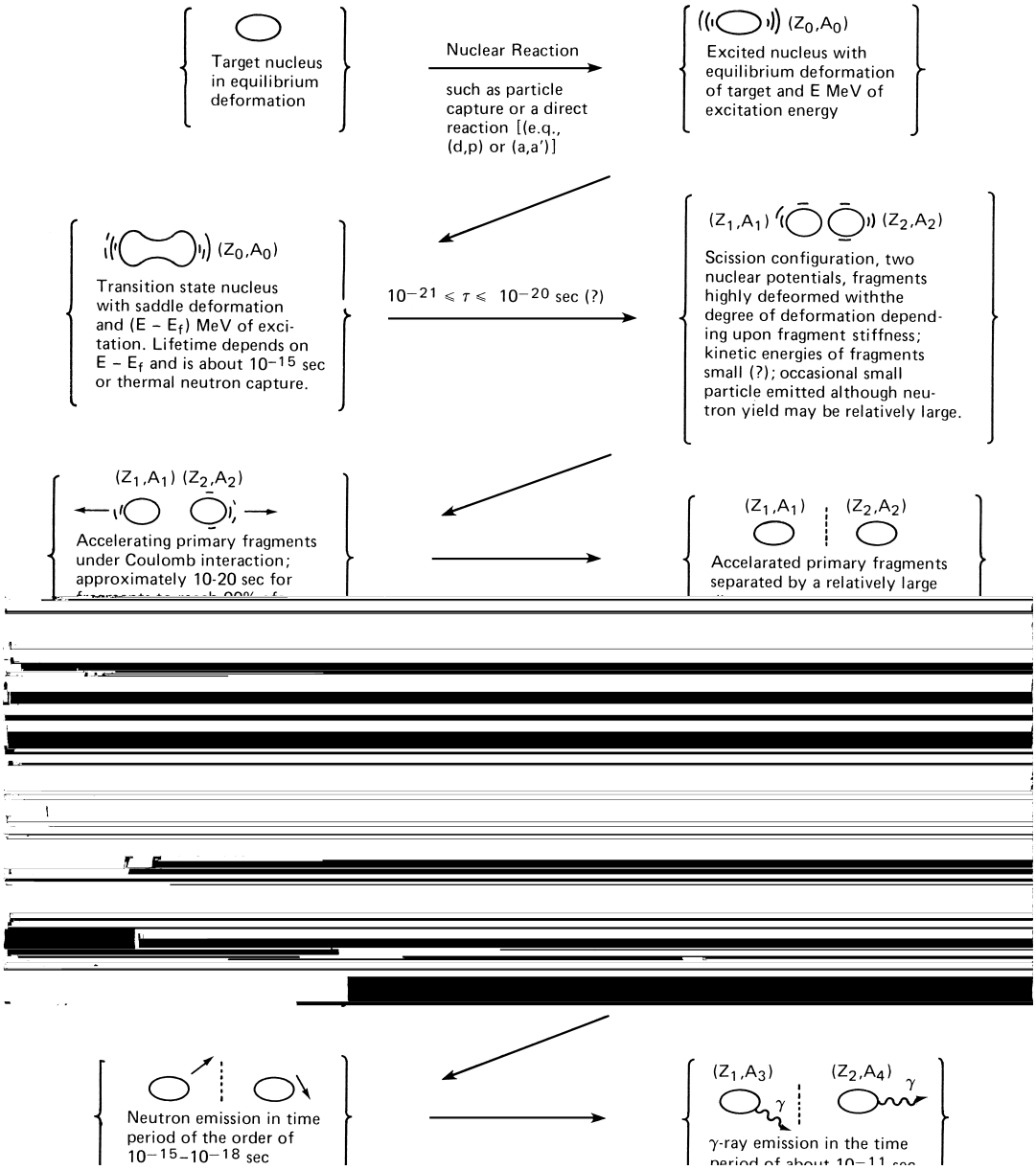


FIGURE 14 Schematic diagram of the fission process. [Reprinted from Gindler, J. E., and Huizenga, J. R. (1968). In "Nuclear Chemistry," Vol. 2, L. Yaffe, ed. Academic Press, New York, p. 4. Copyright 1968 by Academic Press.]

oscillate and deform until it finally reaches the transition state deformation, for which the rate of change of the nuclear surface energy equals the rate of change of the nuclear coulomb energy. Any further deformation beyond this point leads irretrievably to fission. As the neck between the two nascent fragments dissolves, the scission configuration is reached. At this point, two highly charged deformed fragments are in contact, and their mutual coulomb repulsion pushes them away from each other. In about 10^{-20} sec they achieve 90% of their final total kinetic energy, which is the bulk of the energy release in the fission process. As they move away from one another, the highly deformed fragments contract to spheroidal shapes, and in the process they heat up. They get rid of this excess excitation energy by evaporating neutrons, as a hot drop of water evaporates water molecules. These are the bulk of the neutrons emitted in the fission process, the so-called prompt neutrons. Following neutron emission and in competition with the latter stages of this process, the fragments emit γ rays. The final fission fragment nuclei are neutron rich and decay by β emission. Occasionally one of these β decays excites a level in the daughter nucleus that is so high in energy as to be unstable with respect to neutron emission; these later emitted neutrons are called the delayed neutrons. These are the neutrons used to control the operation of nuclear reactors.

When the projectile inducing fission is a low-energy (<1 MeV) neutron and the target nucleus is a heavy element such as U or Pu, a wide range of fragments of differing masses is produced. The fragment mass distribution for typical low-energy fission of actinide nuclei is shown in Figure 15. One can see that most probable fission mass split is asymmetric, with the fragment mass ratio being ~1.4. Because of the near constancy of the left-hand edge of the heavy mass peak at $A = 132\,(Z \sim 50,\ N \sim 82)$ as the mass of the fissioning system changes, it is thought that this preference for asymmetric fission is due to the special stability of the $A \sim 132$ fragments, which result from nuclear shells at $Z = 50$ and $N = 82$. As the excitation energy of the fissioning system increases, the valley between the two mass peaks fills in, and eventually the dominant mass split becomes symmetric at high energies. The total kinetic energy (TKE) release in fission can be understood in terms of the coulomb repulsion between the nascent fragments as they are formed. To a good first approximation, this energy release is given as

$$\mathrm{TKE} = Z_1 Z_2 e^2 \big/ 1.8\left(A_1^{1/3} + A_2^{1/3}\right), \qquad (15)$$

where Z_1, Z_2, A_1, and A_2 refer to the two fragments, and the square of the electronic charge e^2 is 1.44 MeV fm. Approximately 2.5 neutrons are emitted per fission event in the fission of ^{235}U by thermal neutrons.

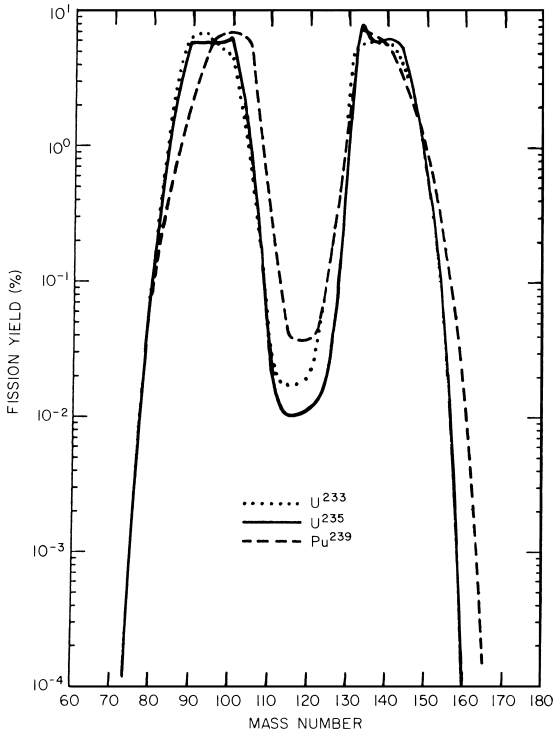


FIGURE 15 Mass yield distributions for typical heavy nuclei.

IX. PRODUCTION OF NEW ELEMENTS

Over 40 years ago, the first transuranium elements, neptunium and plutonium, were synthesized and identified (i.e., "discovered"). The intervening years have witnessed the discovery of 18 more, so this group now consists of 20 known elements ranging from neptunium (element number 93) to the unnamed element with atomic number 112. All of these elements are synthetic in that they do not exist in appreciable quantities in nature. Therefore, these elements represent a 20% expansion of our heritage of the building blocks of nature. Elements 93–101 were synthesized in nuclear reactions involving light projectiles, whereas the others were synthesized using heavy ions.

The history of the discovery of these elements is a fascination story that has been told in a variety of ways and places. The discovery of elements 93–103 is relatively straightforward and is summarized in Table II. From examining accounts of these discoveries, one concludes that there is a continuous gradation in complexity and difficulty of synthesis as the atomic number increases. The identification of mendelevium and higher Z elements was made on a one-atom-at-a-time basis. In all element discovery experiments, it was crucial to the claim of discovery that the Z of the reaction product be cleanly identified using

TABLE II Summary of Actinide Element Synthesis

Element (symbol)	Z	Synthesis reaction	Discoverers and date of discovery
Neptunium (Np)	93	$^{238}\mathrm{U} + \mathrm{n} \rightarrow {}^{239}\mathrm{U} + \gamma$ $^{239}\mathrm{U} \xrightarrow{\beta^-} {}^{239}\mathrm{Np}(t_{1/2} = 2.35\ \mathrm{day})$	E. M. McMillan, P. H. Abelson, 1940
Plutonium (Pu)	94	$^{238}\mathrm{U} + {}^{2}\mathrm{H} \rightarrow {}^{238}\mathrm{Np} + 2\mathrm{n}$ $^{238}\mathrm{Np} \xrightarrow{\beta^-} {}^{238}\mathrm{Pu}(t_{1/2} = 86.4\ \mathrm{yr})$	G. T. Seaborg, E. M. McMillan, J. W. Kennedy, A. C. Wahl, 1940–1941
Americium (Am)	95	$^{239}\mathrm{Pu} + \mathrm{n} \rightarrow {}^{240}\mathrm{Pu} + \gamma$ $^{240}\mathrm{Pu} + \mathrm{n} \rightarrow {}^{241}\mathrm{Pu} + \gamma$ $^{241}\mathrm{Pu} \xrightarrow{\beta^-} {}^{241}\mathrm{Am}(t_{1/2} = 433\ \mathrm{yr})$	G. T. Seaborg, R. A. James, L. O. Morgan, A. Ghiorso, 1944–1945
Curium (Cm)	96	$^{239}\mathrm{Pu} + {}^{4}\mathrm{He} \rightarrow {}^{242}\mathrm{Cm}(t_{1/2} = 162.5\ \mathrm{day}) + \mathrm{n}$	G. T. Seaborg, R. A. James, A. Ghiorso, 1944
Berkelium (Bk)	97	$^{241}\mathrm{Am} + {}^{4}\mathrm{He} \rightarrow {}^{243}\mathrm{Bk}(t_{1/2} = 4.5\ \mathrm{hr}) + 2\mathrm{n}$	S. G. Thompson, A. Ghiorso, G. T. Seaborg, 1949
Californium (Cf)	98	$^{242}\mathrm{Cm} + {}^{4}\mathrm{He} \rightarrow {}^{245}\mathrm{Cf}(t_{1/2} = 44\ \mathrm{min}) + \mathrm{n}$	S. G. Thompson, K. Street, Jr., A. Ghiorso, G. T. Seaborg, 1950
Einsteinium (Es)	99	Mike thermonuclear explosion	A. Ghiorso, S. G. Thompson, G. H. Higgins, G. T. Seaborg, M. H. Studier, P. R. Fields, S. M. Fried, H. Diamond, J. F. Mech, G. L. Pyle, J. R. Manning, C. I. Browne, H. L. Smith, R. W. Spence, 1952
Fermium (Fm)	100	Mike thermonuclear explosion	A. Ghiorso, S. G. Thompson, G. H. Higgins, G. T. Seaborg, M. H. Studier, P. R. Fields, S. M. Fried, H. Diamond, J. F. Mech, G. L. Pyle, J. R. Huizenga, A. Hirsch, W. M. Manning, C. I. Browne, H. L. Smith, R. W. Spence, 1953.
Mendelevium (Md)	101	$^{253}\mathrm{Es} + {}^{4}\mathrm{He} \rightarrow {}^{256}\mathrm{Md}(t_{1/2} = 75\ \mathrm{min}) + \mathrm{n}$	A. Ghiorso, B. G. Harvey, G. R. Choppin, S. G. Thompson, G. T. Seaborg, 1955
Nobelium (No)	102	$^{246}\mathrm{Cm} + {}^{12}\mathrm{C} \rightarrow {}^{258}\mathrm{No} + 4\mathrm{n}$	A. Ghiorso, T. Sikkeland, J. R. Walton, G. T. Seaborg, 1958
Lawrencium (Lr)	103	$\left.\begin{matrix}{}^{250}\mathrm{Cf}\\ {}^{251}\mathrm{Cf}\\ {}^{252}\mathrm{Cf}\end{matrix}\right\} + {}^{11}\mathrm{B} \rightarrow {}^{258}\mathrm{Lr} + \left\{\begin{matrix}3\mathrm{n}\\ 4\mathrm{n}\\ 5\mathrm{n}\end{matrix}\right.$ $\left.\begin{matrix}{}^{250}\mathrm{Cf}\\ {}^{251}\mathrm{Cf}\\ {}^{252}\mathrm{Cf}\end{matrix}\right\} + {}^{10}\mathrm{B} \rightarrow {}^{258}\mathrm{Lr} + \left\{\begin{matrix}2\mathrm{n}\\ 3\mathrm{n}\\ 4\mathrm{n}\end{matrix}\right.$	A. Ghiorso, T. Sikkeland, A. E. Larsh, R. M. Latimer, 1961

chemical or physical techniques. This criterion continues to be applied today to claims of discovery of new elements.

Elements 104–106 were discovered in Berkeley by scientists working at the Lawrence Radiation Laboratory with important contributions being made by scientists working at Dubna in Russia. These elements have been given the names rutherfordium (104), dubnium/hahnium (105), and seaborgium (106). The internationally accepted name for element 105 is dubnium but it is referred to in the United States as hahnium, a reflection of a long-standing controversy over who was the first group to synthesize and identify this element.

Element 107 was produced in 1981 by Münzenberg et al. using the reaction $^{209}\mathrm{Bi}(^{54}\mathrm{Cr}, \mathrm{n})^{262}$ 107. (This element was subsequently named bohrium and given the chemical symbol Bh.). The recoiling product nuclei from the nuclear reaction were passed through a velocity filter (called SHIP), which guaranteed that they had the characteristic velocity of the product of the complete fusion of the projectile and target nuclei. The mass number of the velocity separated product nuclei was roughly determined using a time-of-flight spectrometer and the atomic number and mass number were determined by observing the time correlated α-decay of ^{262}Bh to its decay products. The initial experiment involved the detection of five atoms of Bh, however, subsequent experiments have led to the production of many more atoms.

Element 108 (hassium, chemical symbol Hs) was first observed in 1984 by the team of Münzenberg et al. Using the reaction $^{208}\mathrm{Pb}(^{58}\mathrm{Fe}, \mathrm{n})$ to make three atoms of ^{265}Hs, the products were identified by their α decay with half-lives of the order of milliseconds. This work was especially significant as it showed us that the spontaneous fission decay of these nuclei did not limit their production and identification, portending the synthesis of still heavier nuclei.

Element 109 (meitnerium, chemical symbol Mt) was first produced in 1982 by Münzenberg et al. using the $^{209}\mathrm{Bi}(^{58}\mathrm{Fe}, \mathrm{n})$ reaction to produce one atom of ^{266}Mt. A subsequent experiment in 1988 by the same group produced two more atoms of element 109 and thus confirmed the original work.

Element 110, unnamed at present, was distinctively identified when it was produced by Hofmann et al. in the reaction of ^{62,62}Ni with ^{208}Pb (to make ^{269,271}110). (A prior experiment by Ghiorso et al. had also produced an isotope of element 110, ^{267}110, using the ^{209}Bi(^{59}Co, n) reaction but the identification of the reaction products had not been definitive). Subsequently, element 110 was produced by the actinide-based reaction ^{244}Pu(^{34}S, 5n) ^{273}110 by workers at Dubna. All four isotopes of element 110 agreed with modern predictions of the expected half-lives of the heaviest nuclei. The sharp increase in half-life from a few microseconds (^{267}110) to greater than 100 msec (^{273}110) has been taken as strong evidence for the stabilizing effect of the N = 162 deformed neutron shell.

Elements 111 and 112 were discovered by Hofmann and co-workers in a series of stunning experiments at the GSI laboratories in Darmstadt, Germany. The nuclides ^{272}111 and ^{277}112 were made using the ^{209}Bi(^{64}Ni, n) and ^{208}Pb(^{70}Zn, n) reactions, respectively. The cross sections for these reactions were low enough (~pb) to be discouraging for future attempts to synthesize new elements using cold fusion reactions.

In the study of heavy ion reactions resulting in transuranium reaction products, it is of paramount importance to be able to isolate and uniquely identify the products as to their Z, A, and formation cross section. Indeed, the claim to discovery of a new element must involve identification of Z, whereas the claim of discovery of a new nuclide must involve measurement (and/or deduction) of both Z and A.

For reaction products with the longest half-lives, chemical separation techniques offer convenient methods of isolating individual reaction products and establishing their atomic numbers. These techniques offer the greatest sensitivity of all methods because of the large amounts of target material that can be used.

For species with half-lives in the range $0.1 \leq t_{1/2} \leq 10$ sec, the helium jet is a superior method of isolating reaction products, as witnessed by its use in the discovery of new elements. In this method, reaction products recoiling from the target are thermalized in ~1 atm helium, which exists in the target chamber via a connection to a low-pressure area, creating a jet or stream of helium (Fig. 16). The helium gas stream impinges on a collection device such as a tape, wheel, or drum that moves the activities (radioactive reaction products) to the detectors. The selectivity of the jet system may be improved by performing a gas-phase chemical separation in the jet during transport of the stopped recoils.

Identification of the collected reaction products can be made with a variety of techniques. Perhaps the most

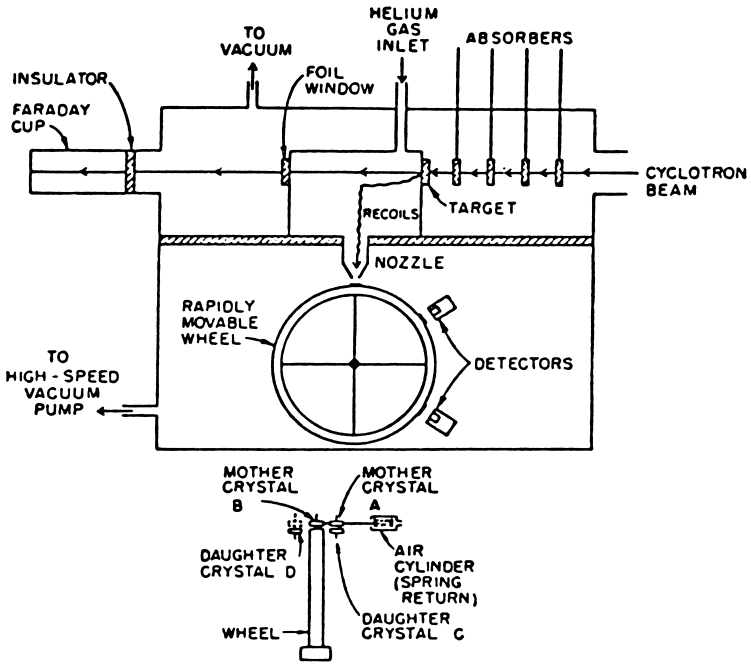


FIGURE 16 Schematic representation of a gas-jet recoil transport assembly. Thermalized product atoms are transported in the He gas stream and collected on the periphery of a wheel or other suitable collection device. Periodically, the wheel is moved to position the spot in front of the detectors. A mother–daughter detector assembly is illustrated in the lower portion of the figure and is used to establish a genetic link.

important of these techniques is the mother–daughter, or double-recoil method, which establishes a genetic link between the unknown reaction product and known daughter and/or granddaughter activities. In this technique (Fig. 16), the recoil heavy atom produced by the α decay of the collected initial reaction product strikes and imbeds itself in a "mother" crystal. The mother crystal is then moved in front of a "daughter" crystal that can detect the α decay of the imbedded atom in the mother crystal. If the α-particle decay characteristics of the daughter nucleus are known, then a genetic link is established, and the (Z, A) of the parent are established. This technique was used in the discovery of several elements and isotopes.

The principal problem with the isolation devices discussed previously (tapes, jets, etc.) is that the reaction product must be stopped and mechanically transported to radiation detectors before product identification can occur. This restricts the use of such devices to studies of nuclei whose $t_{1/2} \geq 1$ msec. For detection and identification of species whose $t_{1/2} \geq 1$ μsec, one employs an instrument based on magnetic and/or electrostatic deflation of target recoils. The most spectacularly successful of these devices in recent years is the velocity filter SHIP (separator for heavy ion reaction products). A schematic diagram of this separator is shown in Figure 17. Evaporation residues produced in a nuclear reaction emerge from the target and pass through a thin carbon foil, which has the effect of equilibrating the ionic charge distribution of the residues. The ions then pass through two filter stages consisting of electric deflectors, dipole magnets, and a quadrupole triplet for focusing. The solid angle of acceptance of the separator is 2.7 msr with a separation time for the reaction products of $\sim$2 μsec. Because complete fusion evaporation residues have very different velocities from target like transfer and deep-inelastic products, the separator with its $\pm$5% velocity acceptance range uniquely separates the evaporation residues from the other reaction products. Following separation, the residues pass through a large area time-of-flight detector and are stopped in an array of seven position-sensitive detectors. From their times of flight and their energy deposits in the position-sensitive detectors, a rough estimate of their mass can be obtained. The final genetic identification of the residues is made by recording the correlations between position in the detector (average residue velocity) and subsequent decay signals (from α or spontaneous fission decay) or even signals from γ- or X-ray detectors placed next to the position-sensitive detector. This device was used in the discovery of elements 107–112.

To detect species with lifetimes that are substantially less than 1 μsec, special techniques must be employed. They include time-of-flight (TOF) techniques which, when combined with a measurement of the product energy, give information about the product mass number. For suitable mass resolution, the time of flight must be $\sim$10 nsec. When searching for rare events, some selection process (such as SHIP) must be employed to reduce the background levels in the apparatus. The decay-in-flight and the crystal-blocking techniques ($10^{-18} \leq t_{1/2} \leq 10^{-14}$ sec) give very little information about the identity of a reaction product other than its existence and its approximate lifetime.

As the masses of the newly synthesized transuranium nuclei have increased, the emphasis has shifted from using light-ion-induced reactions to using heavy-ion-induced reactions as the mode of synthesis. Thus, as mentioned previously, elements 93–101 were synthesized first in reactions induced by neutrons, deuterons, and helium ions, whereas the remaining transuranium element syntheses involved the use of heavy ions such as ^{10}B, ^{11}B, ^{12}B, ^{13}C, ^{15}N, and ^{18}O. Recent developments in accelerator technology have made the use of heavy-ion projectiles as massive as ^{238}U readily available.

The synthesis of heavy nuclei is a two-step process involving the fusion of projectile and target nuclei to form a excited compound nucleus followed by a competition

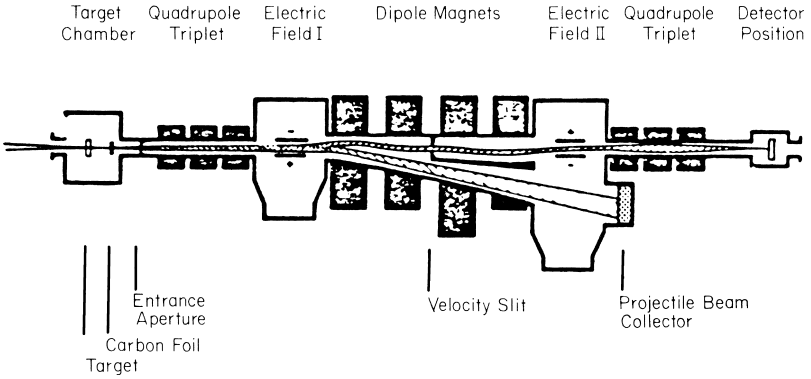


FIGURE 17 Schematic diagram of the velocity filter SHIP at GSI.

between neutron emission and fission in the de-excitation of this compound nucleus. The probability of forming a new heavy nucleus is thus the product of two factors related to the formation of the completely fused system and the survival of the resulting nucleus (against fission). There are two approaches to the synthesis of heavy nuclei: (1) the "cold fusion" reactions involving Pb or Bi target nuclei that produce compound nuclei with low excitation energies ($E^* \sim 13$ MeV), thus ensuring large survival probabilities but at the expense of a reduced formation probability, and (2) the "hot fusion" reactions, involving actinide target nuclei, producing highly excited products ($E^* \sim 30$–60 MeV). The hot fusion reactions do not suffer from the fusion hindrance factors, which may limit the Pb-Bi, based reactions, but because of the low survival probability of the product nuclei against fission, they were not used in the synthesis of elements 107–112.

The scientist who wishes to predict the outcome of a given reaction producing a transuranium product is faced with a difficult chore. In some cases, he or she is forced to predict the second or higher moments of the initial product distributions to estimate the final product yields. In any case, the proper calculation of the survival probability of the initial reaction products may require careful consideration of the fission barriers, nuclear shapes and symmetries, shell effects, and masses, and how they vary with energy, angular momentum, and deformation.

As of 1970, the experimental data on the stability of the transuranium elements seemed to indicate that a practical limit to the periodic table would be reached at approximately element 108 (Fig. 18). By extrapolation of the data existing at that time, one would have concluded that at about element 108, the half-lives of the longest-lived isotopes of the elements would become so short ($<10^{-6}$ sec) due to decay by spontaneous fission as to preclude their production and study. However, during the period from 1966–1972, a number of theoretical calculations based upon modern theories of nuclear structure showed that in the region of proton number $Z = 114$ and neutron number $N = 184$, the spherical ground states of nuclei were stabilized against fission. This stabilization was due to the complete filling of proton and neutron shells and is analogous to the stabilization of chemical elements, such as the noble gases, due to the filling of electronic shells in these atoms. Even more interesting, some of these superheavy nuclei were predicted to have half-lives of the order of the age of the universe, thus stimulating efforts to find these "missing" elements in nature. The superheavy elements were predicted to form an island of relative stability extending both above and below $Z = 114$ and $N = 184$ and separated from the peninsula of known nuclei by a sea of instability (Fig. 19).

One fact should be emphasized from the outset: While the various theoretical predictions about the superheavy nuclei differ as to the expected half-lives and regions of stability, all theoretical predictions are in agreement as to the existence of superheavy nuclei. Thus the search for superheavy nuclei remains as a unique, rigorous test of the predictive power of modern theories of the structure of nuclei.

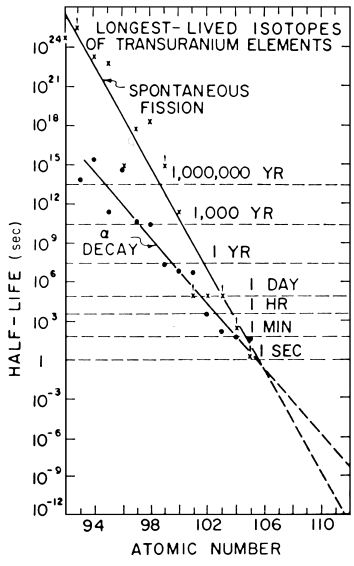


FIGURE 18 The half-life of the longest-lived isotope of a given element versus atomic number, as known in 1970.

X. RECENT DEVELOPMENTS

Prior to 1999, the prospects for synthesizing new elements beyond element 112 did not look promising. Straightforward extrapolations of the cross sections for producing the lighter elements would indicate cross sections of $\sim$1 pb or less for producing further new nuclei. The scientific community was excited and shocked by the report of a team from Dubna and Livermore that they had successfully synthesized element 114. The synthesis reaction chosen was a hot fusion reaction, ^{244}Pu(^{48}Ca, 3n) ^{289}114. Using a gas-filled separator at Dubna, these workers observed a single spectacular decay chain that was taken to signal the birth of a new element. The decay chain involved the emission of alpha particles from species with half-lives that were exceedingly long (30 sec and 15 min). These long lifetimes are the first indications of a close approach to the long sought island of stability thought by some to be near $Z = 114$ and $N = 184$. Later reports indicated the possible synthesis of the other isotopes of element 114. If confirmed and extended, this work represents the fulfillment of the long quest for superheavy nuclei begun in the 1960s.

In May of 1999, the group at Berkeley reported another startling discovery, the successful synthesis of elements

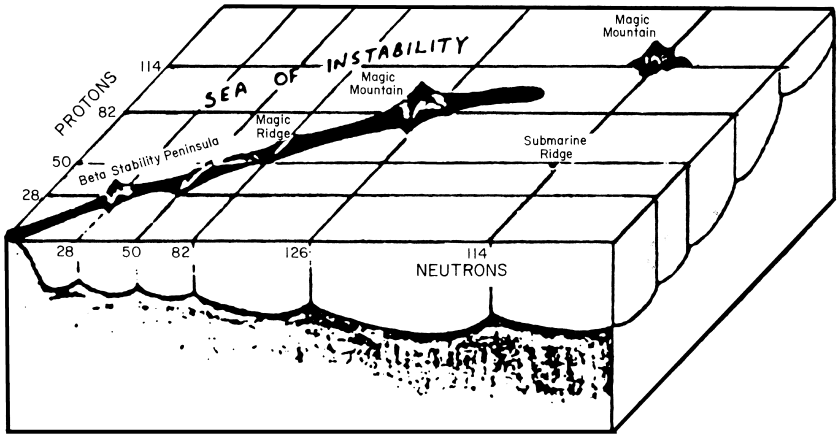


FIGURE 19 Allegorical representation of the stability of nuclei, showing a peninsula of known nuclei and an island of superheavy nuclei (predicted to be relatively stable) in a sea of instability.

116 and 118, using the cold fusion reaction ^{208}Pb(^{86}Kr, n)^{293}118. Three decay chains were observed with element 116 being assigned as a decay product of ^{293}118. The reported production cross section was ~2 pb, an unexpectedly high value given the extrapolation of the existing data on cold fusion. If confirmed and extended, this work would represent a new reaction pathway to the superheavy nuclei, possibly leading to the synthesis of elements 119, 117, 115, and 113.

SEE ALSO THE FOLLOWING ARTICLES

ACCELERATOR PHYSICS AND ENGINEERING • ACTINIDE ELEMENTS • COHERENT CONTROL OF CHEMICAL REACTIONS • DOSIMETRY • FISSION REACTOR PHYSICS • ISOTOPES, SEPARATION AND APPLICATIONS • NUCLEAR PHYSICS • NUCLEAR REACTOR THEORY • PARTICLE PHYSICS, ELEMENTARY • RADIOACTIVITY

BIBLIOGRAPHY

Choppin, G. R. (1964). "Nuclei and Radioactivity," Benjamin, New York.
Choppin, G. R., and Rydberg, J. (1980). "Nuclear Chemistry, Theory and Applications," Pergamon, New York.
Cohen, B. L. (1971). "Concepts of Nuclear Physics," McGraw-Hill, New York.
De Shalit, A., and Feshbach, H. (1974). "Theoretical Nuclear Physics," Wiley, New York.
Enge, H. A. (1966). "Introduction to Nuclear Physics," Addison-Wesley, Reading, MA.
Friedlander, G., Kennedy, J. W., Macias, E. S., and Miller, J. M. (1981). "Nuclear and Radiochemistry," 3rd ed. Wiley, New York.
Harvey, B. G. (1964). "Introduction to Nuclear Chemistry and Physics," 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey.
Harvey, B. G. (1965). "Nuclear Chemistry," Prentice-Hall, Englewood Cliffs, NJ.
Marmier, P., and Sheldon, E. (1970). "Physics of Nuclei and Particles," Vols. 1 and 2. Academic Press, New York.
Segre, E. (1977). "Nuclei and Particles," 2nd ed. Benjamin, Reading, MA.



Nuclear Physics

Christopher R. Gould
North Carolina State University

E. David Davis
University of Kuwait and North Carolina State University

Lawrence Wilets
University of Washington

Philip J. Siemens
Oregon State University

I. Twentieth-Century History
II. The Lightest Nuclei
III. Gross Properties of Nuclei
IV. Nuclear Decays and Limits of Metastability
V. Motion of Nucleons in Nuclei
VI. Collective Modes
VII. Nuclear Reactions
VIII. Energy Sources
IX. Nuclear Forces

GLOSSARY

Atomic mass unit (AMU) $1/12$ the mass of a neutral atom of ^{12}C, equal to 931.5 MeV.

Atomic number (Z) Integer equal to the number of protons in a nucleus; the order of an element in the periodic table.

Atomic weight or nuclear mass number (A) An integer equal to the sum of the number of protons Z and neutrons N.

Isobar Member of a set of nuclides of the same atomic weight A.

Isomer Nuclide in a long-lived state of energy excitation.

Isospin Set of quantum numbers relating to the charge of elementary particles. For the nucleon, the total isospin is $T = \frac{1}{2}$, and the third component is $T_3 = +\frac{1}{2}$ for the proton and $T_3 = -\frac{1}{2}$ for the neutron. For the pion, the total isospin is $T = 1$ and the third components are $T_3 = 1, 0, -1$ for the π^+, π^0, π^-.

Isotone Member of a set of nuclides of the same neutron number N and various Z.

Isotope Elements of the same atomic number Z characterized by various N.

Nuclear magneton ($\mathbf{m_N}$) $e\hbar/2m_pc$.

Nuclide Nucleus characterized by Z and N.

Parity Symmetry of wave function under inversion of the coordinate system: $\mathbf{r} \rightarrow -\mathbf{r}$. The wave function either remains unchanged (even, or $+$, parity) or changes sign (odd, or $-$, parity).

Spin Angular momentum in a (nuclear) state, measured

in units of $\hbar$. The spin can be either an integer (A even) or a half-odd integer (A odd).

THE ATOM consists of a central, massive core, the nucleus, surrounded by an electron cloud. A nucleus is characteristically 1/10,000 the size of the electron cloud and 4000 times as massive. A nucleus contains Z protons and N neutrons. Protons and neutrons are collectively called nucleons. The number of electrons in a neutral atom is equal to the number of protons, and this is also known as the atomic number, the order in which an atom appears in the periodic table of the elements. The sum of the number of protons and neutrons is denoted by A, the mass number. The nuclide of an element of chemical name X is denoted by ${}^{A}_{Z}\mathrm{X}_{N}$, or simply by ${}^{A}\mathrm{X}$ since Z and N can be deduced from X and A. In contrast to atoms, where the size of the electron cloud changes very slowly with atomic number Z, nuclear density is essentially the same for all nuclei, hence the nuclear radius increases as $A^{1/3}$. Neutrons and protons are bound together in nuclei by strong forces, mediated by intermediate-mass elementary particles called mesons. The protons and neutrons themselves possess a substructure consisting of three quarks bound together by gluons. In nature, all elements up to uranium ($Z = 92$) are to be found, with the exception of technetium ($Z = 43$) and promethium ($Z = 61$), which have been produced in the laboratory. Elements up to $Z = 118$ have been produced artificially and identified, but elements 110–118 have not yet been named. Many naturally occurring and artificially produced nuclides are unstable (metastable) and exhibit radioactivity by decay (transmutation) through the emission of electrons (beta particles), alpha particles (helium nuclei), or fission. Nuclear reactions can be induced in the laboratory by bombardment of targets with a wide variety of projectiles: gamma rays, electrons, elementary particles, and other nuclei. The heaviest nuclei can release energy by fissioning into intermediate-weight fragments; the lightest nuclei can release energy by fusion. The latter process is the source of power production in the sun and stars.

I. TWENTIETH-CENTURY HISTORY

The foundations of both modern nuclear physics and modern atomic physics were established by Ernest (Lord) Rutherford through a series of celebrated experiments first published in 1911. He used alpha particles from naturally radioactive emitters as projectiles to bombard a variety of targets, and he detected the scattered alpha particles by visually observing scintillations on a phosphorescent screen. From the distribution of scattered particles, he was able to demonstrate that the interaction of alpha particles with atoms obeyed Coulomb's inverse-square law down to distances on the order of 10^{-13} m = 100 fm (1 fm = 10^{-15} m).

The picture that emerged from Rutherford's experiments was that of an atom consisting of a massive core—the nucleus—of positive electric charge Ze, where $-e$ is the charge on the electron and Z is the atomic number. The nucleus is surrounded by a negatively charged electron gas. Earlier atomic theories fell, most notably J. J. Thomson's model of electrons embedded in a positively charged "jelly." In 1913, Niels Bohr announced his atomic theory of electrons circling the nucleus in quantized planetary orbits. Further studies in atomic physics led to the discovery (invention) of quantum mechanics by Werner Heisenberg (1925) and Erwin Schrödinger (1926).

The discovery of the neutron by James Chadwick in 1932 clarified both the problem of isotopic composition and the connection between atomic weight A and nuclear spin. With protons and neutrons now known to be the building blocks of nuclei, the study of nuclear structure was launched.

In 1935, Hideki Yukawa postulated the existence of a new, intermediate-weight elementary particle, which he called the mesotron, to act as the agent to bind neutrons and protons together in the nucleus. Some confusion ensued when Carl Anderson and Seth Neddermeyer discovered a candidate particle in 1938 that did not seem to interact strongly with nuclei. The problem was resolved in 1947 by Cecil Powell and collaborators who identified two particles, the mu and the pi mesons, the latter being the Yukawa mesotron (now called the pion); the mu meson, or muon, is the Anderson–Neddermeyer particle. This was a remarkable triumph of speculative theoretical induction. It also completed the first phase in the microscopic description of nuclear structure. Subsequently, a host of elementary particles has been found, many of which play important roles in nuclear physics. (See Section VII.B.)

The discovery of fission by Otto Hahn and Fritz Strassmann in 1939 led to the development of the atomic (more properly nuclear) bomb during World War II and the attendant development of fission reactors for electrical power generation. The fusion process, which is the mechanism by which the sun and stars generate their energy, was the basis for development of the hydrogen bomb in the 1950s, and there has been intense research during the subsequent decades to harness thermonuclear fusion as a power source. At the same time, nuclear physics and chemistry have provided radioactive isotopes, radioactive and stable isotope identification techniques, nuclear magnetic resonance, etc. for medical diagnosis and treatment, geological and archaeological dating, tracing of water and

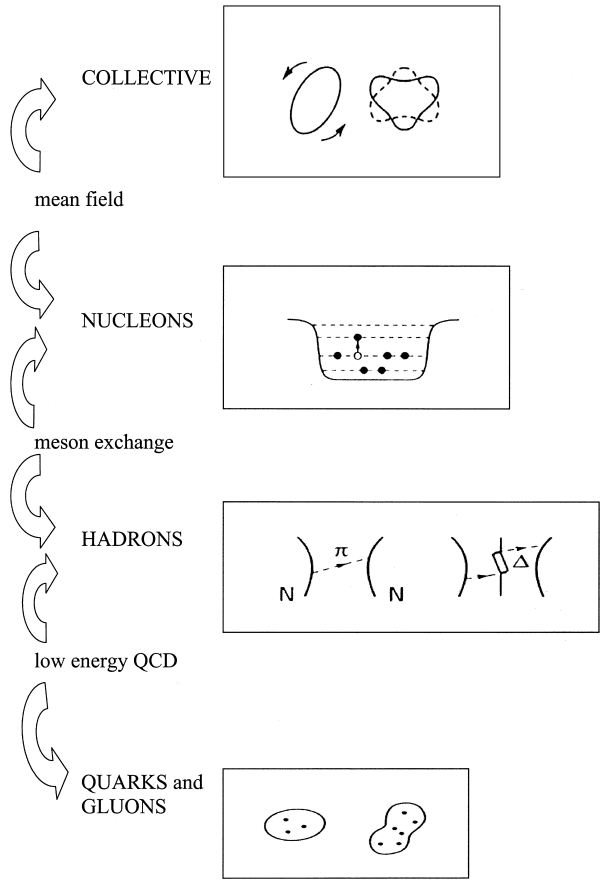


FIGURE 1 Structure of nuclei revealed by projectiles of low to high energy (from top to bottom), probing shorter length scales as the energy increases.

atmospheric flow patterns, planetary and solar system histories, and numerous other applications.

At present, much interest is being concentrated on nuclear substructure, namely, the constituents of the protons, neutrons, and other particles previously considered to be elementary. The subparticles are called quarks; the proton and the neutron each contain three quarks.

Figure 1 summarizes, from top to bottom, the historical evolution of nuclear physics and nuclear phenomena studied with particle accelerations. At the lowest energies (longest length scales), the collective modes of nuclei—rotations and vibrations—are evident. As the energy increases (shorter length scales) the presence of individual nucleons in shell model orbits is revealed, the nucleons themselves interacting via the exchange of mesons. At the highest energies (shortest length scales), the quark and gluon structure of the nucleons is observed. The theory of nucleons interacting via the exchange of mesons is called quantum hadrodynamics (QHD). The theory of quark–gluon interactions is called quantum chromodynamics (QCD). Linking these descriptions of nuclear phenomena is a major challenge for theoretical physics.

II. THE LIGHTEST NUCLEI

A. The Proton

The simplest nucleus is ${}^{1}_{1}\mathrm{H}_{0}$. It has the following characteristics: mass, $m_{\mathrm{p}} = 938.27$ MeV $= 1836\, m_{\mathrm{e}}$; angular momentum $J = \frac{1}{2}\hbar$; magnetic moment $\mu = 2.7928\mu_{\mathrm{N}}$ [$1\ \mu_{\mathrm{N}}$ (nuclear magneton) $= e\hbar/2m_{\mathrm{p}}c$]; root-mean-square charge radius $\langle r^2\rangle^{1/2} = 0.82$ fm.

B. The Neutron

A constituent of nuclei, but not itself a nucleus because it does not form an atom, is the neutron. Its characteristics are $m_{\mathrm{n}} = 939.57$ MeV; $J = \frac{1}{2}\hbar$; $\mu = -1.9130\ \mu_{\mathrm{N}}$; $\langle r^2\rangle = -0.12$ fm^2.

The neutron is heavier than the proton by 1.30 MeV. It decays into a proton, an electron ($m_{\mathrm{e}} = 0.511$ MeV), and an antineutrino with a mean life of 887 sec. The mass of the electron-neutrino or antineutrino has been measured to be less than 3 eV.

C. Light Nuclei

The isotopes of hydrogen have been given special names: deuterium for D$\equiv{}^{2}$H and tritium for T$\equiv{}^{3}$H. The corresponding nuclides are called the deuteron and the triton, respectively. The deuteron is a loosely bound structure (on the nuclear scale), having a binding energy of 2.2 MeV and a root-mean-square charge radius of 2.8 fm. The triton is unstable and decays into ^{3}He, accompanied by the emission of an electron and of an antineutrino with a half-life of 12.33 years. Although ^{3}H is more tightly bound than ^{3}He, the decay occurs because the neutron is heavier than the proton, and ^{3}H is heavier than ^{3}He.

There is no bound state of two neutrons (the dineutron) or two protons (^{2}He).

Helium ($Z = 2$) also comes in two stable isotopes, ^{3}He and ^{4}He. ^{4}He is especially tightly bound, and its central density is the highest of any nucleus. Unstable isotopes of He have been identified through ^{10}He.

There are no stable isotopes of any element with mass number $A = 5$ or 8.

III. GROSS PROPERTIES OF NUCLEI

A. Nuclear Sizes and Shapes

Nuclear density is remarkably constant with respect to Z and A. This leads to the statement thatnuclear volumes

are proportional to A and that the radii are proportional to $A^{1/3}$.

$$R = 1.2\ A^{1/3}\ \text{fm} \tag{1}$$

The electron and muon (which behaves like a heavy electron) are ideal probes for exploring the charge distributions of nuclei. Both have no measured structure of their own (they are point particles) and interact with nuclei only through the electromagnetic field. (The weak interaction, which is responsible for beta decay, is quite negligible here.) Charge distribution experiments are mainly of two classes: atomic spectra and scattering.

Common atoms, of course, contain electrons, and atomic isotope shifts have yielded information on the differences in charge distributions among various isotopes. The muon can also be captured in the electric field of a nucleus to form a hydrogen-like atom. Since the muon is 207 times as massive as the electron, its Bohr orbits are 1/207 times the size of the corresponding electron orbits. Thus, muons can probe nuclear charge distributions more deeply than electrons.

While the density in the interior of a large nucleus is nearly uniform (0.17 nucleons/fm^3), it falls off smoothly at the surface, dropping from 90% to 10% of the interior density over a distance of about 2.5 fm. The approximate constancy of the central density is one aspect of the phenomenon known as *nuclear saturation*.

More precise analyses of the experiments yield detailed differences between nuclides. Most nuclei are not spherical but have intrinsic nonspherical shapes, or distortions. Some nuclei execute oscillations about a spherical shape while others exhibit a permanent intrinsic deformation. Prolate (football-shaped) quadrupole deformations are found for a large number of nuclei. Octupole (pear-shaped) and higher-order deformations have also been observed.

B. Nuclear Masses

From the Einstein relationship $E = mc^2$ there is a basic equivalence between mass and energy, and the two terms are used interchageably: both are frequently measured in the same units, mega-electron-volts (MeV). Atomic masses quoted for nuclides are the masses for the corresponding neutral atoms. Nuclear binding energies can be identified by equating the atomic mass to the following:

$$\begin{aligned} E(Z, N) &= M(Z, N)c^2 \\ &= m_{\text{n}}c^2 N + m_{\text{p}}c^2 Z + m_{\text{e}}c^2 Z - B_{\text{e}}(Z) \\ &\quad - B_{\text{N}}(Z, N), \end{aligned} \tag{2}$$

where the rest-mass energies of the constituent neutrons, protons, and electrons have been explicitly removed. The term B_{e} is the binding energy associated with the atomic electrons. The nuclear physics is contained in the term $B_{\text{N}}(Z, N)$. The minus sign associated with B_{N} is a matter of convention. The condition $B_{\text{N}} > 0$ corresponds to binding, a lowering of the total energy of the system. The lower the energy, the more stable is the nuclide.

Figure 2 shows the distribution of values for B_{N}/A as a function of A. (For any A there are usually several nuclides, or isobars.) The plot shows a maximum around $A = 56$–62. The most stable common nuclide is ^{56}Fe. However, one of the rarer isotopes of nickel ^{62}Ni is

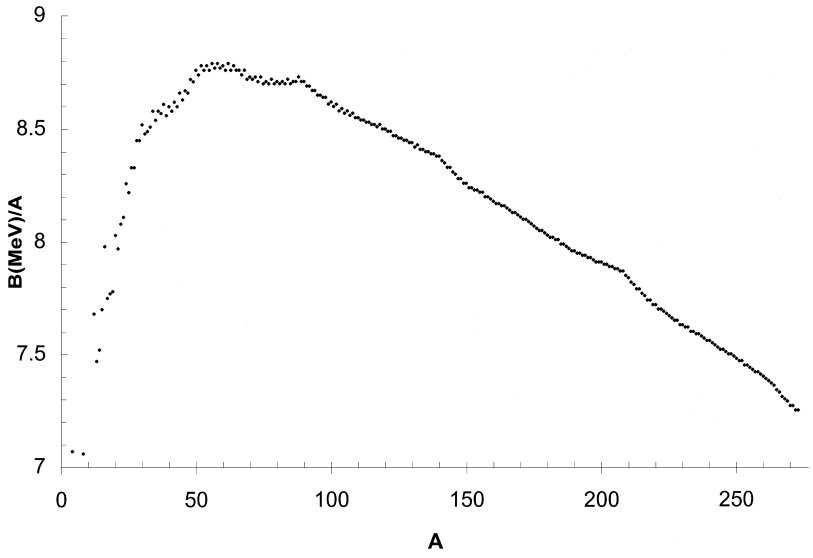


FIGURE 2 Curve of binding energy per nucleon as a function of mass number A. [Reference: Firestone, Richard (private communication).]

the most stable nuclide of all. The shallow bumps around $A = 90$, 140 and 210 are indicative of shell effects.

A physically useful parameterization of nuclear binding energies is given by the von Weizsäcker semiempirical mass formula, inspired by analogy to a classical liquid drop. It reads as follows:

$$-B_{\mathrm{N}}(Z, N) = C_{\mathrm{v}}A + C_{\mathrm{s}}A^{2/3} + C_{\mathrm{sy}}\frac{(N-Z)^2}{A} + C_{\mathrm{C}}\frac{Z^2}{A^{1/3}} + \delta_{\mathrm{p}}(Z, N) + \Delta(Z, N). \quad (3)$$

The various terms have the following interpretation and values:

$C_{\mathrm{v}}A$ is the volume energy; $C_{\mathrm{v}} = -15.68$ MeV would be the energy per nucleon in an infinite nucleus with equal numbers of neutrons and protons if there were no electrostatic (Coulomb) repulsion between the protons. The existence of this term is another manifestation of nuclear saturation.

$C_{\mathrm{s}}A^{2/3}$ is called the surface energy. It represents the surface tension constant times the surface area; $C_{\mathrm{s}} = +18.56$ MeV, and the + sign indicates a loss of binding due to the surface.

$C_{\mathrm{sy}}(N-Z)^2/A$ is the volume symmetry energy. Nuclear matter (in the absence of Coulomb forces) is most strongly bound for symmetric ($N = Z$) nuclei. Deviation from symmetry results in a loss of binding. $C_{\mathrm{sy}} = 28.1$ MeV.

$C_{\mathrm{C}}Z^2/A^{1/3}$ is the Coulomb energy corresponding to a uniformly charged sphere of radius $R \propto A^{1/3}$. The numerical value is $C_{\mathrm{C}} = 0.717$ MeV.

$\delta_{\mathrm{p}}(Z, N)$ is called the pairing energy. In their ground states, even–even (even Z, even N) nuclei are more strongly bound than even–odd or odd–even (i.e., odd A) nuclei, and odd–odd nuclei are less strongly bound yet. This can be approximated by the formula

$$\delta_{\mathrm{p}}(Z, N) \approx \frac{34}{\mathrm{A}^{3/4}}\ \mathrm{MeV}\begin{Bmatrix} +1 & \text{for odd–odd} \\ 0 & \text{for } A \text{ odd} \\ -1 & \text{for even–even} \end{Bmatrix}. \quad (4)$$

$\Delta(Z, N)$ contains further details of nuclear structure, especially of what is known as shell structure. In practice, the parameters associated with the preceding terms are fit to the masses of all known nuclides by the method of least-square deviation; then $\Delta(Z, N)$ is the residual. (Some analyses have included explicit shell correction terms in the semiempirical mass formula.) When the residuals $\Delta(Z, N)$ are plotted against either Z (for various N) or against N (for various Z), the result is a sawtooth curve (or band) that decreases monotonically, with breaks at certain "magic" numbers. The breaks at the magic numbers are characteristically a few mega-electron-volts. Nuclei with magic numbers of protons or neutrons are especially stable, and nuclei with both proton and neutron numbers magic are exceptionally stable.

A better indicator of magic numbers is to be found in the nucleon separation energy, the energy required to remove a neutron or a proton [see Eqs. (12) and (13)]. This becomes large as one approaches a magic number from below, and decreases just above a magic number.

The following numbers are magic for both neutrons and protons: 2, 8, 20, 28, 50, and 82. The value $N = 126$ is also magic. The next higher magic numbers have not yet been observed, but theoretical calculations indicate that $Z = 114$ and $N = 184$ may be magic.

The terminology "magic" is antiquated but colorful. It arose historically because stability of certain proton numbers (irrespective of N) and certain neutron numbers (irrespective of Z) was unexpected. These numbers are now fully understood in terms of the closing of shells in an independent-particle model, quite analogous to the closing of electron shells in atoms at the noble gases with $Z = 2,\ 10,\ 18,\ 36,\ 54$, and 86.

IV. NUCLEAR DECAYS AND LIMITS OF METASTABILITY

A. Beta Decay

The elementary β^- decay reaction

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e} + \bar{\nu} \quad (5)$$

can proceed for the free neutron because the neutron mass is greater than the sum of the masses of the proton and electron combined; the neutrino and antineutrino are massless, or very nearly so. The same process can proceed in nuclei if it is energetically possible. Furthermore, β^+ (positron) decay

$$\mathrm{p} \rightarrow \mathrm{n} + \mathrm{e}^+ + \nu \quad (6)$$

or atomic electron capture

$$\mathrm{p} + \mathrm{e} \rightarrow \mathrm{n} + \nu \quad (7)$$

can also occur in nuclei if energetically possible.

The energy requirement for β^- decay [Eq. (4)] is given in terms of the atomic energies (masses),

$$E(Z, N) - E(Z+1, N-1) \equiv Q_{\beta^-} > 0. \quad (8a)$$

For electron capture it is

$$E(Z, N) - E(Z-1, N+1) \equiv Q_{\mathrm{ec}} > 0. \quad (8b)$$

The rest mass of the electron or positron created in the process is included in the atomic energy. For β^+ decay, however, the condition is

$$E(Z, N) - E(Z-1, N+1) - 2m_{\mathrm{e}}c^2 \equiv Q_{\beta^+} > 0, \quad (8c)$$

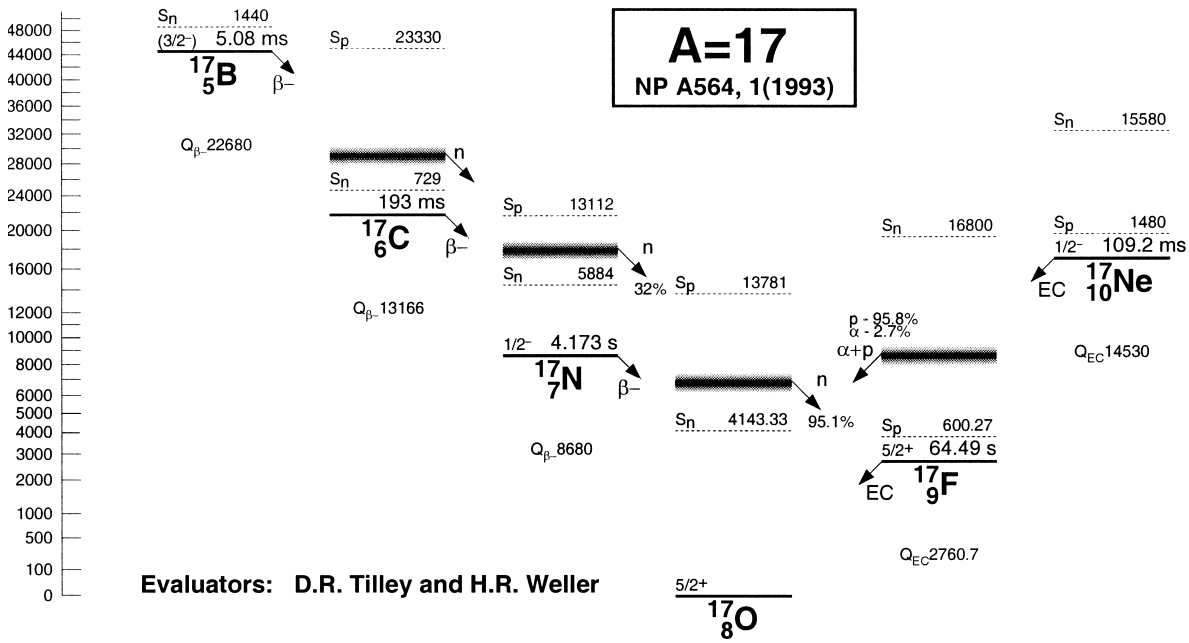


FIGURE 3 Decay properties of nuclei of odd mass $A = 17$. The parabolic shape of the relative binding energies characterizes the valley of stability. [Reference: Firestone, Richard. LBNL Isotopes Project Nuclear Structures Home Page. 25 Jan. 2001 http://ie.lbl.gov/systematics.html.]

where the two electron masses come from keeping track of the number of electrons in the neutral atoms. If β^+ decay is energetically possible, so also is electron capture, but the converse may not be true.

For fixed A, the lowest energy isobar defines the valley of beta stability, (see Fig. 3). Because of the Coulomb energy, the heavier nuclei have $N > Z$. According to the semiempirical mass formula, the equation for the stable valley can be expressed most simply for Z as a function of A:

$$Z = \frac{\frac{1}{2}A}{1 + 0.0064A^{2/3}} \tag{9}$$

An example of the energetics of beta decay is shown in Fig. 3 for a string of odd-A isobars. Note that there is only one stable member of the string; this is generally true of odd-A isobars. The situation is quite different for even-A isobars, as shown in Fig. 4. The odd–odd nuclides lie higher in energy than the even–even ones by twice the pairing energy $2\delta_p$. Several even–even isobars can be stable against beta decay. The odd–odd nuclides can almost always beta decay, in some cases by either β^- or β^+ from the same nuclide. There are no stable odd–odd nuclei heavier than ^{14}N. It is energetically possible for some even–even nuclei to decay by the emission of two electrons (+ or −). Such decays are very slow, and after a long search, double beta decay with emission of two neutrinos has finally been observed for ^{82}Se with a half-life of 10^{20} years;^{76}Ge with a half-life of 10^{21} years and of ^{100}Mo with a half-life of 10^{19} years.

Beta-decay half-lives depend sensitively on the energy release Q, decreasing rapidly with increasing Q, and on details of the nuclear structure. In general, the half-life increases rapidly with increasing change in the nuclear spins between the mother and the daughter. The half-lives tend to decrease as one moves along an isobaric string away from the stable valley.

A search for double beta decay with the emission of no neutrinos is of great interest. Observing the process would prove the neutrino was its own antiparticle (a Majorana neutrino), and would confirm the neutrino rest mass was nonzero, as indicated by observations of neutrino deficits from the sun and from cosmic ray processes.

B. Alpha Decay

Alpha-particle decay is a common phenomenon among heavy nuclei, and all nuclei heavier than ^{209}Bi can decay by α-particle emission (although other modes may dominate). Even when the energy available, Q_α, is positive, the decay is inhibited by the Coulomb barrier

$$2Ze^2/r - Q_\alpha, \qquad r > R \tag{10}$$

which must be penetrated. Such penetration is forbidden in classical mechanics but is possible quantum-mechanically. The inhibition factor depends very sensitively on Q_α, decreasing rapidly with decreasing energy. Although the height of the Coulomb potential at the nuclear surface increases with Z, the energy release increases

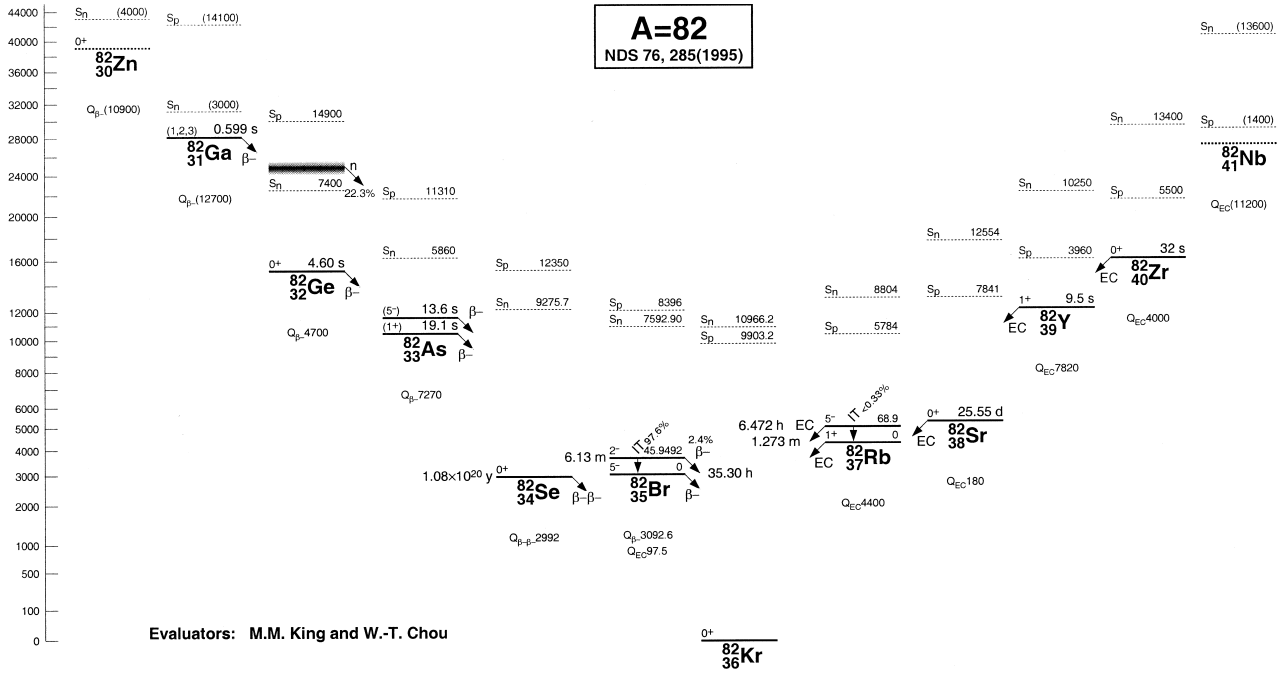


FIGURE 4 Decay properties of nuclei of even mass $A = 82$. Odd–odd nuclei lie on a separate parabola from even–even nuclei, shifted by the pairing energy. [Reference: Firestone, Richard (op. cit.).]

more rapidly, and the lifetimes in general decrease rapidly with Z. This leads to one of the limits of stability discussed in Section IV.D.

C. Spontaneous Fission

Within the concept of the liquid drop model of the nucleus, there is competition between the surface tension, which tends to stabilize the droplet in a spherical shape, and the electrostatic (Coulomb) repulsion, which tends to disrupt the system. A measure of this competition is given by the Bohr–Wheeler fissionability parameter,

$$x = \frac{E_{\text{Coulomb}}}{2E_{\text{surface}}} \approx \frac{Z^2/A}{50} \tag{11}$$

The spherical configuration of a nucleus with $x > 1$ is unstable against deformation along the path which leads path to fission. If $x < 1$, the spherical shape is locally stable, but fission is still energetically allowable for $x > 0.35$ ($x = 0.35$ corresponds to $Z \approx 35$). In the range $0.35 < x < 1$, the process can only proceed by way of quantum-mechanical barrier penetration, so that spontaneous fission half-lives increase rapidly with decreasing x. Odd-A nuclei have a higher barrier against fission than do even–even nuclei. For ^{238}U, $x \approx 0.71$, spontaneous fission does not play a significant role in its radioactivity, which is dominated by alpha decay. In fact, for almost all known nuclei, alpha decay tends to dominate spontaneous fission. However, spontaneous fission eventually is a limiting factor in how high in Z one can go in producing new elements.

D. Limits of Metastability

For the following discussion, please refer to Fig. 5, where β-stable species are plotted for N and Z.

The stable and long-lived nuclides cluster along the valley of beta stability. As one moves away from the valley on either side, the beta decay rates become faster (i.e., shorter half-lives).

To the lower right side of the valley is the neutron-rich region. The energy required to remove a neutron, called the separation energy, is defined by

$$S_n = B_N(Z, N) - B_N(Z, N-1). \tag{12}$$

The separation energy decreases as one moves away from the stable valley. When $S_n < 0$, a neutron can be emitted spontaneously in a time comparable with the transit time for a neutron inside the nucleus, which is on the order of 10^{-22}–10^{-21} sec. The line $S_n = 0$ is called the neutron drip line. Beyond this line, nuclides do not live long enough even to be called metastable.

A similar situation occurs on the proton-rich (upper left) side of the stable valley. For protons, as for alpha particles, there is a Coulomb barrier to be surmounted even when the emission is energetically allowable. Barrier

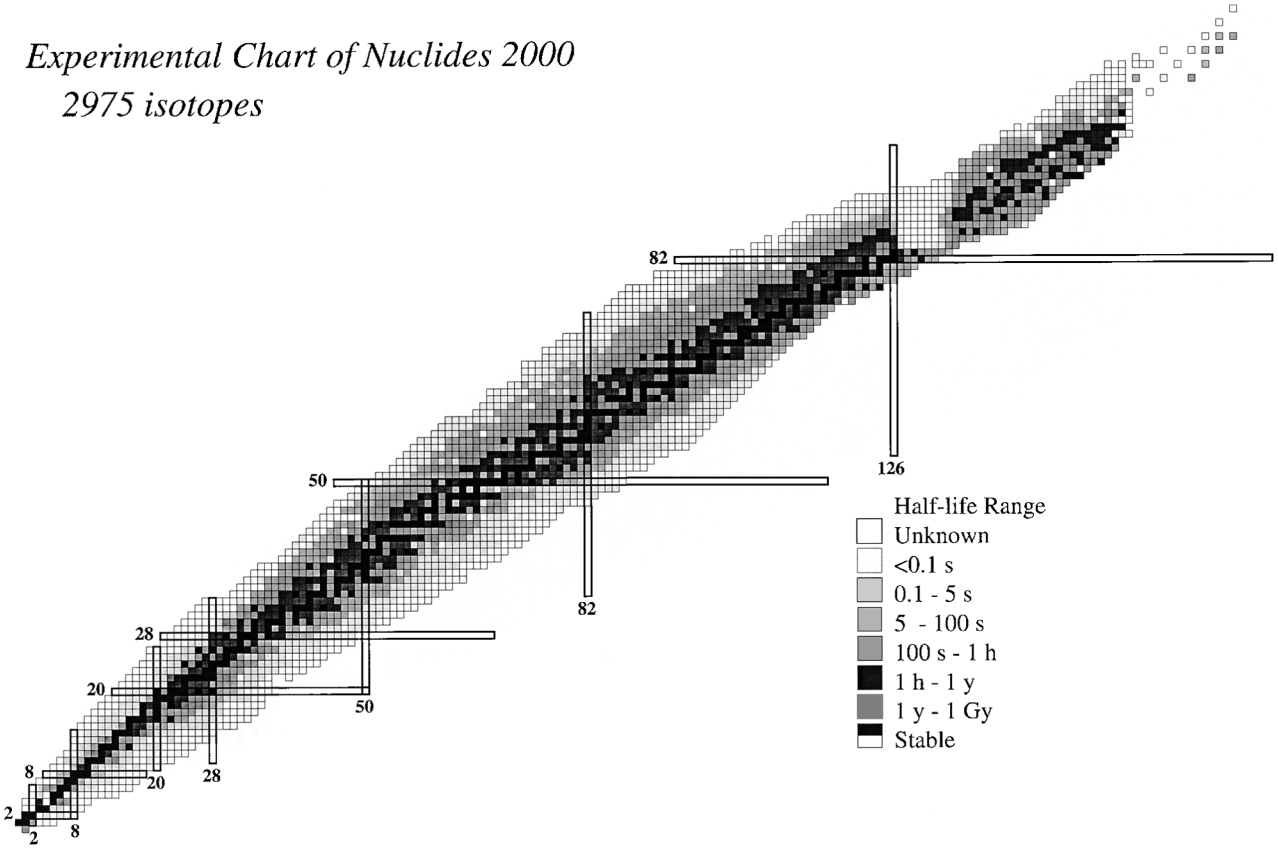


FIGURE 5 Experimental chart of the half lives of the 2975 known nuclides. The proton and neutron drip lives are indicated by the light shading (N horizontal, Z vertical). [Reference: Firestone, Richard (op. cit.).]

penetration is easier for protons than for alpha particles because they have half the charge and one-fourth of the mass. The vanishing of the proton separation energy

$$S_{\mathrm{p}} = B_{\mathrm{N}}(Z, N) - B_{\mathrm{N}}(Z-1, N) \qquad (13)$$

represents a practical limit to metastability, although not as severe as for neutron emission.

The corresponding limit for alpha particles occurs when an alpha particle in a nucleus has an energy greater than its Coulomb barrier, as discussed in Section IV.C. The limit imposed by spontaneous fission is $x = Z^2/50A = 1$.

There is speculation, backed by theoretical calculations, that another region of (meta)stability should occur in the region of closed nucleon shells at $Z = 114$ and $N = 184$. The search for such an "island of stability" has been a topic of intense investigation. (The terminology "island" is poetic, and requires plotting the negative of the atomic masses in order to implement the metaphor. Then the valley of beta stability becomes the "ridge of stability"; beyond lies the "sea of instability." The challenge is to cross the sea from the ridge to the island.)

V. MOTION OF NUCLEONS IN NUCLEI

We have seen that the picture of the nucleus as a droplet of fluid is useful for understanding the gross properties of nuclei, and we shall see in Section VI that the analogy extends to the dynamics of nuclear shapes as well. But nuclear matter is a very special kind of fluid, a quantum liquid in which the nucleons move about freely inside, as molecules in a gas, even though they are held together by short-range forces, like the molecules of a liquid. The reason nucleons can move over long distances is explained by the Pauli principle, which prevents a nucleon from shifting into an orbit already occupied by another nucleon. Thus, even though nucleons are constantly pushing on each other, the forces between them cannot change their motion, since all of the lowest-energy orbits are already occupied. The ability of each nucleon to move almost undisturbed through the nucleus, due to quantum mechanics, gives rise to several characteristic features of nuclei which they share with other quantum liquids, such as liquid helium and electrons in metals. We have already mentioned the shell structure related to magic numbers,

which is analogous to the band structure of electrons in metals. Another striking feature of the quantum nature of nucleonic motion is the fact that nuclei are superconductors, which is the source of the pairing energy $\delta_p(Z, N)$ and also has an important influence on nuclear shapes.

The concept of nucleons moving freely inside a nucleus is one similar to that of electrons moving freely inside an atom. It is called the *independent particle model*. Each nucleon experiences the same potential and moves in an orbital described by a wave function which can be found by solving the Schrödinger equation.

A. Spherical Nuclei

For spherical nuclei, the potential energy of a nucleon due to the other nucleons can be approximated by

$$V(r) = V_c(r) + \gamma \frac{1}{r}\frac{dV_c(r)}{dr}\mathbf{l}\cdot\mathbf{s} \tag{14}$$

with V_c of the Wood–Saxon form

$$V_c(r) = \frac{V_0}{1 + e^{(r-R_v)/a_v}}. \tag{15}$$

The term $\mathbf{l}\cdot\mathbf{s}$ is the spin-orbit operator; γ is a constant equal to a number between 5 and 10. (In atomic physics, $\gamma = -1$.) The protons experience a Coulomb potential as well.

The form of the central potential is similar to that of the charge distribution. The strength is $V_0 \approx -50$ MeV. The size parameter R_v is a few tenths of a fermi greater than the radius parameter for the charge distribution; the surface thickness is about the same.

An orbital is characterized by the quantum numbers $nljm$, where n is the radial quantum number, l is the orbital angular momentum, j is the total (spin plus orbital) angular momentum, and m is the projection of total angular momentum onto (say) some z-axis. The Pauli principle demands that no two identical particles occupy the same orbital. The lowest state of a nucleus is generally obtained by filling the lowest-energy orbitals. The orbitals are degenerate (equal in energy) for different m. Since m can assume the values $-j, -j+1, \ldots +j$, there are $2j+1$ orbitals of the same energy.

The effect of the spin orbit term is to lower states with $j = l + \frac{1}{2}$ with respect to states with $j = l - \frac{1}{2}$ (same l).

A crude but useful estimate of single particle energies and wave functions can be obtained by approximating $V_c(r)$ with a harmonic oscillator potential, $V_c(r) \approx \frac{1}{2}kr^2 + V_0$. The corresponding oscillator frequency would be $\omega = \sqrt{k/m}$, where m is the mass of the neutron or proton. $2\pi/\omega$ is the round-trip transit time of a nucleon in any orbit (in this approximation), which is approximately $A^{1/3} \times 10^{-22}$ sec.

In a schematic rather than rigorous way, one can plot the energy levels for the protons and neutrons in any spherical nucleus on a single diagram. The result corresponds to $\varepsilon_2 = 0$ in Fig. 6. Normally the orbitals are filled, one per nucleon, in the order of increasing energy, up to the number of protons or neutrons in the nucleus. The circled numbers denote the number of orbitals up to the position of the circle, taking into account the $2j+1$ degeneracy. Note that $2j+1$ is an even integer. There are gaps in the spectrum, corresponding to shell closures.

The attraction of the nucleon–nucleon force encourrages like nucleons to pair together in orbitals whose total angular momenta add to zero. This is the basis of nuclear superconductivity, see Section V.C. All ground states of even–even nuclei have total angular momentum 0 and positive parity, 0^+ in the notation J^π. Odd-A nuclei have $J^\pi = j^\pi$, where j^π is the angular momentum and parity of the odd nucleon in the unfilled shell.

B. Deformed Nuclei

Non-closed-shell nuclei often prefer nonspherical shapes. When several nucleons are added to a closed-shell nucleus, they can gain an energy advantage by clustering near each other in preferred regions, which become the elongated ends of a football-shaped nucleus. The bulges in the ends of the football provide an added energy advantage due to quantum mechanics: the extra space allows the nucleons to attain a lower momentum according to the uncertainty principle. As a result, nuclei with half-filled shells are about half again as long as they are thick, like an egg. This nonspherical shape causes these nuclei to have large electric (and material) quadrupole moments. They are larger, by a factor of between 5 and 20, than the moments contributed by a single proton.

In the case of deformed nuclei, the independent particle potential which each nucleon experiences is described by a nonspherical, usually (but not necessarily) axially symmetric potential. In fact, the states can no longer be properly described by the quantum numbers l and j, since the resultant wave functions are linear combinations of states of different l and j. Parity and m, however, remain "good quantum numbers." Furthermore, the single-particle energies are no longer degenerate with respect to m, but (for axial symmetry) there is a degeneracy with respect to the sign of m. The ground states of deformed even–even nuclei are still 0^+, but for odd-A nuclei one finds $J^\pi = m^\pi$, where m^π refers to the last odd nucleon (see Fig. 6).

An understanding of how the nucleus's energy depends on its shape is afforded by the Strutinsky shell correction method: The main part of the energy is given by the liquid drop model. The sum of single particle energies yields a correction associated with nuclear shell structure.

H-6

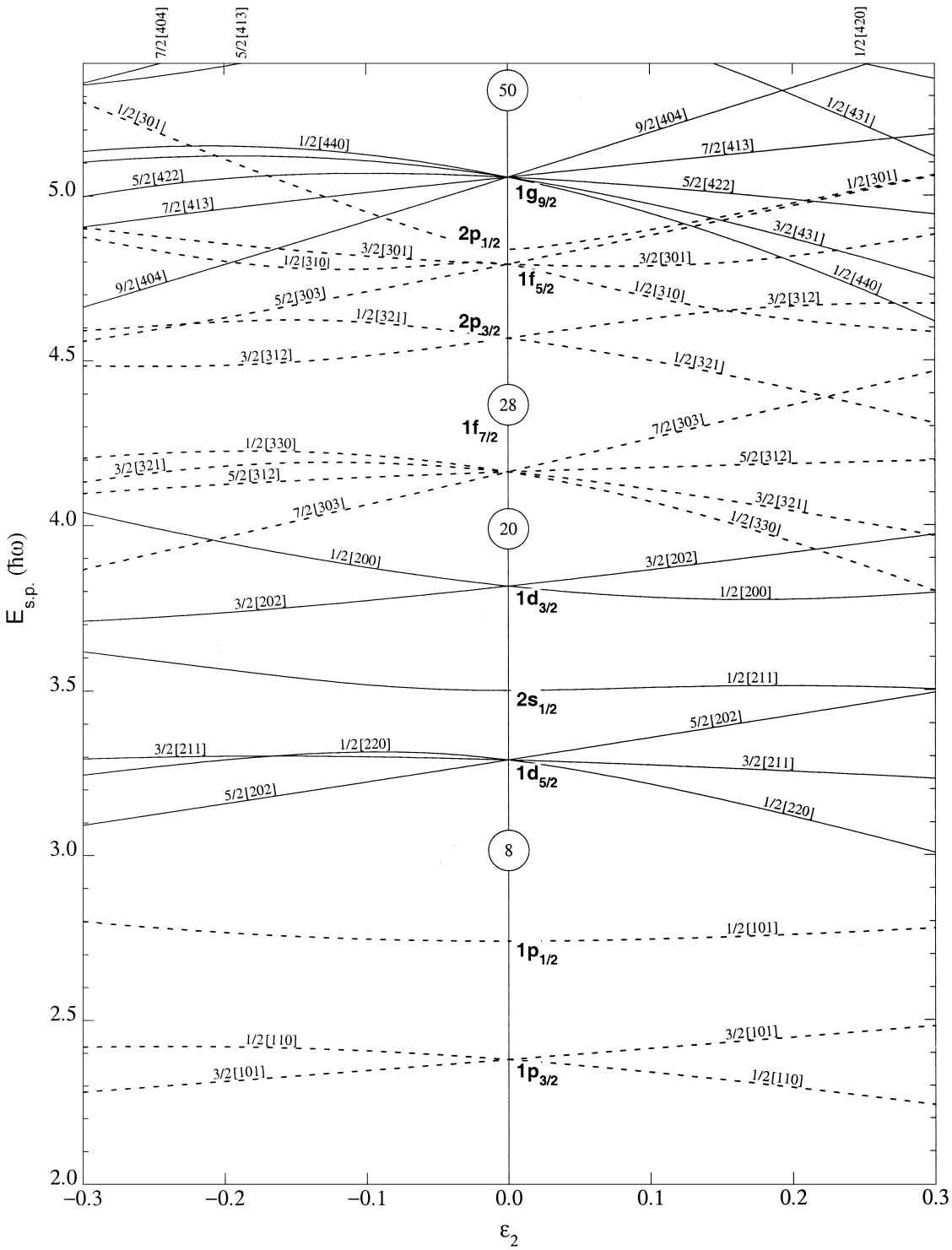


FIGURE 6 Single particle energy levels in the Nilsson model for Z or N < 50. Quadrupole deformations correspond to $e_2 \neq 0$. For spherical nuclei, $e_2 = 0$, and levels are characterized by n, l, j, m. For $e_2 \neq 0$, the m-degeneracy is lifted and levels are characterized by m and by the parity $\pi = (-1)^1$. [Reference: Firestone, Richard (op. cit.).]

In order to utilize both features, the single particle energies are summed over occupied orbitals; subtracted from this sum is a smoothed average so that the resulting shell corrections as a function of deformation fluctuate about zero. To this is added the liquid drop (or some more sophisticated) energy of deformation. Minima in the total energy of deformation correspond to shapes that are stable or metastable. A measure of the distortion is the deformation parameter defined by

$$\varepsilon = \frac{3(c-a)}{(2c+a)}, \tag{16}$$

where a and c are the major axes of the ellipsoid of revolution; c is along the symmetry axis and a perpendicular to the symmetry axis.

Superdeformations, corresponding to c/a approaching 2, are realized in experiment. This second energy minimum, which is common in fissile nuclei, corresponds to superdeformation; it is also realized in intermediate-weight nuclei (e.g., Ce, Dy, and Hg) in high spin states. Stabilization of the second minimum is aided, in the case of heavy nuclei, by the Coulomb repulsion which stretches the nucleus; for intermediate nuclei, the centrifugal force due to rotation in high spin states produces stretching.

These simple considerations of independent-particle motion would imply that all non-closed-shell nuclei should be nonspherical, since there will always be some shape that best fits the occupied single-particle orbitals. Instead, many nuclei with N and/or Z close to the magic numbers are also spherical. The reason for this lies in the phenomenon of pairing, or nuclear superconductivity.

C. Superconducting Nuclei

Superconductivity is a special type of long-range order which occurs in the lowest-energy states of quantum liquids made up of particles which obey the Pauli principle, such as electrons or nucleons. The short-range attractive force between nucleons (see Section IX) makes nucleons want to be near each other. The best way for two nucleons to stay as close as they can is to move in orbits which are identical to each other but moving in opposite directions. This allows the nucleons the best chance to meet frequently. In an axially symmetric nucleus, the orbitals are occupied in pairs, the members of a pair corresponding to time-reversed motion with opposite angular momentum projection m. In a spherical nucleus the chances can be further improved by the nucleons hopping together from orbit to orbit with different m values. Two nucleons moving together in this way are called a pair. The ability to form pairs is the reason for the pairing energy $\delta_{\mathrm{p}}(Z, N)$. The formation of pairs is also the source of electrical superconductivity in some metals, but nuclei are too small for their electrical superconductivity to be seen directly.

The advantage of the spherical shape for forming pairs competes with the advantage of the football shape for non-closed-shell nuclei. The advantage of the football shape increases as more particles are added outside of a closed shell, because all of the added particles benefit from each others' potential energy and the extra room at the ends of the football. This is why nuclei with half-filled shells are deformed while many nuclei with nearly closed shells, or only a few nucleons beyond closed shells, remain spherical.

The formation of pairs has many other consequences for nuclear structure. A useful way to understand the role of the pairs is the Interacting Boson Model, discussed in Section VI.E.

VI. COLLECTIVE MODES

A. Shape Vibrations

Nuclei are nearly always found in nature with the shape that has the minimum possible energy for a given N and Z. The shape of the nucleus can be changed by adding energy, for example, by pushing on a nucleus with the electric field of a charged projectile passing nearby. Once disturbed, the shape of the nucleus will oscillate about its equilibrium spherical equilibrium configuration. The oscillations consist of pure harmonic vibrations, the normal modes. Their frequencies depend on the energy associated with the corresponding shape changes. Just as for the equilibrium shapes discussed in Section V.B, the energy of a shape change comes mostly from the change in energy of the nucleons' orbitals. As a result, the frequencies of the normal-mode vibrations are similar to the frequency Ω_m with which individual nucleons bounce back and forth, namely, about $A^{1/3} \times 10^{-22}$ sec.

All but the lightest nuclei have many normal modes of shape vibration. Typically, for most nuclei the lowest normal mode corresponds to elongation (prolate quadrupole), and its frequency is likely to be between $\frac{1}{2}\Omega_m$ and $\frac{1}{6}\Omega_m$. In some nuclei, for example, ^{208}Pb, the lowest frequency normal mode, corresponds to a pear shape. For spherical nuclei, the normal modes can be characterized by their angular momentum J. The elongation (prolate quadrupole) mode had $J = 2$. The pear-shape, or octupole, mode has $J = 3$. For deformed nuclei, in addition to the elongation mode, there is another quadrupole mode that corresponds to squeezing the middle of the nucleus, which breaks its axial symmetry.

B. Equilibrium Deformations and Rotations

In regions between closed-shell configurations, nucleonic orbits (Section V.B) stabilize the nuclei at nonspherical

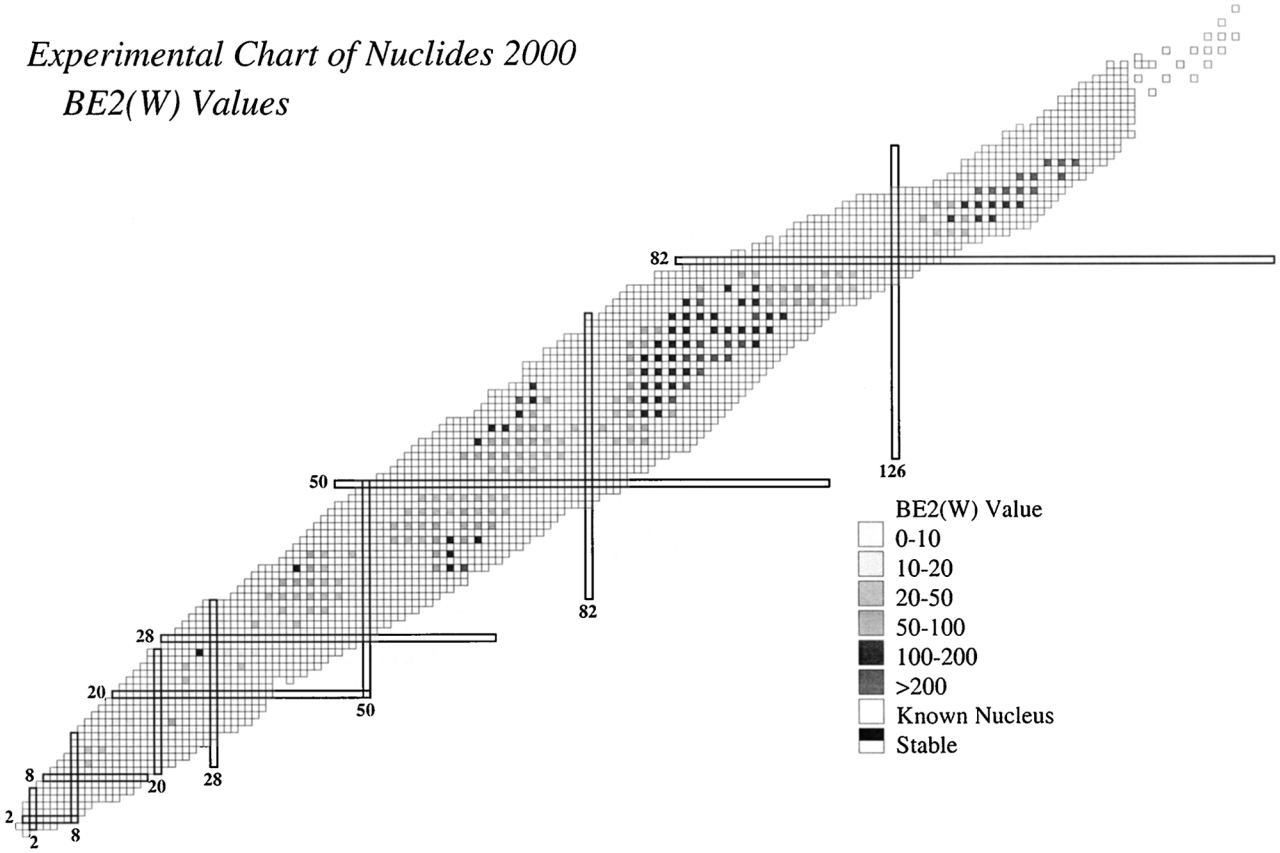


FIGURE 7 Chart of nuclides (Z vertical, N horizontal) showing regions of enhanced electric quadrupole gamma ray decay strength. These regions are characteristic of large nuclear deformations which occur away from closed shell nuclei. [Reference: Firestone, Richard (op. cit.).]

shapes. The overwhelming preference is for a prolate spheroidal (i.e., football) shape. Large permanent deformations occur with great regularity in certain regions of Z and N, (see Fig. 7). Near the stable valley, these are $19 < A < 25$, $150 < A < 185$, and $220 < A$. The two heavier regions correspond, by coincidence, with the chemical rare-earth regions. Large deformations begin abruptly at $N = 90$ (near A of 150) and depend more strongly on N than on Z in the first rare-earth region. Similarly, large deformations begin again at $Z = 88$ (near A of 220) and depend more strongly on Z in the second rare-earth region. Other regions of large deformation occur among the metastable nuclei away from the stable valley.

Permanent deformations lie in the range $0.1 \lesssim \varepsilon \lesssim 0.4$. The identification of such shapes is based on (1) electrical quadrupole moments, first observed in atomic hyperfine studies, (2) electrical quadrupole transitions induced by electric fields from scattered charged particles (especially α particles), (3) the nature of the rotational spectra, (4) beta and gamma transition rates, (5) atomic isotope shifts, (6) spectra of muonic atoms, etc.

Regions of "superdeformation," corresponding to $\varepsilon \approx 0.6$, however, do occur in fissile nuclei (Section VI.C) and in intermediate nuclei of high spin (I up to 60). An understanding of the origin of these is presented in Section V.B.

The lowest mode of excitation of a strongly deformed nucleus is rotational. For axially symmetric shapes, the moment of inertia is very small about an axis of symmetry, and rotation occurs about an axis perpendicular to the symmetry axis, analogous to a rotating dumbbell or diatomic molecule. The rotational motion is then nearly decoupled from the other collective modes. The excitation energy follows the law

$$E_{\text{Rot}} = \hbar^2/2\mathcal{J}[I(I+1) - I_0(I_0+1)] \qquad (19)$$

plus corrections for high spin. All band members have the same parity. Here I_0 is the spin of the lowest member of the band. The spin I assumes the values I_0, I_0+1, I_0+2, ... except that for even–even nuclei one has $I = 0, 2, 4, \ldots$. (For even–even nuclei, the ground-state spin and parity are always 0^+.) The moment of inertia $\mathcal{J}$ is greater than

that of an irrotational fluid, but less than the solid body value.

Rotational energies are small compared with vibrational or particle excitations. The first rotational excitation energy ranges from 50 keV in heavy nuclei to several hundreds of kilo-electron-volts for light nuclei. Rotational bands have been observed with spins as high as 60 in Dy.

Vibrational excitations can also be found on permanently deformed nuclei. Collective excitations for "transition" nuclei (neither good vibrators nor rotators) are more complex.

Nuclei also occur with equilibrium octupole deformations. When compounded with a quadrupole deformation, the nuclei assume a pear-shaped surface deformation. Such have been identified in the even–even nuclei around $_{88}$Ra. An indicator of such shapes is the existence of low-frequency "tunneling" vibration corresponding to inversion of the shape. As a result, even–even nuclei exhibit a band of odd spins, $I = 1, 3, 5, \ldots$, with odd parity, displaced from the even-spin band by the tunneling frequency, which is on the order of 100 keV.

C. Fission

As noted in Section IV.C, nuclei for which the fissionability parameter x is less than unity are stable against small deformation. Nevertheless, the energy release Q_f is positive for nuclei down to $x = 0.35$. The fissile nuclei, thorium and heavier, have $Q_f \gtrsim 190$ MeV. This large amount of energy can be identified with the electrostatic energy of the fragments at the point of scission, $Z_1 Z_2 e^2/(R_1 + R_2)$; during the process of separation the potential energy is converted into kinetic energy of the fragments, neutron emission, and gamma emission.

The potential energy-of-deformation barrier against fission is about 6 MeV for the even–even isotopes of uranium. If this much energy is deposited, the nucleus undergoes rapid fission. This can be achieved by neutron capture on an odd-A isotope (e.g., $^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{236}\mathrm{U}^* \rightarrow$ fission, where the asterisk indicates a state of energy excitation), by gamma absorption, or by any number of nuclear reactions. It is the emission of neutrons ($\gtrsim$2.5) during the fission process that makes possible a sustained chain reaction.

Mass division during the fission process is usually very asymmetric. The two fragments that emerge have a distribution of masses, and division into equal mass fragments is highly unlikely at low excitation energy. Figure 8 shows the mass distribution yields for thermal neutron fission of ^{235}U. With increasing excitation energy, the probability of symmetric mass division increases.

A frequent feature of fission is shape isomerism. It plays an important role in understanding the fission process and in testing nuclear models. For uranium and heavier nuclei, the energy of deformation plotted against (say) the quadrupole deformation parameter along the path from spherical to scission generally is characterized by two peaks and two minima. The lowest point corresponds to the most stable (ground-state) configuration, which is usually nonspherical. There is another local minimum at larger deformation that is about 3 MeV higher. This shape can also be identified with superdeformation (see Section V.B). During nuclear reactions, some fraction of the events results in populating the second minimum. Because of the smaller barrier to fission (compared with the ground state), either spontaneous fission or decay to the ground state can occur with reduced, but measurable, half-life. The observed half-lives range from 10^{-3} to 10^{-9} sec.

D. Isospin and Spin Modes

The most dramatic example of "isospin" waves is the giant dipole resonance. This collective state corresponds to an oscillation separating neutrons from protons, while still maintaining rather constant total nuclear density. The energy of excitation follows the approximate formula

$$E = \hbar\omega \simeq 78 A^{-1/3} \text{ MeV} \tag{20}$$

for $A > 60$. This mode can be excited by various mechanisms, such as gamma ray absorption. The resonance is quite broad, with a full width at half-maximum for gamma-ray excitation cross sections varying from 3 to 10 MeV.

Higher multipoles of various types are possible and have been studied. As an example, nucleons of (say) spin up can oscillate collectively against nucleons of spin down.

E. Interacting Boson Model

A particularly effective and simple picture for describing collective models is provided by the interacting boson model (IBM). Nucleons are spin $\frac{1}{2}$ particles and obey Fermi–Dirac statistics, but a bound pair of nucleons has integer spin and obeys Bose–Einstein statistics at distances large compared with the size of the structure. The basis of the IBM is the strong, short-range attraction between n-n and p-p pairs in states of angular momentum $J = 0^+, 2^+, \ldots$, in order of descending binding energy.

The IBM assumes that the attraction between n-n and p-p pairs in heavy nuclei makes such pairs very correlated and hence behave like bosons. For most purposes, only the most bound pairs, $J = 0$ (s) and $J = 2$(d) pairs are needed to explain the low-lying quadrupole collective modes of even–even nuclei, although $J = 4$ (g) pairs are needed for more detailed agreement. In order to understand cluster states and octupole excitations, $J = 1$ (p) and $J = 3$ (f) bosons have been included; for light nuclei, n-p pairs have also been included. For odd-A nuclei, the odd, unpaired

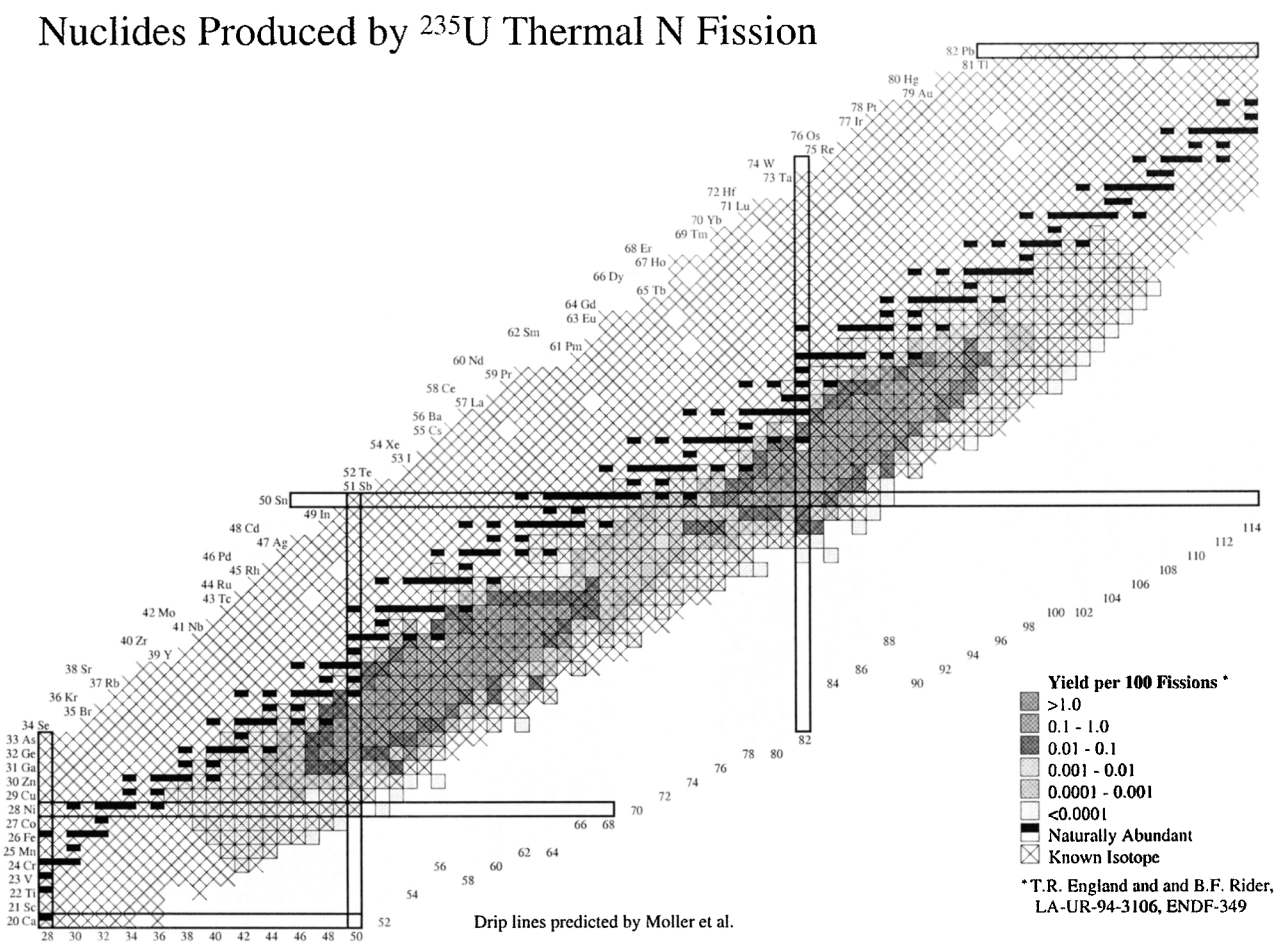


FIGURE 8 The asymmetric distributions of nuclides produced by thermal neutron fission of ^{235}U. [Reference: Firestone, Richard (op. cit.).]

nucleon interacts with the bosons, while for odd–odd nuclei there is one neutron and one proton interacting with the bosons.

The utility of the model lies in the relatively small number of empirically adjustable parameters, the simplicity of calculation (which involves diagonalization of moderately sized matrices), and success in describing low energy, collective properties of nuclei. In particular, the model has given a unified description of the transition from spherical to deformed nuclei and has highlighted the importance of collective motion of the transitional nuclei for which the intrinsic collective motion does not have axial symmetry. Thus, the model is economical for fitting and predicting experimental data.

VII. NUCLEAR REACTIONS

In addition to natural decay processes, nuclear reactions can be induced by any of the elementary particles, by gamma rays, and by other nuclei. Through various reaction mechanisms, nuclei can be excited or transmuted. The resultant products can be analyzed to obtain an understanding of the physics or utilized for practical applications. The following subsections deal with a small part of the wealth of nuclear reactions that have been studied.

A. Nucleon Scattering

Protons and neutrons of energies ranging from a fraction of an electron volt (for neutrons) to many billions of electron volts (for protons) have been used as projectiles to bombard nuclei. The simplest process is that of elastic scattering, where the projectile is deflected but the target nucleus is left undisturbed. The process can be described mathematically by considering the projectile to move in a potential of the form given by Eq. (14). Except at low energies, however, some of the events involve inelastic processes, such as energy excitation followed by decay, or some kind of transmutation. These can be included by

making the potential complex (i.e., containing an imaginary part as well as a real part). The imaginary part of the potential removes probability from the elastic "channel," and the resultant wave function describes only the elastic process. This description is called the optical model in analogy with the scattering of light from a diffractive, absorptive medium.

Nonelastic scattering refers to any process that leads to a different energy state of the target (or projectile, if it is composite) or any rearrangement of the constituents. A special case is inelastic scattering, which retains the composition of the projectile and target but leaves one or the other in an excited state.

It is frequently useful to separate the reaction process into an initial, "direct" reaction, followed by the subsequent decay of the excited intermediate state. If the nucleus passes through complicated intermediate states, one speaks of "compound nucleus" formation and subsequent decay of the compound nucleus.

Strong forward peaked angular distributions and a smooth dependence on the energy of the bombarding particles characterize direct reactions. Relatively isotropic angular distributions but very sharp energy dependence due to the formation of resonance states characterize compound nucleus reactions. The probability that a reaction will occur is parameterized by the cross section, typically measured in barns ($1\text{b} = 10^{-24}$ cm^2) and representations of the effective cross sectional area of the target nucleus. Compound nuclear cross sections can exceed 10^3 b or more, but more typically total nuclear cross sections are of order the nuclear size $\sim \pi R^2$.

The standard notation for nucleus reactions is T(p, p′)T′ for $p + T \rightarrow T' + p'$ or sometimes simply (p, p′). Here p is the incident projectile and T the target; the primed quantities are the corresponding objects at the end of the reaction. In the above, the projectile p is generic; it can be a proton (p), neutron (n), deuteron (d), etc. Direct nucleon-induced reactions include the following: (a) charge exchange, (n, p) or (p, n); (b) capture, $p + {}_Z X_N \rightarrow {}_{Z+1}X^*_N$ or $n + {}_Z X_N \rightarrow {}_Z X^*_{N+1}$; (c) pickup, (p, d) or (n, d).

The final reaction products can include emitted gamma rays (photons), beta particles, further nucleons, alpha particles, fragments, fission, etc.

B. Other Particles

Gamma rays, electrons and positrons, muons, pions and other mesons, and antiprotons have been used as projectiles to bombard nuclei. Each is unique in probing different aspects of nuclear structure: Electrons and muons, as discussed in Section III.A, can be used to explore nuclear charge distributions because they interact as point particles and only through the electromagnetic field. Photons, of course, also interact only electromagnetically, and deposit a discrete amount of energy and momentum upon absorption. Pions interact with nuclei through the strong force; they can be absorbed (depositing the large rest mass of the pion as energy of excitation), scattered (with target excitation), or undergo charge exchange ($\pi^+ + {}_Z X_N \rightarrow \pi^0 + {}_{Z+1}X_{N-1}$, and similarly for the other charged pions) or, more rarely, double charge exchange ($\pi^+ + {}_Z X_N \rightarrow \pi^- + {}_{Z+2}X_{N-2}$, etc.).

Antiprotons interact strongly with nuclei and also annihilate easily. Therefore, they explore the periphery of nuclei. Like other negatively charged particles, they can also form hydrogen-like atoms with nuclei. From the spectrum of gamma rays produced through atomic transitions, information on their interaction with nuclei can be deduced.

C. Nuclei as Projectiles

Nuclei of all masses have been used as projectiles. A heavy nucleus can deposit much more angular momentum and energy than a light projectile. Thus, nuclear projectiles, or heavy ions, are used to study nuclei under extreme conditions. One goal for which heavy ions have been used is to try to make nuclei with exceptionally large masses, the "superheavy nuclei" predicted by many nuclear models. To make these very massive, very fragile objects, the projectile is given as little energy as possible. A certain minimum energy, the Coulomb barrier, is necessary to overcome the electrostatic repulsion of the nuclei. Even this minimal energy has so far proved too disruptive to allow the observation of a superheavy reaction product. The limits of angular momentum and energy that a nucleus can hold are probed when the projectile energy is just a little above the Coulomb barrier.

An interesting and novel process is the collision between two heavy nuclei at energies below the Coulomb barrier, so that the nuclear forces between the objects do not act. Although the two nuclei are close together for only a very short time, it is still long compared with electron transit times and the system behaves as a quasimolecule or a quasi-atom. The high total electric charge of the nuclei can give rise to exotic atomic processes, including spontaneous positron emission and the formation of a "charged vacuum."

If the projectile's velocity when it touches the target nucleus is slow compared to the speed of the nucleons' motion inside the nuclei, many nucleons may be transferred between the target and projectile nuclei. The transferred nucleons carry the projectile's energy and angular momentum to the target; nucleons transferred from the target to the projectile slow the projectile. These processes give rise to a frictional force between the projectile and the target,

and provide a remarkable example of irreversible motion which is usually associated with macroscopic-sized objects. The result is that the target and projectile are heated to very high temperatures, typically 10 to 30 billion K. They also spin very fast. If too much angular momentum is deposited, the target spins apart into two or more large fragments in a special kind of fission driven by the centrifugal force of the nuclear rotation. This fission process limits the ability of heavy projectiles to fuse into very heavy nuclei.

If the speed of the projectile nucleus is similar to the speeds of the nucleons inside it, the reaction becomes quite complicated. Nuclear collisions in this energy regime are the subject of current research and are not yet understood.

When the speed of the projectile becomes larger than the speeds of the nucleons inside it, the reactions simplify again. Some of the projectile nucleons-never notice that the others have hit the target nucleus; similarly, part of the target remains almost undisturbed. These spectator regions emerge from the reaction as smaller nuclei, with the target fragment nearly at rest and the projectile fragment moving with almost the beam velocity. The remaining matter participates in a violent collision in which the temperature may exceed a trillion Kelvin. Of course the hot matter thus formed explodes into many small pieces, mostly neutrons and protons with a few deuterons and other light nuclei. Many pions are also formed. Such collisions have been observed in laboratories since the mid-1970s and are used to study the properties of matter under conditions otherwise known only in supernovae or in the first millisecond of the Big Bang.

At extremely high projectile energies, it is hoped that the temperatures achieved in nuclear collisions may become high enough to cause a radical change in the composition of the nuclear matter. The quarks from which the nucleons are made may become unstuck from each other and form a plasma of freely moving quarks and gluons (see Section IX.D). This state is called a plasma because the free motion of quarks is like the independent motion of electrons and ions in a high-temperature gas or conducting liquid. This program of experiments is in a preliminary phase, using accelerators adapted from particle-physics research. It is not yet known whether the quark plasma can be created; the possibility remains that the nuclei may pass through each other without depositing enough energy to liberate the quarks.

VIII. ENERGY SOURCES

Nuclear reactions are an important source of energy in nature and in technology. Radioactive substances of moderate half-lives have been used, especially in space missions, as a source of energy to operate low-power equipment. More important are the large energy yields of fusion and fission reactions, which are outlined briefly below. Both reactions are exothermic because the most stable nuclei are of intermediate mass, which is the region of maximum binding energy.

A. Fusion

In the early stages of stellar evolution, stars are fueled by the fusion of hydrogen nuclei. This is the case with our own sun. The process requires high temperatures in order to bring the electrically charged nuclei close enough together to interact. In fact, fusion in stars is generally a very slow process, fed by the relatively few particles in the high-velocity tail of the thermal distribution. Two main mechanisms are distinguished in solar hydrogen burning: the direct p + p reaction, and the catalytic C–N–O cycle. Both lead to the overall result

$$4\,^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 2\mathrm{e}^{+} + 2\nu_{\mathrm{e}}. \tag{23}$$

The energy release, including the energy carried off by the neutrinos is 25.7 MeV, or about 0.7% of the rest mass of the four hydrogen atoms.

Fusion is also the source of energy in hydrogen bombs and has been studied as a source of controlled thermonuclear energy generation. A reaction that occurs with high probability and is therefore promising for a power source is

$$^{2}\mathrm{H} + {}^{3}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + \mathrm{n} \tag{24}$$

with an energy release of 17.6 MeV. The specific energy (energy per unit mass of reactant) release in this process is nearly five times that of fission.

The technological problem in controlled fusion is the production of a high-temperature plasma at high density for a sustained period of time. Actually, "high density" here may only be a tiny fraction of 1 atm and confinement times may be only a small fraction of a second. Various techniques are currently under study: magnetic confinement, inertial implosion by laser or particle beams, and muon catalysis.

B. Fission Reactors

The energy release during fission is approximately 210–219 MeV. Coulomb energy of the fragments at separation (scission) accounts for about 80% of this energy, and this is primarily converted into the kinetic energy of the fragments. The remainder is released mainly in the form of gamma rays, neutrons, beta rays, and neutrinos. The specific energy release is about 2 million times that of coal or oil.

Even–even heavy, fissile nuclei have a fission threshold of about 6 MeV. This energy can be injected by the capture of a (slow) neutron on the next lighter even–odd isotope. The number of neutrons released per fission event is about 2.5 for slow neutrons on ^{235}U. The multiplication of neutrons provides the mechanism for a sustained fission reaction: the neutrons emitted can be used to produce more fissions. This is called a chain reaction.

Neutron-capture cross sections at low energies are proportional to the inverse of the neutron velocity. Thus, low-energy neutrons are more effectively captured than high-energy neutrons. However, natural uranium contains only 0.72% of the isotope ^{235}U, and most neutrons would be captured onto ^{238}U and lost to the chain. A moderating substance, ^{2}H or ^{12}C, is often introduced to slow the neutrons before they interact and to take advantage of the fact that slow neutrons have a much higher cross section for fission on ^{235}U than for capture on ^{238}U. Nevertheless, such slow, or moderated, fission reactors require a higher enrichment of ^{235}U.

An alternative to the ^{235}U-enriched reactor is the breeder reactor, which can be either slow (moderated) or fast. Since neutron capture on ^{238}U or ^{232}Th leads, after beta decay, to the fissile nuclei ^{239}Pu or ^{233}U, the captured neutrons are not "wasted" but are utilized to breed new fissile nuclei. However, at least two neutrons are required for each breeding: one for the initial fission reaction, and one for each conversion of a nonfissile nucleus into a fissile one. Of the 2.5 neutrons per fission, 0.5 still remain to be wasted by escape from the vessel or nonuseful absorption. The world supply of ^{238}U and ^{232}Th is, for such purposes, essentially inexhaustible.

IX. NUCLEAR FORCES

A. General Features

Nuclear forces are, to a high degree of accuracy, charge independent. That is to say, except for the explicit electromagnetic part, the neutron–neutron, neutron–proton, and proton–proton interactions are equal—when compared in the same state.

The nucleon–nucleon force is primarily central: the force acts along the line joining the nucleons, and the corresponding potential is a function only of the distance of separation of the nucleons. However, noncentral components are significant, and the interaction also depends on the relative orbital angular momentum and the relative orientation of the two nucleon spins. This dependence dominates when the nucleons are far apart. In fact, the interaction contains all of the complications allowed by the fundamental symmetries of nature.

The central two-body potential decreases more rapidly than the r^{-1} Coulomb form. The long and intermediate range part of the potential is negative and attractive. This is responsible for nuclear binding. The short range part is positive and strongly repulsive; it is sometimes approximated by a hard (infinite) core of radius 0.4–0.5 fm. The strong repulsion plays a crucial role in nuclear saturation.

B. Meson Theory

The "strong" force between nucleons is mediated by the exchange of various mesons, which are so named because they are (usually) intermediate in mass between electrons and nucleons. In a first approximation, the form of the potential energy of interaction is that given by Yukawa,

$$f^2 e^{-\mu r}/r$$

where $\mu = mc/\hbar$ is the inverse of the range of the interaction and m is the mass of the exchanged meson. At short distances ($\mu r \ll 1$), the potential varies as r^{-1}, just as the Coulomb potential. However, the exponential factor causes a more rapid decrease with distance.

The lightest meson is the pion, originally conjectured by Yukawa. The pion comes in three charges: $+$, 0, $-$. Its mass is about one-seventh that of the nucleon, giving a range of interaction of $\mu_\pi^{-1} = 1.4$ fm. The strength of the interaction is $f^2 \approx 0.08\,\hbar c$, which is about 11 times that of the Coulomb interaction at short distances. The pion contribution to nuclear forces dominates at large distances, $r \gtrsim 1$ fm, but is relatively weak compared with the shorter range contributions.

The intermediate range attractive region (0.5 fm $\lesssim r \lesssim$ 1.0 fm) is governed by the exchange of two pions. The short range, repulsive region ($r \lesssim 0.5$ fm) is dominated by the exchange of heavier mesons, especially the rho and omega mesons, which are $5\frac{1}{2}$ times the mass of the pion and hence give a range of 0.25 fm.

It is, however, more appropriate to describe the short-range part of the nucleon–nucleon interaction in the framework of nucleonic substructure. This is an area of current investigation.

C. Subnuclear Structure

The fundamental theory of the strong interactions is quantum chromodynamics (QCD). The particles of the theory are quarks, which are spin-$\frac{1}{2}$ fermions and come in six known *flavors*, named *up*, *down*, *strange*, *charm*, *top*, and *bottom*. Quarks carry electrical charge in multiples of one-third of the electronic charge. They also carry another quantum number called *color*, and associated with color is color charge. There are three colors, conventionally

denoted by the three primary colors red, green, and blue. Antiquarks carry anticolor (e.g., anti-red, anti-green, and anti-blue). The quarks interact with a force field, the quanta of which are called gluons.

The gluon field is very similar to the electromagnetic field with the important difference that, unlike the photons of electromagnetism, the gluons carry (color) charge and therefore interact with each other as well as with quarks. This renders the theory nonlinear and very difficult to solve. General features of the theory can be stated, however. These include the following:

1. Physical particles can exist only in combinations of subparticles that have net color-neutral charge, technically called a color-singlet state. Thus, nucleons consist of three quarks, one of each of the primary colors. Mesons contain one quark and one antiquark that carries the anticolor of the quark. Isolated quarks cannot exist. This property is called color confinement.
2. The electric charges associated with the various quarks are such that only integral multiples of the electronic charge are allowed for physical particles.
3. The color analog of the square of the electric charge, called the "strong coupling constant," is not a true constant but depends on the size of the interacting structure (or the magnitude of the momentum transfer). The smaller the size of the interacting structure (the higher the momentum transfer), the weaker is the interaction. This is called asymptotic freedom. Conversely, the larger the separation, the stronger the interaction.

The lightest quarks are the up and down quarks. They have a mass of only a few thousandths of the mass of the nucleons. The up quark has electric charge $+\frac{2}{3}e$; the down quark has electric charge $-\frac{1}{3}e$. A proton consists of two up quarks and one down quark, for a net charge of $+e$; a neutron consists of one up quark and two down quarks, for a net charge of zero. The mass of the nucleons comes from the interaction of the quarks with the gluon field.

The nucleon is described as consisting of a core of three quarks confined within a sphere (sometimes called a "bag") of radius about 0.5–1.0 fm. Surrounding the core is a cloud of pions, which are themselves composed of quark–antiquark pairs, and can carry electric charge. Within the region of the nucleon core, the strong coupling constant has a value of between 1.0 and 2.0. This is to be compared with the corresponding electromagnetic fine structure constant, $\alpha \equiv e^2/\hbar c \approx 1/137$.

SEE ALSO THE FOLLOWING ARTICLES

ATOMIC PHYSICS • FISSION REACTOR PHYSICS • INTERACTING BOSON MODEL • NUCLEAR CHEMISTRY • NUCLEAR FUSION POWER • NUCLEAR REACTOR THEORY • PARTICLE PHYSICS, ELEMENTARY • PLASMA CONFINEMENT • PLASMA SCIENCE AND ENGINEERING • RADIOACTIVITY • STELLAR STRUCTURE AND EVOLUTION • SUPERCONDUCTIVITY

BIBLIOGRAPHY

Bohr, A., and Mottelson, B. R. (1969). "Nuclear Structure," Vol. I. (1974). "Nuclear Structure," Vol. II, Benjamin, New York.

Heyde, K. (2000). "Basic Ideas and Concepts in Nuclear Physics," Institute of Physics Publishing, Bristol.

"Nuclear Physics: The Core of Matter, the Fuel of Stars," (1999). National Academy Press, Washington, DC.



Particle Physics, Elementary

Brian Martin
University College London

Graham Shaw
Manchester University

I. Particles and Interactions
II. Leptons
III. Quarks and Hadrons
IV. Strong Interactions
V. Electroweak Interactions
VI. Experimental Methods
VII. Beyond the Standard Model

GLOSSARY

Antiparticle Partner of a particle with the same mass and spin, but opposite signs of all charge-like quantum numbers.
Baryon Hadronic bound state of three quarks.
Boson Particle with integer spin.
Colour Entity which plays a role in the strong interaction analogous to electric charge in the electromagnetic interaction.
Electromagnetic interaction Force between particles mediated by photons.
Electroweak interaction Interaction corresponding to the theory which combines the weak and electromagnetic interactions.
Elementary particle Basic particle of the standard model, characterized by being point-like without internal structure or excited states.
Fermion Particle with half-integer spin.
Feynman diagram Pictorial technique to illustrate particle interactions.
Flavour Generic name to describe types of leptons and quarks.
Gauge boson Elementary boson which mediates one of the fundamental forces of nature.
Generation Classification of leptons and quarks into families each of which has two members.
Gluon Elementary particle of the standard model which mediates the strong interaction.
Grand unification Attempt to construct theories which unify the strong, weak, and electromagnetic forces into a single interaction.
Hadron A state composed of quarks bound by the strong interaction.
Higgs boson Particle, as yet undetected, responsible for the generation of mass in the standard model.
Lepton Elementary fermion of the standard model without strong interactions.
Meson Hadronic bound state of a quark and an antiquark.
Photon Elementary particle of the standard model which mediates the electromagnetic interaction.

Quantum chromodynamics (QCD) Theory of the strong interaction.

Quantum electrodynamics (QED) Theory of the electromagnetic interaction.

Quark Elementary fermion of the standard model having strong, weak, and electromagnetic interactions.

Standard model Modern theory of particle physics which attempts to explain all phenomena in terms of the properties and interactions of a small number of elementary particles.

Strong interaction Interaction between quarks mediated by gluons.

Supersymmetry Specific unification theory which postulates new partners for all particles of the standard model: bosons to partner existing fermions and fermions to partner existing bosons.

Weak interaction Interaction between quarks and leptons mediated by $W^{\pm}$ and Z^0 bosons.

$W^{\pm}$ and Z^0 bosons Elementary particles in the standard model which mediate the weak interaction.

ELEMENTARY PARTICLE PHYSICS is the study of the fundamental constituents of matter and their interactions. However, which particles are regarded as fundamental has changed with time as our knowledge has improved. When Dalton proposed his atomic theory in 1803, he assumed atoms were indivisible and unchanging. This view prevailed until 1897, when Thomson discovered the first subatomic particles: the negatively charged *electrons* which, for example, carry the electric current in metals. Within atoms, the electrons orbit a tiny, positively charged *nucleus*, as shown by Rutherford in 1911. These nuclei contain almost the whole of the atomic mass and are themselves composed of two types of particle: *protons*, with mass approximately 2000 times that of an electron and a positive electric charge equal in magnitude to that of the electron; and electrically neutral *neutrons*, discovered by Chadwick in 1932, whose mass is approximately equal to that of the proton. Thus by 1933 atoms had been replaced as the fundamental constituents of matter by electrons, protons, and neutrons. Associated with this change of view was an increased precision in the length scales being explored: the radii of atoms are of order 10^{-10} m, while neutrons and protons have radii of order 10^{-15} m. Nor was this the end of the story. As shorter and shorter distances were explored, neutrons and protons were themselves found to have smaller constituents, as we shall see. In addition, new forms of matter and radiation were discovered, which must also be accounted for.

The modern theory of particle physics is called the *standard model*. In the rest of Section I we shall give a brief overview of this model, leaving fuller explanations to the following sections.

I. PARTICLES AND INTERACTIONS

A. The Standard Model

The standard model attempts to explain all phenomena of particle physics, except gravity, in terms of the properties of a limited number of *elementary particles*, where by elementary we mean point particles, without internal structure or excited states. Such a particle is characterized by, among other things, its mass, its electric charge, and its *spin*. The latter is a permanent angular momentum possessed by particles in quantum theory, even when they are at rest. The maximum value of the spin angular momentum about any axis is $sh/2\pi$, where h is Planck's constant and s is called the spin quantum number, or spin for short. It has a fixed value for particles of any given type, for example $s = \frac{1}{2}$ for electrons, and general quantum mechanical principles restrict the possible values of s to be $0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$. Particles with half-integer spin are called *fermions* and those with integer spin are called *bosons*. In the standard model there are three families of elementary particles, called *leptons, quarks*, and *gauge bosons*. Leptons and quarks are spin-$1/2$ fermions, while the gauge bosons have spin-1. In addition, a further spin-0 particle, called the *Higgs boson*, is postulated to explain the origin of mass.

The most familiar example of a lepton is the *electron*, which is bound in atoms by the *electromagnetic interaction*, one of the four fundamental forces of nature. Another well-known example is the *electron neutrino*, which is a light, neutral particle observed in the decay products of some unstable nuclei (so-called β-decays). The force responsible for β-decay is called the *weak interaction*. In addition to leptons, another class of particles, called *hadrons*, is also observed in nature. Neutrons and protons are examples of hadrons, but as we shall see, many other types exist. In the standard model, hadrons are not considered to be elementary, but are made of quarks bound together by the third force of nature, the *strong interaction*. There is an analogy here with nuclei, which are bound states of protons and neutrons held together by a nuclear force, which is a residual effect of the basic strong interaction between their constituent quarks. The theory is unusual in that the quarks themselves are not directly observable, only their bound states.

The strong, weak, and electromagnetic forces are associated with the elementary gauge bosons of the standard model. Consider, for example, the electromagnetic interaction. In classical physics, the interaction between two

charged particles is transmitted by electromagnetic waves, which are continuously emitted and absorbed. This is an adequate description at long distances, but at short distances the quantum nature of the interaction must be taken into account. In quantum theory, the interaction is transmitted discontinuously by the exchange of spin-1 bosons called *photons*. Photons are the gauge bosons, or "force carriers," of the electromagnetic interaction. The weak and strong interactions are also associated with the exchange of spin-1 gauge bosons. For the weak interaction, they are very massive and of two types: the charged W^+ and W^- *bosons* and the neutral Z^0 *bosons*. The equivalent particles for the strong interaction are called *gluons* and have zero mass and electric charge, like the photon.

In addition to the strong, weak, and electromagnetic interactions between quarks and leptons, there is a fourth known fundamental force of nature—*gravity*. There is no consistent quantum theory of gravity at present, although attempts to construct one invariably predict the existence of gauge bosons with spin-2, called *gravitons*, which have never been observed. However, the gravitational interaction between individual elementary particles is negligibly small compared to the other three and we shall ignore it in what follows.

B. Relativity and Antiparticles

Elementary particle physics is often called *high-energy* physics. One reason for this is that if we wish to produce new particles in a collision between two other particles, then because of the well-known relativistic mass–energy relation $E = mc^2$, high energies are needed, at least as great as the rest masses of the particles produced. The second reason is that to explore the structure of a particle requires a probe whose wavelength λ is at least as small as the structure to be explored. By the de Broglie relation $\lambda = h/p$, this implies that the momentum p of the probing particle, and hence its energy, must be large. For example, to explore the internal structure of the proton using electrons requires wavelengths which are much smaller than the radius of the proton, which is roughly 10^{-15} m. This in turn requires electron energies which are greater than 10^3 times the rest energy of the electron, implying electron velocities very close to the speed of light. Hence any explanation of the phenomena of elementary particle physics must take account of the requirements of the theory of special relativity, in addition to those of quantum theory.

Constructing a quantum theory of elementary particles which is consistent with special relativity leads to the conclusion that for each charged particle of nature, whether it is an elementary particle or a hadron, there must exist an associated particle, called an *antiparticle*. This has the same mass as the corresponding particle, but the opposite electric charge. Experimental evidence confirms this important theoretical prediction. If we write the particle as P, then the antiparticle is in general written with a bar over it, i.e., $\bar{P}$. For example, associated with every quark, q, is an antiquark, $\bar{q}$. However, for very common particles the bar is often omitted. Thus, for example, associated with the negatively charged electron e^- is an antiparticle e^+, called the *positron*. In this case the superscript denoting the charge makes explicit the fact that the antiparticle has the opposite electric charge to that of its associated particle. Electric charge is just one example of a so-called *quantum number* (spin, introduced earlier, is another), which is a quantity that characterizes a particle, whether it is elementary or composite. Many quantum numbers differ in sign for particle and antiparticle, and electric charge is an example of this. We will meet other quantum numbers in Sections II and III. When brought together, particle–antiparticle pairs can annihilate each other, releasing their combined rest energy $2mc^2$ as photons or other forms of radiation. Finally, we note that there is symmetry between particles and particles, and it is a convention which is which; for example, we could call the positron the particle, and the electron the antiparticle. That we do not do so merely reflects the fact that the matter around us contains electrons rather than positrons, rather than the other way round.

C. Particle Reactions

Reactions involving elementary particles and/or hadrons are summarized by equations in which the different particles are represented by symbols, which sometimes, but not always, have a superscript to denote their electric charge. In the reaction

$$\nu_e + n \rightarrow e^- + p,$$

for example, an electron neutrino ν_e collides with a neutron n to produce an electron e^- and a proton p; while the equation

$$e^- + p \rightarrow e^- + p$$

represents an electron and proton interacting to give the same particles in the final state, but traveling in different directions. This latter type of reaction, in which the particles remain unchanged, is called *elastic scattering*, while the first reaction is an example of *inelastic scattering*. Collisions between given initial particles do not always lead to the same final state, but can lead to different final states with different probabilities. For example, an electron–positron collision can give rise to elastic scattering

$$e^- + e^+ \rightarrow e^- + e^+$$

or annihilation to give either two or three photons in the final state

$$e^+ + e^- \rightarrow \gamma + \gamma$$
$$\rightarrow \gamma + \gamma + \gamma.$$

Finally, some particles are unstable and spontaneously decay to other, lighter particles. An example of this is the neutron, which decays by the β–decay reaction

$$n \rightarrow p + e^- + \bar{\nu}_e$$

with a mean lifetime of about 900 sec.

D. Feynman Diagrams

Particle reactions, like those previously described, are brought about by the fundamental weak, electromagnetic, and strong forces between the elementary particles involved. A convenient way of illustrating this is to use *Feynman diagrams*. There are mathematical techniques associated with these, which enable them to be used to calculate the quantum mechanical probabilities for given reactions to occur, but in this article we will simply use them as a convenient pictorial description of reaction mechanisms. We first illustrate them for the case of electromagnetic reactions, which arise from the emission and/or absorption of photons.

For example, the dominant interaction between two electrons is due to the exchange of a single photon, which is emitted by one electron and absorbed by the other. This mechanism, which gives rise to the familiar Coulomb interaction at large distances, is illustrated in the Feynman diagram (Fig. 1a). In such diagrams, by convention, the initial particles are shown on the left and the final particles to the right. Spin-1/2 fermions (such as the electron) are drawn as solid lines and photons are drawn as wiggly lines. The arrow heads pointing to the right indicate that the solid lines represent electrons. In the case of photon exchange between two positrons, which is shown in Fig. 1b, the arrow heads on the antiparticle (positron) lines are conventionally shown as pointing to the left.

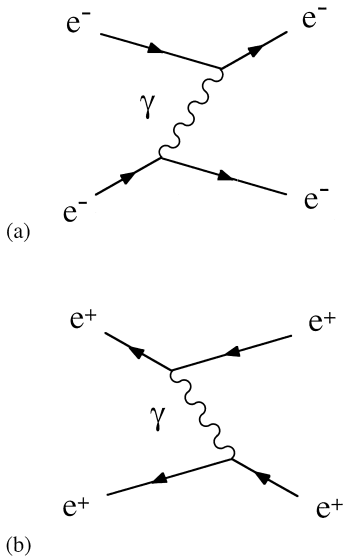


FIGURE 1 Feynman diagrams for (a) electron–electron elastic scattering and (b) positron–positron elastic scattering, via the exchange of a single photon.

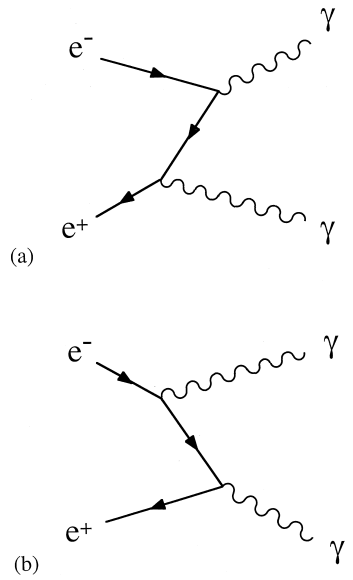


FIGURE 2 The dominant Feynman diagrams for electron–positron annihilation to give two photons in the final state.

The dominant contributions to the annihilation reaction $e^+e^- \rightarrow \gamma\gamma$ are shown in Fig. 2. In the first diagram, the positron emits a photon and then annihilates with an electron to produce the second photon; in the second the electron emits the photon before annihilating with the positron to produce the second photon. (In practice, it is usual to draw just one of these diagrams, leaving the other implied.) In interpreting these diagrams, it is important to remember that the direction of the arrows on fermion lines do not indicate their direction of motion, but merely whether the fermions are particles or antiparticles; and that the initial particles are always to the left and the final particles to the right.

A feature of these diagrams is that they are constructed from a single fundamental three-line vertex. This is characteristic of electromagnetic processes. Each vertex has a line corresponding to a single photon being emitted or absorbed; while one fermion line has the arrow pointing toward the vertex and the other away, guaranteeing charge conservation at the vertex. For example, a forbidden vertex like Fig. 3 would correspond to a process in which an

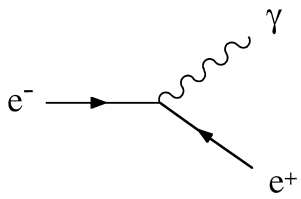


FIGURE 3 A forbidden electromagnetic vertex in which both arrows point into the vertex. Such a vertex would violate electric charge conservation.

electron emitted a photon and turned into a positron, violating charge conservation. Associated with each vertex of a Feynman diagram is a dimensionless parameter, the so-called *coupling constant*, which represents the basic strength of the interaction. For electromagnetic interactions like Figs. 1 and 2, this is the *fine structure constant* α, defined by

$$\alpha \equiv e^2/4\pi\varepsilon_0\hbar c \approx 1/137,$$

where e is the charge on the proton, c is the speed of light, ε_0 is the permittivity of free space, and $\hbar \equiv h/2\pi$.

The number of vertices in any diagram is called the *order* n, and when the probability associated with any given Feynman diagram is calculated, it always contains a factor of α^n. The single-photon exchange diagrams of Fig. 1 thus contain a factor of α^2. In addition, one can easily draw diagrams corresponding to two-photon exchange contributions to the same processes, which contain four vertices, as illustrated in Fig. 4. The contribution from two-photon exchange is therefore of order α^4 and is very small compared to the contribution from single-photon exchange. This is again a general feature of electromagnetic interactions. Because the fine structure constant is very small, only the lowest-order diagrams which contribute to a given process need be taken into account, and more complicated higher-order diagrams with more vertices can to a good approximation be ignored.

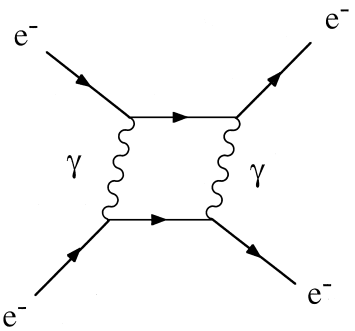


FIGURE 4 A two-photon exchange contribution to electron–electron scattering. Because the fine structure constant α is so small, this makes a very small contribution compared to one-photon exchange, and can usually be ignored.

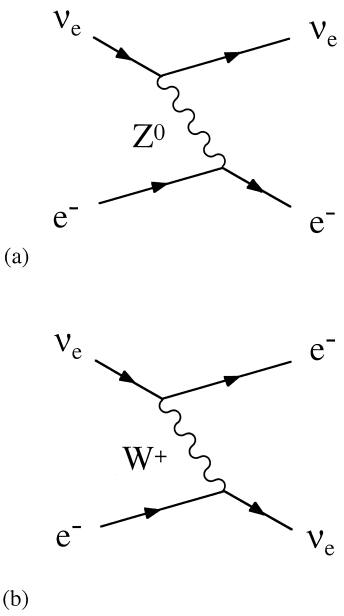


FIGURE 5 The dominant contributions to elastic neutrino electron scattering $\nu_e + e^- \rightarrow \nu_e + e^-$.

E. Ranges of the Fundamental Forces

Feynman diagrams can also be used to describe the weak interactions, which are mediated by the W^+, W^- and Z^0 bosons, rather than the photon. This is illustrated by Fig. 5, which shows the dominant contributions to the elastic scattering reaction

$$\nu_e + e^- \rightarrow \nu_e + e^-.$$

However, a striking difference to the corresponding photon exchange contribution to elastic electron scattering Fig. 1 arises because the W and Z bosons are very massive. In general, it can be shown that the range R of the force arising from the exchange of a particle of rest mass M is

$$R = \hbar/Mc.$$

More precisely, the strength of the force dies away exponentially for distances greater than R and rapidly becomes negligible. Because the photon has zero mass, the range R is infinite for the electromagnetic interaction. However, the W and Z bosons have rest masses which are approximately 85 and 97 times as great as the rest mass of the proton, respectively, leading to ranges $R \approx 2 \times 10^{-18}$ m. This range is tiny compared to, for example, the size of the proton, which is of order 10^{-15} m. It has important consequences for the study of weak interactions, as we shall see when we discuss them in more detail later.

The strong interactions are mediated by the exchange of zero mass, electrically neutral gauge bosons, the gluons. For example, the simplest strong interaction between two quarks is due to the exchange of a single gluon and this is

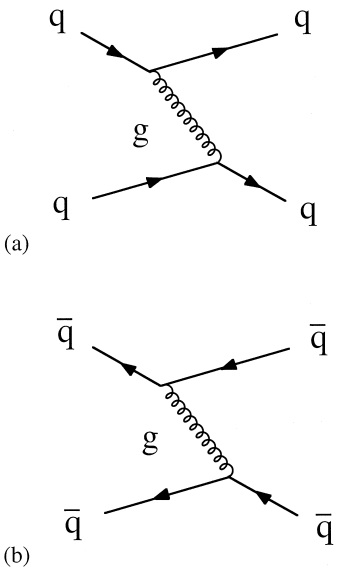


FIGURE 6 Feynman diagrams for (a) quark–quark elastic scattering and (b) antiquark–antiquark elastic scattering via the exchange of a single gluon. The gluon is represented by a "corkscrew" line to distinguish it from a photon.

illustrated in Fig. 6a; while the corresponding diagram for quark–antiquark scattering is shown in Fig. 6b. These diagrams are very similar to the electromagnetic interactions of Fig. 1, and because gluons are massless, the strong interaction between quarks is of infinite range. Nonetheless, the residual interactions between their bound states, the hadrons, are of finite range (about 10^{-15} m), as we shall see.

F. Units: Length, Mass, and Energy

Most branches of science introduce special units which are convenient for their own purposes. Particle physics is no exception. Distances tend to be measured in femtometres or, equivalently *fermis*, with $1\text{ fm} \equiv 10^{-15}$ m. In these units, the radius of the proton is 0.8 fm. The range of the nuclear force between protons and neutrons is of order 1–2 fm, while the range of the weak force is of order 10^{-3} fm. For comparison, the radii of atoms are of order 10^5 fm.

Energies in particle physics are invariably specified in terms of the electron volt, eV, defined as the energy required to raise the electric potential of an electron or proton by 1 V. In terms of S.I. units, $1\text{ eV} = 1.6 \times 10^{-19}$ joules. The subsidiary units $\text{MeV} = 10^6$ eV, $\text{GeV} = 10^9$ eV, and $\text{TeV} = 10^{12}$ eV are also often used. In terms of these units, atomic ionization energies are typically a few electron volts, nuclear binding energies are typically 8 MeV per particle, and the highest particle energies produced in present accelerators are of order 1 TeV.

In order to create a new particle of mass M, an energy at least as great as its rest energy Mc^2 must be supplied. The rest energies of the electron and proton are 0.51 MeV and 0.94 GeV, respectively, whereas the W and Z^0 bosons have rest energies of 80 and 91 GeV, respectively. Correspondingly their masses are conveniently measured in MeV/c^2 or GeV/c^2, so that, for example,

$$M_e = 0.51\,\text{MeV/c}^2, \quad M_p = 0.94\,\text{GeV/c}^2,$$
$$M_W = 80.3\,\text{GeV/c}^2, \quad M_Z = 91.2\,\text{GeV/c}^2$$

In terms of S.I. units, $1\text{ MeV/c}^2 = 1.78 \times 10^{-30}$ kg.

II. LEPTONS

Leptons are one of the three classes of elementary particles in the standard model. They are spin-1/2 fermions with no strong interactions. In this section we take a first look at the leptons.

A. The Six Known Leptons

There are six known leptons, and they occur in pairs, called *generations*, which we write as doublets:

$$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}, \quad \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}, \quad \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}.$$

Each generation comprises a *charged lepton* with electric charge $Q = -e$, and a neutral lepton or *neutrino*. The three charged leptons (e^-, μ^-, τ^-) are the familiar electron, together with two more massive particles, the *mu-lepton* or *muon*, and the *tau-lepton* or *tauon*. The associated neutrinos (ν_e, ν_μ, ν_τ) have very small masses and are called the *electron neutrino*, *mu-neutrino*, and *tau-neutrino* respectively. In addition to the leptons there are six corresponding antiparticles (antileptons):

$$\begin{pmatrix}e^+\\ \bar{\nu}_e\end{pmatrix}, \quad \begin{pmatrix}\mu^+\\ \bar{\nu}_\mu\end{pmatrix}, \quad \begin{pmatrix}\tau^+\\ \bar{\nu}_\tau\end{pmatrix}.$$

Ignoring gravity, the charged leptons interact via both electromagnetic and weak forces, whereas for the ghostly neutrinos, only weak interactions have been observed. Because of this, the neutrinos, which all have extremely small masses, can be detected only with considerable difficulty.

The masses and lifetimes of the leptons are listed for convenience in Table I. The electron and the neutrinos are stable, for reasons which will become clear shortly. The muons decay by the weak interaction processes

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu; \qquad \mu^- \to e^- + \bar{\nu}_e + \nu_\mu$$

with lifetimes 2.2×10^{-6} sec. The tauons also decay by the weak interactions, with a much shorter lifetime $(2.91 \pm 0.02) \times 10^{-13}$ s. Because it is heavier than the muon, it can decay to many different final states, which can include both hadrons and leptons. However, about 35% of decays again

TABLE I Properties of Leptons[a]

Name and symbol	Mass	Q	L_e	L_μ	L_τ	Lifetime(s)	Major decays
Electron e^-	0.511	−1	1	0	0	Stable	None
Electron neutrino ν_e	<3 eV/c^2	0	1	0	0	Stable	None
Muon (mu) μ^-	105.7	−1	0	1	0	2.197×10^{-6}	$e^-\bar{\nu}_e\nu_\mu$ (100%)
Muon neutrino ν_μ	<0.19	0	0	1	0	Stable	None
Tauon (tau) τ^-	1777.0	−1	0	0	1	2.91×10^{-13}	$\mu^-\bar{\nu}_\mu\nu_\tau$ (17.4%)
							$e^-\bar{\nu}_e\nu_\tau$ (17.8%)
							ν_τ + hadrons (~64%)
Tauon neutrino ν_τ	<18.2	0	0	0	1	Stable	None

[a] All have spin-1/2. Masses are given units of MeV/c^2. The antiparticles (not shown) have the same masses as their associated particles, but the electric charges (Q) and lepton numbers (L_ℓ, $\ell = e, \mu, \tau$) are reversed in sign.

lead to purely leptonic final states, via reactions which are very similar to muon decay, for example:

$$\tau^+ \to \mu^+ + \nu_\mu + \bar{\nu}_\tau; \quad \tau^- \to e^- + \bar{\nu}_e + \nu_\tau.$$

These decays illustrate the fundamental principle of lepton number conservation, to which we now turn.

B. Lepton Number Conservation

Associated with each generation of leptons is a conserved quantum number called a *lepton number*. The first of these *lepton numbers* is the *electron number*, defined for any state by

$$L_e \equiv N(e^-) - N(e^+) + N(\nu_e) - N(\bar{\nu}_e),$$

where $N(e^-)$ is the number of electrons present, $N(e^+)$ is the number of positrons present and so on. For single-particle states, $L_e = 1$ for e^- and ν_e; $L_e = -1$ for e^+ and $\bar{\nu}_e$; and $L_e = 0$ for all other particles. The *muon number* and *tauon number*, are defined similarly by

$$L_\mu \equiv N(\mu^-) - N(\mu^+) + N(\nu_\mu) - N(\bar{\nu}_\mu)$$

and

$$L_\tau \equiv N(\tau^-) - N(\tau^+) + N(\bar{\nu}_\tau) - N(\bar{\nu}_\tau)$$

respectively, and their values for all single particle states are summarized for convenience in Table I. They are zero for all particles other than leptons, such as photons, protons, or neutrons, and for multiparticle states the lepton numbers of the individual particles are simply added. For example, the final state in neutron β-decay (i.e., $n \to pe^-\bar{\nu}_e$) has

$$L_e = L_e(p) + L_e(e^-) + L_e(\bar{\nu}_e) = (0) + (1) + (-1) = 0$$

like the initial state, which has $L_e(n) = 0$.

The value of each lepton number is conserved in any reaction. In electromagnetic interactions, this reduces to the conservation of $N(e^-)$–$N(e^+)$, $N(\mu^-)$–$N(\mu^+)$, and $N(\tau^-)$–$N(\tau^+)$, since neutrinos are not involved. This implies that the charged leptons can only be created or annihilated in particle–antiparticle pairs. For example, in the electromagnetic reaction

$$e^+ + e^- \to \mu^+ + \mu^-$$

an electron pair is annihilated and a muon pair is created by the mechanism of Fig. 7.

In weak interactions more general possibilities are allowed which still conserve the lepton numbers. For example, in the tau-decay process $\tau^- \to e^- + \bar{\nu}_e + \nu_\tau$, a tauon converts to a tauon neutrino and an electron is created together with an antineutrino, rather than a positron. The dominant Feynman graph corresponding to this process is shown in Fig. 8.

Lepton number conservation, like electric charge conservation, plays an important role in understanding reactions involving leptons. Observed reactions always conserve lepton numbers, while reactions which violate lepton number conservation are "forbidden" and are never observed. For example, the neutrino scattering reaction

$$\nu_\mu + n \to \mu^- + p$$

is observed experimentally, while the apparently similar reaction

$$\nu_\mu + n \to e^- + p$$

which violates both L_e and L_μ conservation, is not.

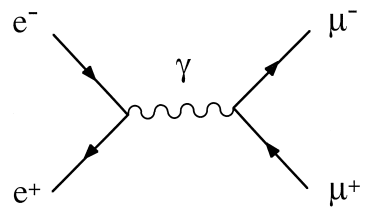


FIGURE 7 One-photon exchange contribution to the reaction $e^+ + e^- \to \mu^+ + \mu^-$.

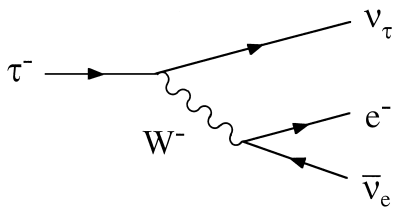


FIGURE 8 The dominant contribution to the process $\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau$.

Finally, conservation laws explain the stability of the electron and the neutrinos. The electron is stable because electric charge is conserved in all interactions and the electron is the lightest charged particle. Hence decays to lighter particles which satisfy all other conservation laws, like $e^- \rightarrow \nu_e + \gamma$, are necessarily forbidden by electric charge conservation. In the same way, lepton number conservation implies that the lightest particles with nonzero values of the three lepton numbers—the three neutrinos—are stable, whether they have zero masses or not.

C. Neutrinos

The electron is very familiar, but the existence of the *electron neutrino* ν_e was first postulated by Pauli in 1930. He did this in order to understand the observed β-decays

$$(Z, N) \rightarrow (Z + 1, N - 1) + e^- + \bar{\nu}_e$$

and

$$(Z, N) \rightarrow (Z - 1, N + 1) + e^+ + \bar{\nu}_e,$$

where (Z, N) denotes an atomic nucleus containing Z protons and N neutrons. The neutrinos and antineutrinos emitted in these decays are not observed experimentally, but are inferred from energy and angular momentum conservation. In the case of energy, if the antineutrino were not present in the first of the reactions, the energy E_e of the emitted electron would be equal to the difference in rest energies of the two nuclei

$$E_e = \Delta M c^2 = [M(Z, N) - M(Z + 1, N - 1)]c^2,$$

where for simplicity we have neglected the extremely small kinetic energy of the recoiling nucleus $(Z + 1, N - 1)$. However, if the antineutrino is present, the electron energy will not be unique, but lie in the range

$$m_e c^2 \leq E_e \leq \left(\Delta M - m_{\bar{\nu}_e}\right)c^2$$

depending on how much of the kinetic energy released in the decay is carried away by the neutrino. Experimentally, the observed energies span the whole of the previously cited range and a measurement of the energy of the electron near its maximum value of $E_e = (\Delta M - m_{\bar{\nu}_e})c^2$ determines the neutrino mass. The most accurate results are compatible with zero mass neutrinos, and when experimental errors are taken into account yield the experimentally allowed range

$$0 \leq m_{\bar{\nu}_e} < 3\,\mathrm{eV/c^2} \approx 6 \times 10^{-6}\, m_e.$$

The masses of both ν_μ and ν_τ can similarly be inferred from e^- and μ^- the energy spectra in the leptonic decays of muons and tauons using energy conservation. The results from these and other decays show that the neutrino masses are very small compared with the masses of the associated charged leptons; and again they are consistent with zero. The present limits are given in Table I.

Small neutrino masses, compatible with these limits, can be ignored in most circumstances, and there are theoretical attractions in assuming neutrino masses are precisely zero. However, there is now strong evidence for physical phenomena which could not occur if the neutrinos had zero mass (see Section VII). The possibility of small masses compatible with the experimental limits cannot be ignored.

Neutrinos can only be detected with extreme difficulty. For example, electron neutrinos and antineutrinos can in principle be detected by observing the *inverse β-decay* processes

$$\nu_e + n \rightarrow e^- + p$$

and

$$\bar{\nu}_e + p \rightarrow e^+ + n$$

However, because neutrinos only interact via the weak interaction, the probability for these and other processes to occur is extremely small. In particular, the neutrinos and antineutrinos emitted in β-decays, with energies of order 1 MeV, have mean free paths in matter of order 10^6 km. Nevertheless, if the neutrino flux is intense enough and the detector is large enough, the reactions can be observed. In particular, uranium fission fragments are neutron rich, and decay by electron emission to give an antineutrino flux which can be of order or 10^{17} m^{-2} s^{-1} or more in the vicinity of a nuclear reactor. These antineutrinos will occasionally interact with protons in a large detector, enabling examples of the inverse β-decay reaction to be observed. Electron neutrinos were first detected in this way in a classic experiment by Reines and Cowan in 1959, and their interactions have been studied in considerably more detail since.

The mu-neutrino ν_μ has been detected using the reaction $\nu_\mu + n \rightarrow \mu^- + p$ and other reactions. In this case, well-defined, high-energy mu-neutrino beams can be created in the laboratory by exploiting the decay properties of particles called *pions*, which we shall meet in Section III. The probability of neutrinos interacting with matter increases rapidly with energy, and for large detectors, neutrino

events initiated by such beams are so copious they have become an indispensable tool in studying both the fundamental properties of weak interactions and the internal structure of the proton. Finally, in 2000, a few examples of tau-neutrinos were reported, so that almost 70 years after Pauli first suggested the existence of a neutrino, all three types have been directly detected.

D. Universal Lepton Interactions

The three neutrinos have similar properties, but the three charged leptons are strikingly different. For example, the magnetic moment of the electron is roughly 200 times greater than that of the electron; high-energy electrons are mostly stopped by 1 cm of lead, while muons are the most penetrating form of radiation known, apart from neutrinos; and the tauon lifetime is many orders of magnitude smaller than the muon lifetime, while the electron is stable. It is therefore a remarkable fact that all experimental data are consistent with the assumption that the interactions of the electron and its associated neutrino are identical with those of the muon and its associated neutrino and of the tauon and its neutrino, *provided the mass differences are taken into account*. This property, called *universality*, can be verified with great precision, because we have a precise theory of electromagnetic and weak interactions, which enables us to predict the mass dependence of all observables. Thus the magnitude of the magnetic moment μ_P of any point-like, spin-1/2 particle P of mass m is given by

$$\mu_P = q\hbar/2m,$$

where q is its charge and h is Planck's constant. (In quoting this and other results, we ignore tiny but well-understood corrections which are typically of order 1 part/1000.) The agreement of this prediction with experiment both explains the differences in the observed magnetic moments and confirms the point-like nature of the leptons.

The *radiation length*, which is a measure of how fast a charged particle loses energy in the form of electromagnetic radiation when traversing matter, is proportional to the mass squared. Hence it is about 4×10^4 times greater for muons than for electrons, explaining their much greater penetrating power in matter. Finally, the rates for weak decays are extremely sensitive to the kinetic energy released in the decay; and the ratio of the decay rates Γ for muon and tauon leptonic decays is predicted to be

$$\frac{\Gamma(\tau^- \to e^- + \bar{\nu}_e + \nu_\tau)}{\Gamma(\mu^- \to e^- + \bar{\nu}_e + \nu_\mu)} = 1.34 \times 10^6.$$

This is in excellent agreement with the experiment, and accounts very well for the huge difference between the tauon and muon lifetimes when the other decay modes of the tauon are also taken into account. The electron is stable because it is the lightest charged particle, as we have already noted.

The previously cited are just some of the most striking manifestations of the universality of lepton interactions. More generally, the three generations of leptons tell not three stories, but in all essential points, one story three times.

The question naturally arises as to whether there are more generations, with identical interactions, waiting to be discovered. This question has been answered, under reasonable assumptions, by an experimental study of the decays of the Z^0 boson. This particle, which has a mass of 91 GeV/c^2, is one of the two gauge bosons associated with the weak interaction. It decays, among other final states, to neutrino pairs

$$Z^0 \to \ell + \bar{\ell} \qquad (\ell = e, \mu, \tau)$$

with a total rate which is proportional to the number of neutrino species, if we assume universal lepton interactions and neutrino masses which are small compared to the mass of the Z^0. The measured total decay rate is consistent with the expectation for three, but not four, neutrino species. Only three generations of leptons can exist, if we assume universal lepton interactions and exclude very large neutrino masses.

Why there are just three generations of leptons, and not fewer or more, remains a mystery.

III. QUARKS AND HADRONS

We turn now to the strongly interacting particles—the quarks and their bound states, the hadrons. These also interact by the weak and electromagnetic interactions, although such effects can often be neglected compared to the strong interactions. To this extent we are entering the realm of "strong interaction physics."

A. Introduction

More than 200 strongly interacting particles have now been observed, all with zero or integer electric charges: 0, ± 1, or ± 2 in units of e. These particles are called *hadrons*. They are all bound states of the fundamental spin-1/2 particles called *quarks*, whose electric charge is either $+2/3$ or $-1/3$, and/or antiquarks, with charges $-2/3$ or $+1/3$. The quarks themselves have never been directly observed as single, free particles, but there is compelling evidence for their existence. The evidence comes from three main areas: *hadron spectroscopy*, *lepton scattering* and *jets*.

1. Hadron Spectroscopy

This is the study of the static properties of hadrons: their masses, lifetimes, and decay modes, and especially the values of their quantum numbers, including their spins, electric charges, and many more. The existence and properties of quarks were first inferred from hadron spectroscopy by Gell-Mann and Zweig in 1964; and the close correspondence between the experimentally observed hadrons and those predicted by the quark model remains one of the strongest reasons for our belief in the existence of quarks.

2. Lepton Scattering

In 1911, Rutherford and collaborators established the existence of the nucleus by studying the large angle scattering of relatively high-energy charged particles from atoms. In much the same way, the large angle scattering of high-energy leptons (electrons, muons, and neutrinos) from protons and neutrons reveals the existence of point-like quarks within them. This phenomenon was first observed at SLAC in 1969.

3. Jets

High-energy collisions can cause the quarks within hadrons, or newly created quark–antiquark pairs, to fly apart from each other with very high energies. Before they can be observed, these quarks are converted by relatively gentle interactions into jets of hadrons, whose production rates and angular distributions reflect those of the quarks from which they originated. They were first clearly identified in electron–positron collisions at the DESY laboratory in Hamburg in 1979, and an example of a "two-jet" event observed is shown in Fig. 9. The production rate and angular distribution of the observed jets closely matches that of quarks produced in the reaction

$$e^+ + e^- \rightarrow q + \bar{q}$$

by the mechanism of Fig. 10. Such jets have now been observed in many reactions, and are the closest thing to a quark "track" we are ever likely to see.

As this last remark implies, despite the overwhelming evidence for the existence of quarks, it is extremely unlikely that the failure to detect free quarks is an experimental problem. There are several reasons for believing this. First, free quarks would be easily distinguished from other particles by their fractional charges. Second, electric charge conservation implies that a fractionally charged particle cannot decay to a final state composed entirely of particles with integer electric charges. Hence the lightest fractionally charged particle, i.e., the lightest free quark, would be stable. Finally, some of the quarks are not very massive (see following) and because they interact by the strong interaction, one would expect free quarks to be copiously produced in, for example, high-energy proton–proton collisions. However, despite careful and exhaustive searches in ordinary matter, in cosmic rays and in high-energy collision products, free quarks have never been observed. The obvious conclusion—that quarks exist solely within hadrons and not as isolated free particles—is called *confinement*. The modern theory of strong interactions is called *quantum chromodynamics* and offers at least a qualitative account of confinement, although the details elude us due to the extreme difficulty of performing accurate calculations. In what follows, we shall assume confinement and use the properties of quarks to interpret the properties of hadrons.

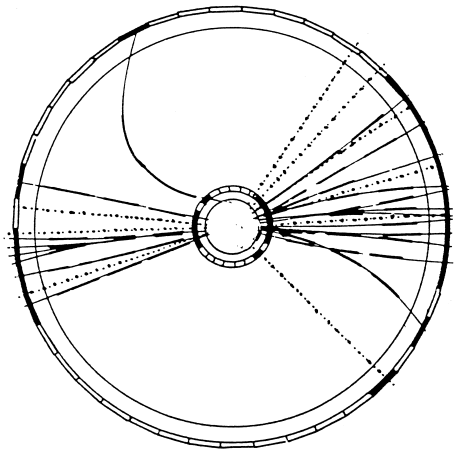


FIGURE 9 Computer reconstruction of a "two-jet" event observed in e^+e^- annihilation in a detector called a drift chamber (see Section VI). The interaction takes place at the center of the chamber, where the two beams collide head-on—one into the paper and the other out of the paper. The solid lines indicate the trajectories of charged particles and the dotted lines those of neutral particles, the latter being detected outside the chamber by other means.

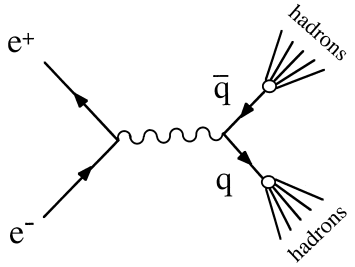


FIGURE 10 Basic mechanism of two-jet production in e^+e^- annihilation.

B. The Six Known Quarks

Six distinct types, or *flavors*, of spin-1/2 quarks are now known to exist. Like the leptons, they occur in pairs, or *generations*, denoted

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix}.$$

Each generation consists of a quark with charge $+\frac{2}{3}$ (u, c, or t) together with a quark of charge $-\frac{1}{3}$ (d, s, or b), in units of e. They are called the *down* (d), *up* (u), *strange* (s), *charmed* (c), *bottom* (b), and *top* (t) quarks. The corresponding antiquarks are denoted

$$\begin{pmatrix} \bar{d} \\ \bar{u} \end{pmatrix}, \quad \begin{pmatrix} \bar{s} \\ \bar{c} \end{pmatrix}, \quad \begin{pmatrix} \bar{b} \\ \bar{t} \end{pmatrix}.$$

with charges $+1/3$ ($\bar{d}$, $\bar{s}$, or $\bar{b}$) and $-2/3$ ($\bar{u}$, $\bar{c}$, or $\bar{t}$).

Approximate quark masses are given in Table II. Except for the top quark, these masses are inferred indirectly from the observed masses of their hadron bound states, together with models of quark binding. The stability of quarks in hadrons—like the stability of protons and neutrons in atomic nuclei—is influenced by their interaction energies. However, for the s, c, and b quarks these effects are small enough for them to be assigned approximate lifetimes of 10^{-8}–10^{-10} s for the s-quark and 10^{-12}–10^{-13} s for both the c- and b-quarks. The top quark is much heavier than the other quarks and its lifetime is of order 10^{-25} sec. This lifetime is so short that, when top quarks are created, they decay too quickly to form observable hadrons. In contrast to the other quarks, our knowledge of the top quark is based entirely on observations of its decay products.

TABLE II Properties of Quarks[a]

Name	Symbol	Mass	Q	Lifetime(s)	Major decays
Down	d	$m_d \approx 0.3$	$-1/3$		
Up	u	$m_u \approx m_d$	$2/3$		
Strange	s	$m_s \approx 0.5$	$-1/3$	10^{-8}–10^{-10}	$s \to u + X$
Charmed	c	$m_c \approx 1.5$	$2/3$	10^{-12}–10^{-13}	$c \to s + X$
					$c \to d + X$
Bottom	b	$m_b \approx 4.5$	$-1/3$	10^{-12}–10^{-13}	$b \to c + X$
Top	t	$m_t = 180 \pm 12$	$2/3$	$\sim 10^{-25}$	$t \to b + X$

[a] All have spin-1/2. Masses are given units of GeV/c^2. The antiparticles (not shown) have the same masses as their associated particles, but the electric charges (Q) are reversed in sign. In the major decay modes X stands for other particles.

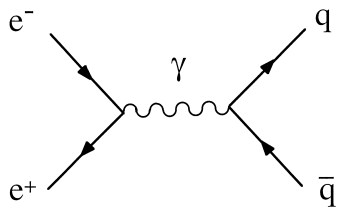


FIGURE 11 The dominant Feynman diagram for the reaction $e^+ + e^- \to q + \bar{q}$.

C. Quark Numbers

In strong and electromagnetic interactions, quarks can only be created or destroyed as particle–antiparticle pairs. This implies, for example, that in electromagnetic processes corresponding to the Feynman diagram of Fig. 11, the reaction

$$e^+ + e^- \to c + \bar{c}$$

which creates a $c\bar{c}$ pair, is allowed, but the reaction

$$e^+ + e^- \to c + \bar{u}$$

producing a $c\bar{u}$ pair, is forbidden. More generally, it implies conservation of each of the six *quark numbers*

$$N_f \equiv N(f) - N(\bar{f}) \qquad (f = u, d, s, c, b, t),$$

where $N(f)$ is the number of quarks of type f present and $N(\bar{f})$ is the number of $\bar{f}$-antiquarks present. For example, for single particle states; $N_c = 1$ for the c-quark; $N_c = -1$ for the $\bar{c}$ antiquark; and $N_c = 0$ for all other particles. Similar results apply for the other quark numbers N_f, and for multi-particle states the quark numbers of the individual particles are simply added. Thus a state containing the particles u, u, d, has $N_u = 2$, $N_d = 1$, and $N_f = 0$ for the other quark numbers with $f = s, c, b, t$.

In weak interactions, more general possibilities are allowed, and only the total quark number

$$N_q \equiv N(q) - N(\bar{q})$$

is conserved, where $N(q)$ and $N(\bar{q})$ are the total number of quarks and antiquarks present, irrespective of their flavor. This is illustrated by the decay modes of the quarks themselves, some of which are listed in Table II, which are all weak interaction processes. For example, one of the *main* decay modes of the charmed quark is

$$c \to s + u + \bar{d}$$

in which a c-quark is replaced by an s-quark and a u-quark is created together with a $\bar{d}$ antiquark. This clearly violates conservation of the quark numbers N_c, N_s, N_u, and N_d, but the total quark number N_q is conserved.

In practice, it is convenient to replace the total quark number N_q in discussions by the *baryon number*, defined by

$$B \equiv N_q/3 = [N(q) - N(\bar{q})]/3.$$

Like the electric charge and the lepton numbers introduced in the last section, the baryon number is conserved in all known interactions.

D. Baryons and Mesons

The principal properties of atoms and nuclei can be explained in terms of their proton, neutron, and electron constituents, although in practice many details are too complicated to be accurately calculated. However, the properties of these constituents can be determined without reference to atoms and nuclei, by studying them directly as free particles in the laboratory. In this sense atomic and nuclear physics are no longer fundamental, although they are still very interesting and important if we want to understand the world we live in.

In the case of hadrons the situation is more complicated. Their properties are explained in terms of a few fundamental quark constituents; but the properties of the quarks themselves can only be studied experimentally by appropriate measurements on hadrons. Whether desirable or not, studying quarks without hadrons is not an option.

The observed hadrons are of three types. These are the *baryons*, which have half-integer spin and are assumed to be bound states of three quarks ($3q$); the *antibaryons*, which are their antiparticles and are assumed to be bound states of three antiquarks ($3\bar{q}$); and the *mesons* which have integer spin and are assumed to be bound states of a quark and an antiquark ($q\bar{q}$). The baryons and antibaryons have baryon numbers 1 and –1 respectively, while the mesons have baryon number 0. Hence the baryons and antibaryons can annihilate each other in reactions which conserve baryon number to give mesons or, more rarely, photons or lepton–antilepton pairs, in the final state. Some examples of baryons and mesons, together with their quark compositions, are shown in Table III.

The lightest known baryons are the proton and neutron with the quark compositions

$$p = uud, \qquad n = udd.$$

These particles are already familiar as constituents of atomic nuclei, and the birth of particle physics as a new subject, distinct from atomic and nuclear physics, dates from 1947, when hadrons other than the neutron and proton were first detected. These were the *pions* and the *kaons* discovered in cosmic rays by groups in Bristol and Manchester, respectively.

TABLE III Some Examples of Baryons and Mesons, with Their Quark Compositions and Major Decay Modes[a]

Particle	Mass	Lifetime(s)	Major decays
$\pi^+(u\bar{d})$	140	2.6×10^{-8}	$\mu^+\nu_\mu$ (~100%)
$\pi^0(u\bar{u}, d\bar{d})$	135	8.4×10^{-17}	$\gamma\gamma$ (~100%)
$K^+(u\bar{s})$	494	1.2×10^{-8}	$\mu^+\nu_\mu$ (64%)
			$\pi^+\pi^0$ (21%)
$K^{*+}(u\bar{s})$	892	$\sim 1.3 \times 10^{-23}$	$K^+\pi^0$, $K^0\pi^+$ (~100%)
$D^-(d\bar{c})$	1869	1.1×10^{-12}	Several seen
$B^-(b\bar{u})$	5278	1.6×10^{-12}	Several seen
$p(uud)$	938	Stable	None
$n(udd)$	940	887	$pe^-\bar{\nu}_e$ (100%)
$\Lambda(uds)$	1116	2.6×10^{-10}	$p\pi^-$ (64%)
			$n\pi^0$ (36%)
$\Lambda^{++}(uuu)$	1232	$\sim 0.6 \times 10^{-23}$	$p\pi^+$ (100%)
$\Omega^-(sss)$	1672	0.8×10^{-10}	ΛK^- (68%)
			$\Xi^0\pi^-$ (24%)
$\Lambda_c^+(udc)$	2285	2.1×10^{-13}	Several seen

[a] Masses are in MeV/c^2.

The discovery of pions was not totally unexpected, since Yukawa had famously predicted their existence and their approximate masses in 1935, in order to explain the observed range of nuclear forces. There are three types of pion, denoted $\pi^{\pm}(140)$, $\pi^0(135)$, where here and in what follows we give the hadron masses in brackets in units of MeV/c^2 and use a superscript to indicate the electric charge in units of e. They are the lightest known mesons and have the quark compositions

$$\pi^+ = u\bar{d}, \quad \pi^0 = u\bar{u}, \quad d\bar{d}, \quad \pi^- = d\bar{u}.$$

While the charged pions have a unique composition, the neutral pion is composed of both $u\bar{u}$ and $d\bar{d}$ pairs in equal amounts. Nowadays, all the pions can be copiously produced in high-energy collisions at accelerators by strong interaction processes such as

$$p + p \rightarrow p + n + \pi^+.$$

In contrast to the discovery of the pions, the discovery of the kaons was totally unexpected, and they were almost immediately recognized as a completely new form of matter, because they had supposedly "strange" properties. Eventually (1954) it was realized that these properties were precisely what would be expected if kaons had nonzero values of a hitherto unknown quantum number, called *strangeness*, which was conserved in strong and electromagnetic interactions, but not conserved in weak interactions. Particles with nonzero strangeness were christened *strange particles*, and with the advent of the quark model in 1964, it was realized that strangeness

S was, apart from a sign, the strange quark number introduced in the last section, i.e.,

$$S = -N_s.$$

Kaons are the lightest strange mesons, with the quark compositions:

$$K^+(494) = u\bar{s}, \qquad K^0(498) = d\bar{s},$$

where K^+ and K^0 have $S = +1$ and K^- and $\bar{K}^0$ have $S = -1$, while the lightest strange baryon is the *lambda*, with the quark composition

$$\Lambda = uds.$$

Subsequently, hadrons containing c and b quarks have also been discovered, with nonzero values of the *charm* and *beauty* quantum numbers defined by

$$C \equiv N_c \equiv N(c) - N(\bar{c})$$

$$\text{and} \quad \tilde{B} \equiv -N_b \equiv -N(b) - N(\bar{b}).$$

The previously cited examples illustrate just some of the different combinations of quarks which form baryons or mesons. However, for each of these quark combinations there exist not one, but many, states. This is illustrated in Fig. 12, which shows all the known $u\bar{d}$ states with masses below 1.5 GeV/c^2. Each of these states is labeled by its spin and by its *parity*, which is a quantum mechanical observable related to the behavior of the state under a mirror reflection. It can take the value $+1$ or -1, and the notation 1^- is used to indicate a particle of spin-1 with negative parity, and so on. The lowest lying state shown in Fig. 12 has spin-parity 0^- and is the π^+ meson discussed above. It can be regarded as the "ground state" of the $u\bar{d}$ system. The other "excited" states can have different spin-parities depending on the different states of motion

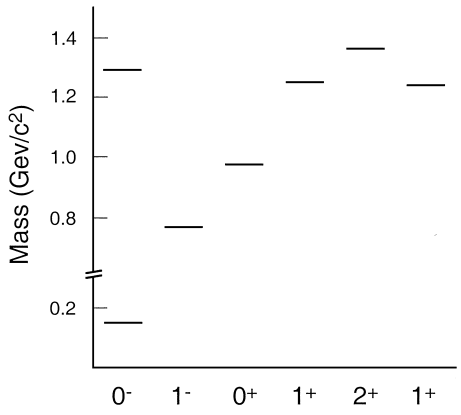


FIGURE 12 Observed bound states of the $u\bar{d}$ system with masses below 1.5 GeV/c^2. The state with spin-parity 0^- is the ground state $\pi^+(140)$. The others can be regarded as its excited states.

of the quarks within the hadron. They are often called *resonances* and they have very short lifetimes, of order 10^{-23} sec. It is part of the triumph of the quark model that it successfully accounts for the excited states of the various quark systems, as well as their ground states, when the internal motion of the quarks is properly taken into account.

E. Hadron Lifetimes

Hadrons have typical radii r of order 1 fm, with an associated time scale r/c of order 10^{-23} sec. The vast majority are highly unstable resonances, corresponding to excited states of the various quark systems, and decay to lighter hadrons by the strong interaction with lifetimes of this order. A typical example is the $K^{*+}(890) = u\bar{s}$ resonance, which decays to $K^+\pi^0$ and $K^0\pi^+$ final states with a lifetime of 1.3×10^{-23} sec. The quark description of the process $K^{*+} \to K^0 + \pi^+$, for example, is

$$u\bar{s} \to d\bar{s} + u\bar{d}.$$

From this we see that the final state contains the same quarks as the initial state, plus an additional $d\bar{d}$ pair, so that the quark numbers N_u and N_d are separately conserved. This is characteristic of strong and electromagnetic processes, which are only allowed if all the quark numbers N_u, N_d, N_s, N_c, and N_b are separately conserved.

Since leptons and photons do not have strong interactions, hadrons can only decay by the strong interaction if lighter states composed solely of other hadrons exist with the same quantum numbers. While this is possible for the majority of hadrons, it is not in general possible for the lightest state corresponding to any given quark combination. These hadrons, which cannot decay by strong interactions, are long-lived on a timescale of order 10^{-23} sec and are often called *stable particles*. Here we shall call them *long-lived particles*, because except for the proton they are not absolutely stable, but decay by either the electromagnetic or weak interaction.

The proton is stable because it is the lightest particle with nonzero baryon number and baryon number is conserved in all known interactions. A few of the other long-lived hadrons decay by electromagnetic interactions to final states which include photons. These decays, like the strong interaction, conserve all the individual quark numbers. An example of this is the neutral pion, which has $N_u = N_d = N_s = N_c = N_b = 0$ and decays by the reaction

$$\pi^0(u\bar{u}, d\bar{d}) \to \gamma + \gamma$$

with a lifetime of 0.8×10^{-16} sec. However, most of the long-lived hadrons have nonzero values for at least one of the quark numbers, and can only decay by the weak

interaction, which violates quark number conservation. For example, the positive pion decays with a lifetime of 2.6×10^{-8} sec by the reaction

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

while the $\Lambda(1116) = uds$ baryon decays mostly by the reaction

$$\Lambda \rightarrow p + \pi^-$$

with a lifetime of 2.6×10^{-10} sec. The quark interpretations of these reactions are

$$(u\bar{d}) \rightarrow \mu^+ + \nu_\mu,$$

in which a u-quark annihilates with a $\bar{d}$-antiquark, violating both N_u and N_d conservation; and for lambda decay

$$sud \rightarrow uud + d\bar{u},$$

in which an s quark turns into a u quark and a $u\bar{d}$ pair is created, violating N_u, N_d, and N_s conservation.

We see from the previous equations that the strong, electromagnetic, or weak nature of a given hadron decay can be determined by inspecting quark numbers. The resulting lifetimes can then be summarized as follows. Strong decays lead to lifetimes which are typically of order 10^{-23} sec. Electromagnetic decay rates are suppressed by powers of the fine structure constant α relative to strong decays, leading to observed lifetimes in the range 10^{-16}–10^{-21} sec. Finally, weak decays give longer lifetimes, which depend sensitively on the characteristic energy of the decay. A useful measure of this characteristic energy is the *Q-value*, which is the kinetic energy released in the decay of the particle at rest. For neutron decay ($n \rightarrow p + e^- + \bar{\nu}_e$), the Q-value

$$Q = m_n - m_p - m_e - m_{\bar{\nu}_e} = 0.79\,\text{MeV}$$

is very small, leading to a lifetime of about 10^3 sec. However, Q-values of the order of 10^2–10^3 MeV are more typical, leading to lifetimes in the range 10^{-7}–10^{-13} sec. Thus hadron decay lifetimes are reasonably well understood and span some 27 orders of magnitude, from about 10^{-24} to about 10^3 sec. The typical ranges corresponding to each interaction are summarised in Table IV.

TABLE IV Typical Lifetimes of Hadrons Decaying by the Three Interactions

Interaction	Lifetimes(s)
Strong	10^{-22}–10^{-24}
Electromagnetic	10^{-16}–10^{-21}
Weak	10^{-7}–10^{-13}

F. Flavor Independence

Flavor independence is one of the most fundamental properties of the strong interaction. It is the statement that the strong force between two quarks at a fixed distance apart is independent of which quark flavors u, d, c, b, t are involved. Thus, for example, the strong forces between us and ds pairs are identical. The same principle applies to quark–antiquark forces, which are, however, not identical to quark–quark forces. Flavor independence does not apply to the electromagnetic interaction, since the quarks have different electric charges, but compared to the strong force between quarks, the electromagnetic force is a small correction. In addition, in applying flavor independence one must take account of the quark mass differences, which can be nontrivial. However, there are cases where these corrections are small or easily estimated, and the phenomenon of flavor independence is plain to see.

One of the striking things about hadrons is that they occur in families of particles with approximately the same masses, called *charge multiplets*. Within a given family, all particles have the same spin-parity and the same strangeness, charm, and beauty, but differ in their electric charges. Examples are the triplet of pions, (π^+, π^0, π^-), and the nucleon doublet (p, n). This behavior reflects an approximate symmetry between u and d quarks. This arises because these two quarks have the same mass, apart from a small correction

$$m_d - m_u = (3 \pm 1)\,\text{MeV/c}^2$$

so that in this case, mass corrections can to a good approximation be neglected. For example, consider the case of the proton and neutron, with quark contents

$$p(938) = uud, \qquad n(940) = uud.$$

If we neglect the small mass difference between the u and d quarks and also the electromagnetic interactions, which is equivalent to setting all electric charges to zero, so that the forces acting on the u and d quarks are exactly equal, then replacing the u quark by a d quark in the proton would produce a "neutron" which is essentially identical to the proton. Another example is the K mesons

$$K^+(494) = u\bar{s}, \qquad K^0(498) = d\bar{s},$$

where again, interchanging a u and d quark interchanges K^+ and K^0. Of course the symmetry is not exact because of the small mass difference between the u and d quarks and because of the electromagnetic forces, and it is these that lead to the small differences in mass within multiplets. The symmetry between u and d quarks is called *isospin symmetry* and greatly simplifies the interpretation of hadron physics.

A case where the mass differences between quarks are large, but relatively easily taken into account, is the comparison of the $c\bar{c}$ and $b\bar{b}$ quark systems. These are called *charmonium* and *bottomium*, respectively, and they are important because, in this case, the quarks are so heavy that they move slowly enough within the resulting hadrons to be treated nonrelativistically to a first approximation. This means that the rest energies of the bound states, and hence their masses, can be calculated from the static potential between the quarks in exactly the same way that the energy levels in the hydrogen atom are calculated from the Coulomb potential. In this case, however, the procedure is reversed, with the aim of determining the form of the static potential from the rather precisely measured energies of the bound states. To cut a long story short, one finds that the potentials required to describe the system are the same within the reasonably small uncertainties of the method, confirming again the flavor independence of the strong force.

IV. STRONG INTERACTIONS

A. Color

The quark model account of the hadron spectrum is very successful. However, it begs several questions. One is why only $3q$, $3\bar{q}$, and $q\bar{q}$ states are observed? Another arises from a particular assumption which must be made in order to obtain agreement with the observed hadron spectrum. This is: if two quarks of the same flavor uu, dd, ss ... are in the same spatial state, they must also be in the same spin state, with their spins parallel. Indeed for some states containing three like-quarks, such as the $\Omega^- = sss$, all three like-quarks have the same space and spin states. However, there is a fundamental theorem in quantum physics, called the Pauli exclusion principle, which states that identical fermions cannot simultaneously be in the same state. The three s-quarks in the Ω^- therefore cannot be in the same state, so how do they differ?

In order to resolve this last question, it is necessary to assume a new degree of freedom exists for quarks, but not leptons, which is somewhat whimsically called *color*. The basic properties of color are as follows.

(a) Any quark $u, d, s, \ldots$ can exist in three different color states, which we denote r, g, b for "red," "green," and "blue," respectively. There is direct experimental evidence that just three such states exist, as we shall see in Section D.3.

(b) Each of these states is characterized by the corresponding values of two conserved *color charges*, denoted I_3^C and Y^C, which are strong interaction analogs of the electric charge in electromagnetic interactions. These charges depend only on the color state r, g, b and not on the flavor $u, d, s, \ldots$. The particular values for quarks and antiquarks are given in Table V, and are a consequence of a fundamental symmetry of the strong interaction, with the imposing name of SU(3) color symmetry, which we will not pursue. For multiparticle states, the color charges of the individual states are simply added.

TABLE V Values of the Color Charges I_3^C and Y^C for the Color States of Quarks and Antiquarks

	Quarks			Antiquarks	
(a)	I_3^C	Y^C	(b)	I_3^C	I_3^C
r	1/2	1/3	$\bar{r}$	−1/2	−1/3
g	−1/2	1/3	$\bar{g}$	1/2	−1/3
b	0	−2/3	$\bar{b}$	0	2/3

(c) Only *color singlet* states with zero values for the color charges are observable as free particles. This is the hypothesis of *color confinement*. It can be derived, at least approximately, from the theory of strong interactions we shall describe shortly.

Returning to the quark model, we can see from Table V that a $3q$ state can only have both $I_3^C = 0$ and $Y^C = 0$ if we have one quark in an r state, one in a g state and one in a b state. Hence in the Ω^-, for example, all three s-quarks are necessarily in different color states, as required by the Pauli principle.

One can also see from Table V that free quarks and fractionally charged combinations like qq and $qq\bar{q}$ are forbidden by color confinement, in accordance with experimental observation. On the other hand the combinations $q\bar{q}$ and $3q$ used in the simple quark model are allowed. More unusual are combinations like $qq\bar{q}\bar{q}$ and $qqqq\bar{q}$, which could give rise to "exotic" mesons and baryons, respectively, and are not forbidden by color confinement. Although there is no convincing experimental evidence for such states, they may well play a minor role in hadron physics.

B. Quantum Chromodynamics (QCD)

The theory which describes strong interactions in the standard model is called *quantum chromodynamics*, or QCD for short. Although QCD is not tested to the same extent or precision as quantum electrodynamics (QED), it is nevertheless in impressive agreement with a large body of experimental data, and is not contradicted by any known experiment. QCD is similar to QED in that both describe interactions which are mediated by massless spin-1 bosons, gluons in the former case and photons in the latter.

Theories of this type are called *gauge theories*. Gluons have zero electric charge, like photons, but unlike photons which couple to electric charge, gluons couple to *colour* "charges." This leads immediately to the flavor independence of strong interactions introduced in Section III.F, that is, the different quark flavors $a = u, d, s, c, b$, and t must have identical strong interactions, because they exist in the same three color states r, g, b with the same possible values of the color charges.

A second property of strong interactions that follows from this picture is that the forces between the quarks must be of long range, because the gluons have zero mass. This does not imply that the forces between hadrons are also long range. The forces between the "colorless" hadrons are the residues of the forces between their quark constituents, and cancel when the hadrons are far apart.

QED and QCD both describe interactions, albeit of very different strengths, which are mediated by massless spin-1 bosons which couple to conserved charges. However, there is a crucial difference between them, which profoundly effects the characters of the resulting forces. While the photons which couple to the electric charge are themselves neutral, gluons have nonzero values of the color charges to which they couple. This is illustrated in Fig. 13, which shows a particular example of a quark–quark interaction by gluon exchange. In this diagram, the color states of the two quarks are interchanged, and the gluon has nonzero values of the color quantum numbers, whose values follow from color charge conservation at the vertices. The first thing implied by this is that gluons, like quarks, are confined and cannot be observed as free particles. The second is that since gluons couple to particles with nonzero color charges, and since gluons themselves also have nonzero color charges, then by implication gluons couple to other gluons. The two types of gluon self-coupling which occur in QCD are illustrated in Fig. 14, which shows the two lowest order contributions to gluon–gluon scattering. The first is a gluon exchange process in analogy to Fig. 6(a)

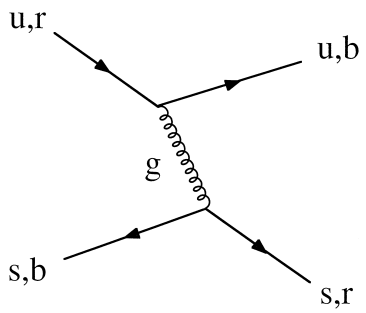


FIGURE 13 Example of quark–quark scattering by gluon exchange. In this diagram the quark flavor *u* or *s* is unchanged on gluon emission, but the color state can change, as shown.

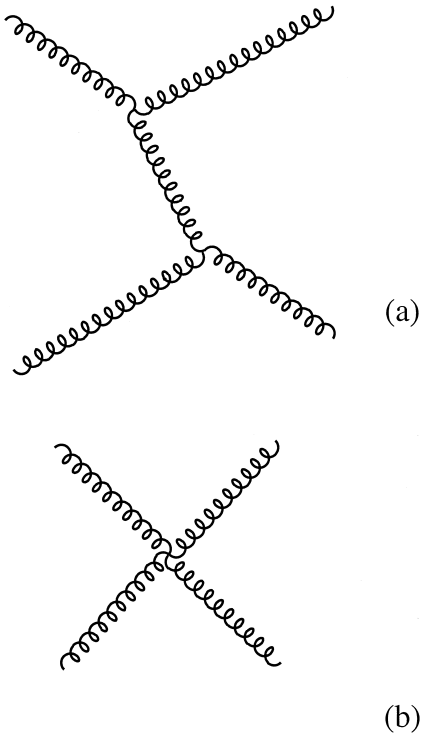


FIGURE 14 The two lowest-order contributions to gluon–gluon scattering in QCD: (a) one-gluon exchange, (b) contact interaction.

for quark–quark scattering, while the second involves a so-called "zero range" or "contact" interaction.

These gluon–gluon interactions have no analog in QED, and it can be shown that they lead to properties of the strong interaction which differ markedly from those of the electromagnetic interaction. These properties are *color confinement*, which we have discussed earlier, and *asymptotic freedom*, which means that the interaction gets weaker at short distances. Conversely, as the distance between the quarks increases, the interaction gets stronger. In this strong interaction regime the situation is very complicated, and it has not yet been possible to evaluate the theory precisely. We therefore have to rely on results obtained by numerical simulations of the theory, which approximate space by a limited number of discrete lattice points. This approach is called *lattice gauge theory*, and it requires very large computers. While very precise results are difficult to obtain, the demonstration of confinement in QCD rests largely on such simulations.

These features are conveniently illustrated by considering the static potential between a heavy quark and an antiquark. At short interquark distances $r \lesssim 0.1$ fm, the interaction is dominated by one-gluon exchange and we might expect a Coulomb-like potential analogous to that

arising from one-photon exchange in QED. In fact it can be shown that the potential is given by

$$V(r) = -\frac{4}{3}\frac{\alpha_s}{r}\hbar c \qquad (r \leq 0.1\,\text{fm}),$$

where the *strong coupling constant* α_s is a measure of the strength of the interaction analogous to the fine structure constant α in QED. Because of asymptotic freedom, the strength of the interaction, and hence α_s, decreases with decreasing r, but $r < 0.1$ fm for this variation is slight and can in many applications be neglected. At distances beyond 0.1 fm, however, the strength of the interaction increases with increasing r, and one-gluon exchange no longer dominates. In this region, we are forced to rely on numerical calculations of very limited precision. These show that at large distances the potential rises approximately linearly

$$V(r) \approx \lambda r \qquad (r \geq 1\,\text{fm}),$$

where the constant λ cannot be calculated precisely, but is of order 1 Gev fm^{-1}. This is an example of a *confining potential* in that it does not die away with increasing separation, and the force between the quark and antiquark cannot be neglected, even when they are very far apart.

C. The Strong Coupling Constant

The strong interaction derives its name from the strong forces acting at distances of order 1 fm which, among other things, bind quarks into hadrons. However, some remarkable phenomena depend on the fact that the interaction gets weaker at short distances; that is, on asymptotic freedom. Such short-distance interactions are associated with large momentum transfers $|\mathbf{q}|$ between the particles, with

$$|\mathbf{q}| = O(\hbar/r),$$

where r is the distance at which the interaction occurs. In such cases, the strength of the interaction can be shown to depend on the squared energy–momentum transfer

$$Q^2 \equiv \mathbf{q}^2c^2 - E_q^2$$

which is a Lorentz invariant generalization of the squared momentum transfer $\boldsymbol{q}^2$ to which it reduces when the energy exchanged E_q is zero. Specifically, it can be shown that the QCD coupling constant α_s is given to a good approximation by

$$\alpha_s = \frac{12\pi}{(33 - 2N_f)\ell n(Q^2/\Lambda^2)}$$

for $Q^2 \gg \Lambda^2$, where N_f is the number of quark flavors $u, d, s, \ldots$, with $4m_q^2c^4 < Q^2$, and Λ is a scale parameter

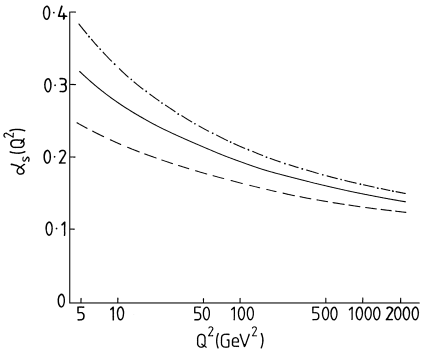


FIGURE 15 The running coupling constant α_s corresponding to four flavors and a scale parameter $\Lambda = 0.2 \pm 0.1$ GeV. The dashed, solid, and dot–dashed curves correspond to $\Lambda = 0.1$, 0.2., and 0.3, respectively.

which must be determined from experiment. This has been done by measuring the coupling constant in a variety of processes giving

$$\Lambda = 0.2 \pm 0.1\,\text{GeV}$$

corresponding to the values of $\alpha_s(Q^2)$ plotted in Fig. 15. Because α_s varies with Q^2, it is often referred to as the *running coupling constant*. However, the variation is small at large Q^2 and over limited Q^2 regions it can often be neglected. In this large Q^2 region, the coupling is sufficiently weak that calculations can be performed with reasonable accuracy by retaining only diagrams of lowest and next-to-lowest order; and sometimes the short-range strong interaction can be neglected to a first approximation, as we shall immediately see.

D. Jets and Gluons

A striking feature of many high-energy particle collisions is the occurrence of jets of hadrons in the final state, like those shown in Fig. 9. These are formed in many reactions, and are the closest thing to quark and gluon tracks we are likely to see. Here we shall illustrate their formation by considering high-energy electron–positron annihilation reactions of the type $e^+ + e^- \rightarrow$ hadrons.

Such reactions have been extensively studied in *colliding beam* experiments, in which high-energy electrons and positrons collide head on, with equal and opposite momenta, so that the total momentum of the hadrons produced must balance out to zero in order to conserve momentum.

1. Quark Jets

In the center-of-mass energy range 15–40 GeV, electron–positron annihilation into hadrons is dominated by

processes like that shown in Fig. 10. These can be regarded as occurring in two stages: a primary electromagnetic process $e^+ + e^- \to q + \bar{q}$, leading to the production of a quark–antiquark pair; followed by a strong interaction process, called *fragmentation*, which converts the high energy $q\bar{q}$ pair into two jets of hadrons. These jets are emitted in opposite directions in the center-of-mass frame in order to conserve momentum, and a typical example of such an event observed in an electron–positron colliding beam experiment is shown in Fig. 9.

The fragmentation process which converts the quarks into hadrons is very complicated, and the composition of the jets—i.e., the numbers and types of particles in the jet and their momenta—varies from event to event. However, the direction of a jet, defined by the total momentum

$$\mathbf{P} = \sum_i \mathbf{p}_i ,$$

where the sum extends over all the particles within the jet, reflects closely the parent quark or antiquark direction. This is because the QCD interaction is relatively weak at very short distances (asymptotic freedom), and the quark and antiquark do not interact strongly until they are separated by a distance r of order 1 fm. At these relatively large distances, only a relatively small momentum can be transferred, and hence the jets which subsequently develop point almost exactly in the initial quark and antiquark directions. That is, the jet angular distribution relative to the electron beam direction reflects the angular distributions of the quark and antiquark in the basic reaction $e^+ + e^- \to q + \bar{q}$, as illustrated in Fig. 16. The latter can be easily calculated, and is indeed in excellent agreement with the observed angular distribution of the jets.

2. Gluon Jets

The dominant process in electron–positron annihilation into hadrons is the formation of two "back to back" jets, as discussed earlier. However, occasionally we would expect a high-momentum gluon to be emitted by the quark or antiquark before fragmentation occurs, in much the same way as a high-energy electron sometimes emits a photon. The quark, antiquark, and gluon then fragment into hadrons, leading to a three-jet event. A computer reconstruction of such an event is shown in Fig. 17. Such events provided the first unambiguous evidence for gluons, since the relative angular distributions of the jets are in good agreement with the theoretical expectation for spin-1 gluons, but are inconsistent with what would be expected if, for example, the third jet originated from a particle of spin zero. The ratio of three-jet to two-jet events can also be calculated, assuming that the third jet is a gluon, because the probability that a quark or antiquark will emit a gluon is determined by the strong coupling α_s, in much the same way that the probability that an electron or positron will emit a photon is determined by the fine structure constant α. The result is in excellent agreement with the measured value, confirming again the correctness of the interpretation.

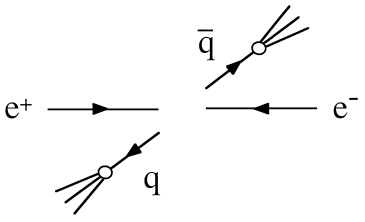


FIGURE 16 Schematic diagram representing two-jet formation in electron–positron annihilation in the center-of-mass frame.

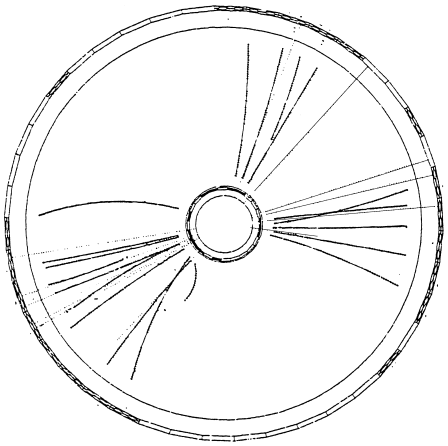


FIGURE 17 Computer reconstruction of a "three-jet" event observed in e^+e^- in annihilation in a drift chamber. (See caption to Fig. 9 for full details.)

3. Color Counting

The *cross-section* σ for a reaction is a measure of the probability of its occurrence. The cross sections for electron–positron annihilation to hadrons and for electron–positron annihilation to muons both decrease rapidly with energy, but their ratio

$$R \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)}$$

is almost energy independent. The near constancy of this ratio follows from the dominance of the two-step mechanism of Fig. 10, with the total annihilation rate being determined by that of the initial reaction $e^+ + e^- \to q + \bar{q}$. The value of the ratio R directly confirms the existence of three color states, each with the same electric charge, for each quark flavor.

To understand this, let us suppose that each quark flavor $f = u, d, s, \ldots$ exists in N_C color states, so that $N_C = 3$ according to QCD, while $N_C = 1$ if the color degree of freedom does not exist. Since the different color states all have the same electric charge, they will all be produced equally readily by the mechanism of Fig. 10, and the rate for producing quark pairs of any given flavor $f = u, d, s, \ldots$ will be proportional to the number of colors N_C. The cross section is also proportional to the squared charge of the produced pair, and since muon pairs are produced by an identical mechanism shown in Fig. 7, we obtain

$$\sigma(e^+e^- \to q\bar{q}) = N_C e_f^2 \sigma(e^+e^- \to \mu^+\mu^-),$$

where e_f is the electric charge on a quark of flavor f. Hence if hadron production were completely dominated by the two-step process of Fig. 10, we would have

$$R = R_0 \equiv N_C\left(e_u^2 + e_d^2 + e_s^2 + e_c^2 + e_b^2\right) = 11N_C/9$$

since the top quark is too heavy to be produced, even at the high energies we are considering. When the small contribution from the three-jet events and other corrections of order α_s are taken into account, this expression for R is modified to

$$R = R_0(1 + \alpha_s/\pi),$$

giving rise to a weak energy dependence from the running of α_s described in Section IV.C. Although these corrections of order α_s are very small compared to the dominant contribution, they must be included if the most precise experimental data on R, which have errors of order 2 or 3%, are to be accounted for. The data are in excellent agreement with the theoretical prediction for the value $N_C = 3$, confirming the basic assumption of QCD. The same result is also obtained from other reactions, although the arguments are more complicated.

V. ELECTROWEAK INTERACTIONS

A. Introduction

Like the strong and electromagnetic interactions, the weak interaction is also associated with elementary spin-1 bosons, which act as "force carriers" between quarks and/or leptons. Until 1973 all observed weak interactions were consistent with the hypothesis that they were mediated by the exchange of heavy charged bosons $W^\pm$ only. However, in the 1960s, a theory was developed which unified electromagnetic and weak interactions in a way which is often compared to the unification of electric and magnetic interactions by Faraday and Maxwell a century earlier. This new theory made several remarkable predictions, including the existence of a heavy neutral vector boson

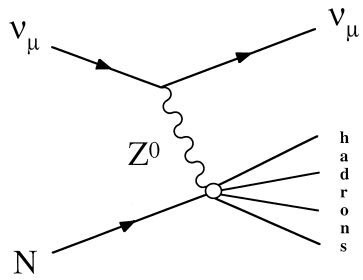


FIGURE 18 Feynman diagram for the weak neutral current reaction $\nu_\mu + N \to \nu_\mu + X$, where X denotes any hadrons allowed by the conservation laws.

Z^0, and of weak reactions arising from its exchange. The latter processes are called *neutral current* reactions to distinguish them from the so-called *charged current* reactions arising from charged $W^\pm$ boson exchange. In particular, neutral current reactions of the type $\nu_\mu + N \to \nu_\mu + X$ were predicted to occur via the mechanism of Fig. 18, where N is a nucleon and X, is any set of hadrons allowed by the conservation laws. Although difficult to detect, such reactions were first observed in 1973.

The prediction of the existence and properties of neutral currents, prior to their discovery, is only one of many spectacular successes of the unified theory of electromagnetic and weak interactions. Others include the prediction of the existence of the charmed quark, prior to its discovery in 1974; and the prediction of the masses of the $W^\pm$ and Z^0 bosons prior to the long-awaited detection of these particles in 1983. In general, the theory is in complete agreement with all data on both weak and electromagnetic interactions, which are now often referred to collectively as the *electroweak interaction*, in the same way that electric and magnetic interactions are referred to collectively as electromagnetic interactions. However, the new unification only becomes manifest at very high energies, and at lower energies weak and electromagnetic interactions can still be clearly separated. Furthermore, the theory predicts the existence of a new scalar boson, the so-called *Higgs boson*, which is associated with the origin of particle masses within the model. Although a possible signal has recently been reported, the evidence for its existence is not conclusive and its detection remains the outstanding experimental problem in the standard model.

B. $W^\pm$ and Z^0 Bosons

There are three "weak intermediate vector bosons," the two charged bosons W^+ and W^- the neutral Z^0. They were all discovered at CERN in 1983 in the reactions

$$\bar{p} + p \to W^+ + X^-, \qquad \bar{p} + p \to W^- + X^+,$$

$$\text{and} \quad \bar{p} + p \to Z^0 + X^0,$$

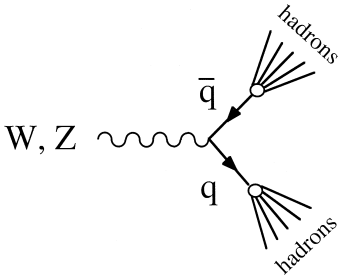


FIGURE 19 Dominant decays of W and Z bosons to jets of hadrons.

where $X^{\pm}$ and X^0 are arbitrary hadronic states allowed by the conservation laws. The $W^{\pm}$ and Z^0 boson masses are measured to be

$$M_W = 80.6\ \text{GeV/c}^2, \qquad M_Z = 91.2\ \text{GeV/c}^2,$$

while the lifetimes are about 3×10^{-25} sec. The dominant decays lead to jets of hadrons via the mechanism of Fig. 19, but the leptonic decays

$$W^+ \to \ell^+ + \nu_\ell, \qquad W^- \to \ell^- + \bar{\nu}_\ell$$

and

$$Z^0 \to \ell^+ + \ell^-, \qquad Z^0 \to \nu_\ell + \bar{\nu}_l,$$

where $\ell = e$, μ or τ as usual, are also important. The W^- decays by the mechanism of Fig. 20, and there are similar diagrams for the other leptonic decays. They all conserve lepton numbers, so that while, for example, $W^- \to e^- + \bar{\nu}_e$ and $Z^0 \to \mu^+ + \mu^-$ are allowed, $W^- \to e^- + \bar{\nu}_\mu$ and $Z^0 \to e^+ + \mu^-$ are not.

In contrast to the zero mass photons and gluons, which mediate the electromagnetic and strong interactions, the $W^{\pm}$ and Z^0 are very massive particles. Correspondingly the interactions resulting from their exchange have ranges

$$R_W \approx R_Z \approx 2 \times 10^{-3}\ \text{fm},$$

which is very small, even when compared to the size of the nucleon (see Section I.E). Examples of such exchange processes are neutrino electron scattering Fig. 5 and tauon decay Fig. 8. The first step in Fig. 8 is the process $\tau \to W + \nu_\tau$, which obviously involves a massive violation of energy conservation of order M_W for tauons at rest, since tauons are much lighter than W bosons. In quantum mechanics an energy violation of order ΔE is allowed, provided it persists only for a time of order $\hbar/\Delta E$, when energy conservation must be restored. Hence in Fig. 8, the second vertex, where energy conservation is restored, must be very close to the first vertex, leading to a very short-range interaction as cited previously.

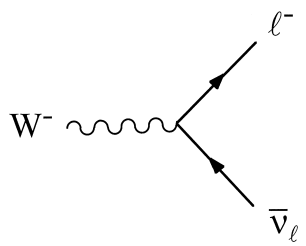


FIGURE 20 Leptonic decay of the W^- boson.

Another important feature of an exchange interaction is its strength. As in the case of electromagnetism (The reader may wish to refer to of Sec. IE at this point.), Feynman diagrams are constructed from fundamental three line vertices, as illustrated in Figs. 5 and 8. At each vertex a boson is emitted or absorbed; while both fermion lines belong to the same generation $\ell = e$, μ or τ, with one arrow pointing inwards and one out to guarantee conservation of each lepton number N_e, N_μ, and N_τ. Finally, associated with each vertex is a dimensionless parameter with the same value

$$\alpha_w = g_w^2/4\pi\hbar c \approx 1/400$$

for all three generations. This constant is the weak analogue of the fine structure constant $\alpha \approx 1/137$ in electromagnetic interactions (see the discussion of Section I.D) with g_W the weak analog of the electronic charge e in appropriate units.

We see from the previous that, despite its name, the weak interaction has a similar intrinsic strength to the electromagnetic interaction. Its apparent weakness in many reactions, which gives the interaction its name, is solely a consequence of its short range, since the particles are only rarely close enough to interact at all. Indeed at energies when the de Broglie wavelengths $\lambda = h/p$ of the particles are large compared to the range of the weak interaction, which is an excellent approximation for all lepton and hadron decays, the range can be neglected altogether. In this approximation the weak interaction becomes a *point* or *zero range* interaction, as illustrated in Fig. 21 for τ decay. The effective strength of this point interaction can be shown to be

$$\alpha_{eff} = \alpha_w(\bar{E}/M_W)^2, \qquad \bar{E} \ll M_W,$$

where $\bar{E}$ is a typical energy scale for the process in question. Thus we see that the interaction is both weak and

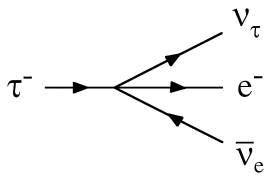


FIGURE 21 Tauon decay in the zero-range or "point" approximation.

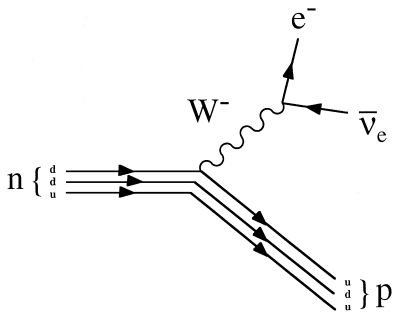


FIGURE 22 Quark diagram for the decay $n \to p + e^- + \bar{\nu}_e$.

very energy dependent at "low energies," but becomes comparable in strength with the electromagnetic interaction at energies on the scale of the W-boson mass.

C. Quarks and Hadrons

The weak decays of hadrons are understood in terms of basic processes in which $W^\pm$ bosons are emitted or absorbed by their constituent quarks. For example, neutron decay is essentially the process

$$d \to u + e^- + \bar{\nu}_e$$

as illustrated in Fig. 22 while in the pion decay process

$$\pi^-(d\bar{u}) \to \mu^- + \bar{\nu}_\mu$$

the initial quarks annihilate to produce a W-boson as shown in Fig. 23. However, the weak interactions of quarks are more complicated than those of leptons, and are best understood in terms of two ideas: *lepton–quark symmetry*, and *quark mixing*. In its simplest form, for two generations, lepton–quark symmetry asserts that the first two generations of quarks

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} s \\ c \end{pmatrix}$$

and the first two generations of leptons

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}$$

have identical weak interactions. That is, one can obtain the basic $W^\pm$-quark vertices by making the replacements $\nu_e \to u$, $e^- \to d$, $\nu_\mu \to c$, $\mu^- \to s$ in the basic $W^\pm$-lepton vertices, leaving the coupling constant g_W unchanged. For example, for the leptons the fundamental processes include $e^- + \bar{\nu}_e \to W^-$ and $\mu^- + \bar{\nu}_\mu \to W^-$, while $e^- + \bar{\nu}_\mu \to W^-$ and $\mu^- + \bar{\nu}_e \to W^-$ are forbidden by lepton number conservation. Quark symmetry in the simple form cited here then implies that the fundamental processes $d + \bar{u} \to W^-$ and $s + \bar{c} \to W^-$ occur with the same couplings g_W as the corresponding leptonic processes, while the processes $s + \bar{u} \to W^-$ and $d + \bar{c} \to W^-$ are forbidden. This works quite well for many reactions, like the pion decay $\pi^- \to \mu^- + \bar{\nu}_\mu$ (Fig. 23). However, many decays which are forbidden in this simple scheme are observed to occur, albeit at a rate which is suppressed relative to the "allowed" decays. An example of this is the kaon decay $K^- \to \mu^- + \bar{\nu}_\mu$, which requires a nonzero $s + \bar{u} \to W^-$ vertex, as illustrated in Fig. 24. All these suppressed decays can be successfully incorporated into the theory by introducing *quark mixing*. According to this idea, the d and s quarks participate in the weak interactions via the linear combinations

$$d' = d\cos\theta_C + s\sin\theta_C$$

and

$$s' = -d\sin\theta_C + s\cos\theta_C,$$

where the parameter θ_C called the Cabibbo angle. That is, lepton–quark symmetry is assumed to apply to the doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c \\ s' \end{pmatrix}$$

This hypothesis enables theory and experiment to be brought into good agreement by choosing a value $\theta_C = 13^\circ$ for the Cabibbo angle. In particular, one then finds that the rates for the previously "allowed" decays occur at a rate which is suppressed by a factor $\cos^2\theta_C \approx 0.95$, while the previously "forbidden" decays are now allowed, but with a rate which is suppressed by a factor $\sin^2\theta_C \approx 0.05$.

Historically, the most remarkable thing about these ideas is that they were formulated before the discovery

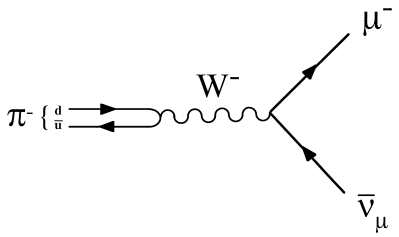


FIGURE 23 Quark diagram for the decay $\pi^- \to \mu^- + \bar{\nu}_\mu$.

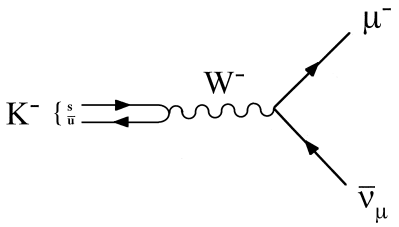


FIGURE 24 Quark diagram for the decay $K^- \to \mu^- + \bar{\nu}_\mu$.

of the charmed quark. In 1971 only seven fundamental fermions were known: the four leptons ν_e, e^-, ν_μ, and μ^-; and the three quarks u, d, s. This led Glashow, Iliopolous, and Maiani to propose the existence of a fourth quark c to complete the lepton–quark symmetry and to solve problems associated with neutral currents. The charmed quark was subsequently discovered in 1974 and its measured weak couplings are consistent with the predictions of lepton–quark symmetry and quark mixing.

Since 1974, events have moved on and there are now six known leptons

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}$$

and six known quarks

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}.$$

When the third generation is taken into account, the mixing scheme becomes more complicated as we must allow for the possibility of mixing between all three "lower" quarks d, s, and b instead of just the first two. This is done by generalizing the equation which defined the quark combinations d' and s' to

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

where $V_{\alpha\beta}$ ($\alpha = u, c, t$; $\beta = d, s, b$) is called the CKM matrix. Its elements are not independent of each other, but can be expressed in terms of three angles, one of which is the Cabibbo angle; and one phase δ, which is of special significance, as we shall see in the following. Lepton–quark symmetry is then applied to the doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix} \quad \begin{pmatrix} c \\ s' \end{pmatrix} \quad \begin{pmatrix} t \\ b' \end{pmatrix},$$

and the CKM matrix elements $V_{\alpha\beta}$ are determined from experimental data.

For the first two generations, the results of this complex scheme lead to very small corrections to the simple "Cabibbo scheme" discussed earlier. However, one of these corrections is the occurrence of *CP violation*, which arises solely, within the standard model, from nonzero values of the phase δ in the CKM matrix. A brief explanation of CP violation is as follows.

In strong and electromagnetic interactions there is symmetry under the parity transformation P, corresponding to a mirror reflection; and under the so-called C-parity transformation, which corresponds to changing all particles into their antiparticles. In weak interactions, by contrast, both these symmetries were found to be violated in 1957, but symmetry under the combined CP transformation seemed to be respected. However, in 1964, Fitch and Cronin discovered a tiny violation, of the order or 1 or 2 parts per 1000, in the decays of neutral kaons. To date no violations in any other systems have been found. However, this may change very soon, since an intense experimental effort is underway to search for similar effects in the decays of neutral mesons containing b quarks. The purpose of this is to discover whether the CP violation which is expected in this system conforms to the predictions of quark mixing; or whether it requires completely new phenomena, beyond the standard model.

D. Neutral Currents and the Unified Theory

Neutral current reactions are those which involve the emission, absorption, or exchange of Z^0 bosons. The unified electroweak theory predicted the existence of such reactions before their experimental discovery in 1973. This theory was proposed mainly in order to solve problems associated with Feynman diagrams in which more than one W boson was exchanged, like that shown in Fig. 25 which contributes to the reaction $e^+ + e^- \rightarrow \mu^+ + \mu^-$. Such contributions are expected to be small because they are higher order in the weak interaction, and this appears to be confirmed by experimental data, which are in good agreement with simple theoretical predictions which neglect them entirely. However, when these contributions are explicitly calculated, they are found to be proportional to divergent integrals, i.e., they are infinite. In the unified theory, this problem is automatically solved when diagrams involving the exchange of Z^0 bosons and photons are taken into account. These also give infinite contributions, but when all the diagrams of a given order are added together the divergences cancel, giving a well-defined and finite contribution overall. This cancellation is not accidental, but is a consequence of a fundamental symmetry relating the weak and electromagnetic interactions, called an SU(2) × U(1) gauge symmetry. We shall not pursue this, since it is complicated, but simply comment on some phenomenological consequences of the theory.

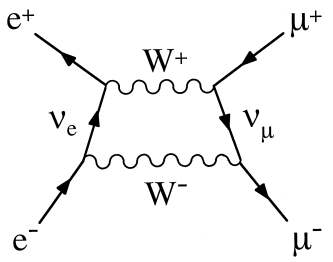


FIGURE 25 Higher-order contribution to the reaction $e^+ + e^- \rightarrow \mu^+ + \mu^-$ from the exchange of two W mesons.

The first is that the theory requires a relation between the weak and electromagnetic couplings, called the *unification condition*. The unification condition is

$$\frac{e}{2\sqrt{2}\varepsilon_0^{1/2}} = g_w \sin\theta_W = g_z \cos\theta_W,$$

where the *weak mixing angle* θ_W is given by

$$\cos\theta_W \equiv M_W/M_Z \qquad (0 < \theta < \pi/2),$$

and g_z is a coupling constant which characterizes the strength of the neutral current vertices. The unification condition relates the strengths of the various interactions to the W and Z masses, and historically was used to predict the latter from the former before the $W^{\pm}$ and Z^0 bosons were discovered.

Second, just as all the charged current interactions of leptons can be understood in terms of the basic $W^{\pm}$-lepton vertices, in the same way all known neutral current interactions can be accounted for in terms of basic Z^0-lepton vertices. The corresponding quark vertices can be obtained from the lepton vertices by using lepton–quark symmetry and quark mixing, in the same way that $W^{\pm}$-quark vertices are obtained from the $W^{\pm}$-lepton vertices. When this is done, the results show that neutral current interactions, like electromagnetic interactions, conserve the quark numbers N_u, N_d, N_s ..., in contrast to the charged current interactions which do not conserve them. This prediction is confirmed by experiment.

Finally, in any process in which a photon is exchanged, a Z^0 boson can be exchanged as well. This is illustrated by the lowest-order diagrams for the particular case of the muon pair production reaction $e^+ + e^- \to \mu^+ + \mu^-$ in Fig. 26. At energies which are small compared to the Z^0 mass, the Z^0-exchange contributions can be neglected compared to the corresponding photon-exchange contributions, and these reactions can be regarded as purely electromagnetic to a high degree of accuracy. However, at very high-energy and momentum transfers, Z^0-exchange contributions become comparable with photon exchange, and we are therefore dealing with genuinely electroweak processes which involve both weak and electromagnetic interactions to a comparable degree.

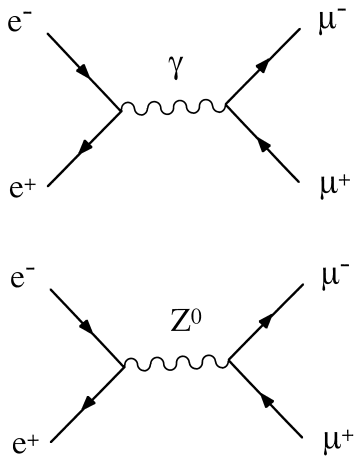


FIGURE 26 Lowest-order contributions to the reaction $e^+ + e^- \to \mu^+ + \mu^-$ from γ and Z^0 exchanges.

These points are beautifully illustrated by the cross section for the muon pair production reaction $e^+ + e^- \to \mu^+ + \mu^-$, where we assume that the energy is large enough for the lepton mases to be neglected. Simple dimensional arguments can be shown to give $\sigma_\gamma \approx \alpha^2/E^2$ for the one-photon exchange contribution and $\sigma_Z \approx G_Z^2 E^2$ for the contribution of the Z^0-exchange diagram at energies $E^2 \ll M_Z^2$, where G_Z is the effective low-energy constant. The ratio of these gives

$$\frac{\sigma_Z}{\sigma_\gamma} \approx \frac{G_Z^2 E^4}{\alpha^2} \approx \frac{E^4}{M_Z^4},$$

where we have neglected factors of order unity.

We thus see that the one-photon exchange diagram dominates at low energies, and the cross-section falls as E^{-2}. This is in agreement with the observed behavior shown in Fig. 27 and justifies our neglect of the Z^0-exchange contribution at low energies. However, the relative importance of the Z^0-exchange contribution increases rapidly with energy and at beam energies of about 25 GeV it begins to make a significant contribution to the total cross section. At still higher energies, the cross section is dominated by a very large peak at an energy corresponding to the Z^0 mass, as illustrated in Fig. 27. At this energy the low-energy approximation is irrelevant and Fig. 27 corresponds to the formation of physical Z^0 bosons in the process $e^+ + e^- \to Z^0$ followed by the subsequent decay $Z^0 \to \mu^+ + \mu^-$ to give the final-state muons. Finally, beyond the peak we once again regain the electroweak

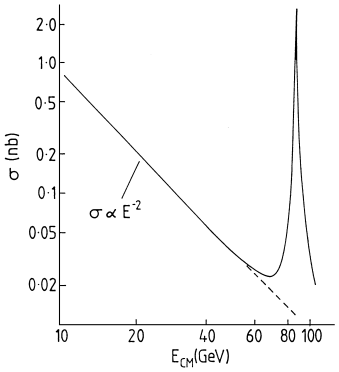


FIGURE 27 Cross section for the reaction $e^+ + e^- \to \mu^+ + \mu^-$ showing the contributions from γ and Z^0 exchanges and the peak at the mass of the Z^0.

regime where both contributions are comparable and neither dominates.

E. The Higgs Boson

The Higgs boson is a neutral spin-0 boson whose existence is predicted by the unified electroweak theory, but which has not yet been observed. It is required because of a fundamental symmetry associated with theories in which the force carriers are spin-1 bosons. This symmetry is called gauge invariance, and it can be shown to require that the spin-1 "gauge bosons" have zero masses if they are the only bosons in the theory. This is fine for QED and QCD, since the gauge bosons are the photons and gluons and they do indeed have zero masses. Gauge invariance also plays an important role in the unified electroweak theory, where it is needed to ensure the cancellation of the divergences which occur in individual Feynman diagrams. In this case the result is even stronger and it can be shown that gauge invariance requires that the fundamental particles—the quarks, leptons, and gauge bosons—*all* have zero masses if gauge bosons are the only bosons in the theory. This problem, called the *origin of mass*, is overcome by assuming that the various particles interact with a new type of field, called the *Higgs field*, whose existence can be shown to have two consequences. The first is that gauge bosons can acquire masses without violating the gauge invariance of the interaction. The second is that there are electrically neutral quanta associated with the Higgs field, called *Higgs bosons*, in the same way that there are quanta associated with the electromagnetic field, i.e., photons.

The existence of the Higgs boson is the most important prediction of the standard model which has not been verified by experiment, and searches for it are a high priority. A problem in designing suitable experiments is that its mass is not predicted by the theory. However, its couplings to other particles *are* predicted, and *are* essentially proportional to the mass of the particle to which it couples. The Higgs boson therefore couples very weakly to light particles like neutrinos, electrons, muons, and u, d, s quarks; and much more strongly to heavy particles like $W^{\pm}$ and Z^0 bosons, and presumably b and t quarks. Hence attempts to produce Higgs bosons are made more difficult by the need to first produce the very heavy particles to which they couple.

The failure to observe Higgs bosons in present experiments leads to limits on their mass. The best results came from the electron–positron collider LEP at CERN. This machine had a maximum energy of 208 GeV, which was enough to produce Higgs bosons with masses up to almost 120 GeV/c^2 in the reaction

$$e^+ + e^- \rightarrow H^0 + Z^0.$$

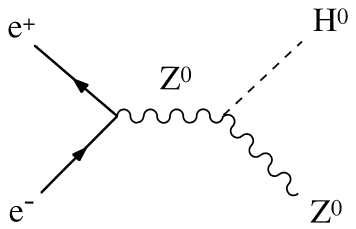


FIGURE 28 Dominant contribution to the production of the Higgs boson.

Assuming that Higgs bosons exist, this reaction is predicted to occur by the mechanism of Fig. 28, while the bosons themselves can be detected via their decay products, which are quite distinctive. For the mass range under consideration, the Higgs boson is predicted to decay initially almost exclusively to $b\bar{b}$, where the quarks would be observed as jets containing short-lived hadrons with nonzero beauty. The Z^0 can be most clearly identified by its decays to electron–positron or muon pairs, but others decays also give useful information.

Thus the previously cited reaction is an ideal method for producing and detecting Higgs bosons. The results are tantalizing. By the time LEP closed down to make way for another project in November 2000, it had shown that no Higgs bosons could exist with a mass less than 113.5 GeV/c^2; and had obtained some evidence for the existence of a Higgs boson with a mass of 115 ± 1 GeV/c^2. Unfortunately, while this signal is statistically much more likely to be a genuine result rather than a statistical fluctuation, the latter cannot be completely ruled out and it may well be several more years before we can be certain whether a Higgs boson of this mass exists or not.

Finally, we note that in the previous discussion we have considered only the simplest case of a single, electrically neutral, Higgs boson, as required by the standard model. However, some interesting extensions of the standard model, discussed in Section VII, require more than one Higgs particle, including electrically charged varieties. Whatever the truth, it is clear that experimental investigation of the Higgs sector will play a central role in the future of particle physics.

VI. EXPERIMENTAL METHODS

In particle physics, experimental projects typically involve hundreds of physicists and take on the order of 10 years to construct, followed by a similar period of data taking. Most, but not all, use machines called *accelerators*, whose sizes are measured in kilometers, to create beams of very high-energy particles. These beams are then directed onto targets, and the interactions between the particles in the beam and in the target are observed. In general, many

different reactions can occur, which typically result in the creation of many new particles. These particles are identified, and their momenta measured, using large multicomponent *detectors*, which supply information at a rate which can only be handled thanks to the development of modern computers.

In what follows we shall not focus on the complexity of accelerators and detector systems, but on the physical principles underlying them.

A. Accelerators and Beams

High-energy particles are needed both to create new particles and to explore the structure of hadrons. The primary sources of such particles are *particle accelerators*. All accelerators use electromagnetic forces to boost the energy of stable charged particles, which are injected into the machine from a high-intensity source of low-energy particles, for example, a proton ion source. Accelerators may conveniently be divided into *linear* and *cyclic* varieties. The former are also known as *linacs*, and at energies above about 1 GeV cyclic machines are usually of a type known as *synchrotrons*.

In a linac, bunches of particles pass through a straight evacuated waveguide with a periodic array of gaps, or cavities. Radio frequency oscillations in the cavities are used to establish a moving electromagnetic wave in the structure, with a longitudinal component of the electric field moving in phase with the particles. So long as this phase relationship can be maintained, the particles will be continuously accelerated. In a synchrotron, the beam of particles is constrained in a circular or near circular path by an array of dipole magnets called bending magnets, which act like optical prisms; and acceleration is achieved as the beam repeatedly traverses one or more cavities placed in the ring. Since the particles travel in a circular orbit they continuously emit radiation, called in this context *synchrotron radiation*, which eventually limits the useful energies that can be achieved using cyclic machines. Stability of the orbit is vital, to ensure that the particles continue to be accelerated, and that they do not strike the sides of the vacuum tube.

In addition to the *primary beams* produced directly by the accelerator, *secondary beams* may also be formed. Typically a primary beam is directed onto a metal target and from the secondary particles produced those of a particular type and momentum are selected. If the secondary particles are unstable, then further beams may be formed from their decay products. Secondary beams can be neutral and/or unstable, provided only that they live long enough to travel appreciable distances in the laboratory. An example is a beam of muons formed from the decays of charged pions.

In addition to the energy of the beam, one is also concerned to produce a beam of high intensity, so that interactions will be plentiful. The intensity is ultimately limited by defocusing effects, e.g., the mutual repulsion of the particles in the beam, and a number of technical problems have to be overcome which are outside the scope of this brief account.

Both linear and cyclic accelerators can be subdivided into *fixed-target* and *colliding beam* machines. The latter are also known as *colliders*, or in the case of cyclic machines, *storage rings*. In fixed-target machines, particles are accelerated to the highest operating energy and then the beam is extracted from the machine and directed onto a stationary target, which is usually a solid or liquid. The intensity of the beam is such that large numbers of interactions can be produced, which can either be studied in their own right or used to produce secondary beams.

The main disadvantage of fixed-target machines becomes apparent with the need to work at high energies in the center-of-mass reference frame, in which the momenta of all the particles add up to zero. The center-of-mass energy is important because it is a measure of the energy available to create new particles. In the laboratory frame, defined as the reference frame in which the target is stationary, at least some of the final state particles must be in motion to conserve momentum. Consequently, at least some of the initial beam energy must reappear as kinetic energy of final state particles and is unavailable for particle production. In contrast, in the center-of-mass frame the total momentum is zero, and in principle all the energy is available for particle production. In a fixed-target experiment, the total center-of-mass energy is given by

$$E_{CM} = \left[m_b^2 c^4 + m_t^2 c^4 + 2 m_t c^2 E_L\right]^{1/2},$$

where $m_b(m_t)$ is the mass of the beam (target) particle, and E_L is the energy of the beam particle in the laboratory. At high energies this increases only as $E_L^{1/2}$ and most of the beam energy is unavailable for particle production.

In a colliding-beam accelerator, two beams of particles traveling in almost opposite directions are made to collide at a small or zero crossing angle. For simplicity, if we assume that the colliding particles have the same mass and collide at zero crossing angle with the same energy E_L, the total center of mass energy is

$$E_{CM} = 2E_L,$$

which increases linearly with the energy of the accelerated particles E_L, and is a significant improvement on the fixed-target result. Colliders are not, however, without their own disadvantages. The colliding particles have to be stable, which limits the interactions which can be studied, and the collision rate in the intersection region is generally smaller

than that achieved in fixed-target experiments, because the beam densities are low compared to a solid or liquid target.

In a *fixed-target* experiment the target is stationary in the laboratory. Alternatively, the target may itself be moving in the form of another beam of particles, in which case we have a *colliding beam* experiment. Finally, details of the particles produced in the collision (e.g., their momenta) are deduced by observing their interactions with the material of *detectors*, which are placed in the vicinity of the interaction region.

B. Particle Interactions with Matter

In order to be detected a particle must undergo an interaction with the material of a detector. The first possibility is that the particle interacts with an atomic nucleus. For example, this could be via the strong nuclear interaction if it is a hadron, or by the weak interaction if it is a neutrino. Both are *short-range interactions*. If the energy is sufficiently high, new particles may be produced, and such reactions are often the first step in the detection process. In addition to these short-range interactions, a charged particle will also excite and ionize atoms along its path, giving rise to *ionization energy losses*, and emit radiation, leading to *radiation energy losses*. Both of these processes are due to the long-range electromagnetic interaction. They are important because they form the basis of many detectors for charged particles. Photons are also directly detected by electromagnetic interactions, and at high energies their interactions with matter lead predominantly to the production of e^+e^- pairs.

For hadrons, the most important short-range interactions with nuclei are due to the strong nuclear force, which, unlike the electromagnetic interaction, is as important for neutral particles as for charged ones. Neutrinos and antineutrinos also have short-range interactions with nuclei, but because such processes are weak interactions, neutrinos can penetrate huge thickness of matter without detection. Nonetheless, in the absence of other possibilities such reactions are the basis for detecting neutrinos.

Ionization energy losses are important for all charged particles, and for particles other than electrons and positrons they dominate over radiation energy losses at all but the highest attainable energies. The losses are due dominantly to Coulomb scattering from the atomic electrons. The rate of loss of energy with distance traversed falls rapidly as the velocity increases from zero. All particles have a region of "minimum ionization" after which it rises slowly. The magnitude of the energy loss is proportional to the density of the medium and to the squared charge of the particle.

When a charged particle traverses matter it can also lose energy by radiative collisions, especially with nuclei. The electric field of a nucleus will accelerate and decelerate the particles as they pass, causing them to radiate photons, and hence lose energy. This process is called *bremsstrahlung* (literally "braking radiation" in German) and is a particularly important contribution to the energy loss for electrons and positrons. A detailed calculation shows that for relativistic electrons, the average rate of energy loss is given by

$$-\frac{dE}{dx} = \frac{E}{L_R},$$

where the constant L_R, called the *radiation length*, is the average thickness of material which reduces the mean energy of an electron or positron by a factor $e = 2.718$. At high energies the radiation losses are proportional to E/m_P^2, for an arbitrary charged particle of mass m_P and completely dominate the energy losses for electrons and positrons at high enough energies. On the other hand, radiation losses are much smaller than ionization losses for all particles other than electrons and positrons at all but the highest energies.

In contrast to heavy charged particles, photons have a high probability of being absorbed or scattered through large angles by the atoms in matter. Several processes contribute to the absorption, but at high energies pair production in the field of a nucleus dominates.

C. Particle Detectors

To be useful, particle detection must be done with a resolution sufficient to enable particles to be separated in both space and time in order to determine which are associated with a particular event. We also need to be able to identify each particle and measure its energy and momentum. No single detector is optimal with respect to all these requirements, and in practice experiments frequently use several different types of detector in combination. At both fixed-target machines and colliders, the modern trend is to build very large multi-component detectors which integrate many different subdetectors in a single device. Such systems rely heavily on fast electronics and computers to monitor and control the subdetectors, and to coordinate, classify, and record the vast amount of information flowing in from different parts of the apparatus.

Most detector systems need a *timing device* and this is commonly provided by the *scintillation counter*. In materials called "scintillators," a small fraction of the energy of excited atomic electrons produced by the passage of a charged particle re-emerges as visible light during deexcitation. In a scintillation counter this light passes down the scintillator and is directed onto the face of a photomultiplier by multiple internal reflections along a shaped solid plastic tube called a light guide, and the whole assembly

is made light-tight to prevent background light reaching the photomultiplier tube. Electrons, emitted from the cathode of the photomultiplier by the photoelectric effect, are amplified to give an electronic pulse, which is very short because of the very short decay time of the scintillator. The scintillation counter is thus an ideal timing device and is widely used for "triggering" other detectors, i.e., its signal is used to decide whether or not to activate other detectors, and whether to record information from the event. Commonly used scintillators are inorganic single crystals (e.g., sodium iodide) or organic liquids and plastics, and a modern complex detector may use several tons of detector in combination with thousands of photomultiplier tubes.

The determination of *position* is based on detecting ionization, either by collecting the total ionization products onto electrodes using an electric field or by making the ionization track visible in some form. One common device which does this is the *drift chamber*. This is based on the observation that if an electric field is established in a gas, then the electrons released as part of electron–ion pairs by the passage of a charged particle will drift toward the anode. If the field is strong enough, an electron will gain sufficient energy to cause secondary ionization, and a chain of such processes leads to an avalanche of secondary electrons which can be collected as a pulse on the anode. The fact that the liberated electrons take time to drift from their point of production to the anode implies that there is a time delay between the passage of a charged particle through the chamber and the creation of a pulse at the anode and this is related to the distance between the particle trajectory and the anode wire. In practice, a reference time has to be defined, which, for example, could be done by allowing the particle to pass through a scintillator positioned elsewhere in the experiment. The electrons drift for a time and are then collected at the anode, thus providing a signal that the particle has passed. If the drift time can be measured accurately and the drift velocity is known, then excellent spatial resolution may be obtained. Arrays of drift chambers, or other devices which measure position, are called *track chambers* because they can be used to reconstruct the trajectories of any charged particles which pass though the chamber.

Another detector of position is the semiconductor detector, which is essentially a solid-state ionization chamber, with electron–hole pairs playing the role of electron–ion pairs in a gaseous detector. In the presence of an electric field, the electrons and holes separate and collect at the electrodes, giving a signal proportional to the energy loss of the incident charged particle. For example, in a silicon microstrip detector, narrow strips of active detector are etched onto a thin slice of silicon. Arrays of such strips can then be used to form detectors with superb spatial resolution. They can, for example, be placed close to the interaction vertex in a colliding beam experiment, with a view to studying events involving the decay of very short-lived particles.

The *momentum* of a charged particle is usually determined from the curvature of its track in an applied magnetic field. It is common practice to enclose track chambers in a magnetic field to perform momentum analysis. An apparatus which is dedicated to measuring momentum is called a *spectrometer*.

Methods of *identification* are often based on determining the mass of the particle by simultaneous measurements of its momentum together with some other quantity. At low velocities, measurements of the rate of energy loss dE/dx can be used, while muons may be characterized by their unique penetrating power. Other methods concentrate on measuring the velocity or energy, assuming always that the momentum is known.

One important identification method for high-energy particles is based on the Cerenkov effect. When a charged particle with velocity u traverses a dispersive medium of refractive index n, excited atoms in the vicinity of the particle become polarized, and if u is greater than the speed of light in the medium c/n, a part of the excitation energy reappears as coherent radiation, called *Cerenkov radiation*. This phenomenon is the optical analogue of the familiar shock wave heard when an aircraft "breaks the sound barrier." The radiation is emitted at a characteristic angle θ to the direction of motion given by

$$\cos\theta = c/nu.$$

A determination of θ is thus a direct measurement of the velocity.

Finally, detectors called *calorimeters* measure the *energy* (and position) of a particle by its total absorption. They differ from most other detectors in that the nature of the particle is changed by the detector, and in that they can detect neutral as well as charged particles. A calorimeter may be a homogeneous absorber/detector, such as a block of lead glass used to detect photons by Cerenkov radiation. Alternatively, it can be a sandwich construction with separate layers of absorber (e.g., a metal such as lead) and detector (e.g., a scintillator). During the absorption process the particle will interact with the material of the absorber, generating secondary particles which will themselves generate further particles and so on, so that a cascade or shower, develops. Eventually all, or almost all, of the primary energy is deposited in the calorimeter, and gives a signal in the detector part of the device. One of the reasons why calorimeters are important, especially at high energies is that they can detect neutral particles, by detecting the charged secondaries. The relative precision of energy measurements is also much better than

that obtained by high-energy spectrometers and the signal produced can be very fast.

The previously cited are by no means a complete set of the detectors currently in use. However, they illustrate the variety of information available from different detectors, with each type of detector being suited for a particular sort of information—time, position, momentum, energy or speed—for particular classes of particle. Their full power is only realized when they are used in combination, in multi-component detector systems, to analyze any and all of the particles produced in a particular event.

VII. BEYOND THE STANDARD MODEL

The standard model has been phenomenally successful. Nonetheless, many questions remain unanswered, and there could well be phenomena awaiting discovery, especially at higher energies, which are not described by the standard model. In this section we will briefly discuss some possibilities currently being investigated.

A. Grand Unification

With the success of the unified theory of electroweak interactions, it became natural to ask whether the strong interaction could also be incorporated with the weak and electromagnetic interactions into a single so-called grand unified theory. The much greater strength of the strong interaction compared to the electroweak at presently accessible energies would seem to make this a hopeless task. However, the strength of an interaction depends on the distance over which it acts, or more precisely on the magnitude of the energy–momentum transfer Q^2 (see Fig. 15). In particular, the strong interaction coupling decreases with Q^2, while the electroweak couplings vary much more slowly, and a naive extrapolation from their low-energy values suggests that the various couplings might become equal at an enormous value $Q^2 = M_X^2 c^4$, where the so-called unification mass M_X is of order 10^{15} GeV/c^2. This is illustrated in Fig. 29, where g_s is related to the QCD coupling constant α_s by

$$\alpha_s = g_s^2/4\pi$$

and the electroweak couplings are related to those defined in Sections V.B and D by

$$g = 2\sqrt{2}g_w, \qquad g' = 2\sqrt{2}g_z.$$

In grand unified theories all three interactions are united into a single interaction, characterized by a single coupling constant, at the unification mass; differences between them emerge as one interpolates downward to currently available energies. Of course, this interpolation assumes

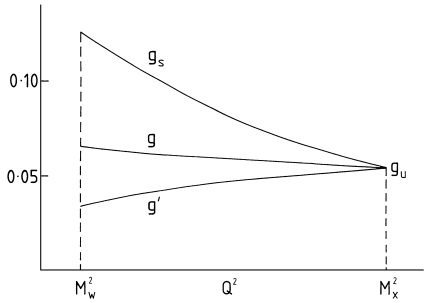


FIGURE 29 The behaviour of the strong (g_s) and electroweak (g, g') couplings as functions of the squared energy–momentum transfer Q^2 in a typical grand unified theory.

that nothing totally unexpected will happen between energies on the order of 10^2 and 10^{15} GeV which could spoil the predictions.

There are many ways in which grand unified theories may be constructed so that they contain the predictions of both quantum chromodynamics and the unified electroweak theory at currently attainable energies. The simplest grand unified model is the so-called SU(5) model, which incorporates quarks and leptons into common families. For example, in the standard model the three color states of the down quark, which are conveniently denoted (d_r, d_g, d_b) can be converted into each other by gluon emission as illustrated in Fig. 30a; while the positron and antineutrino $(e^+, \bar{\nu}_e)$ can be converted into each other by $W^\pm$ emission, as illustrated in Fig. 30b. In the SU(5) model both sets of particles are associated into a single family $(d_r, d_g, d_b, e^+, \bar{\nu}_e)$ with five members, and the quarks and leptons can convert into each other by processes like

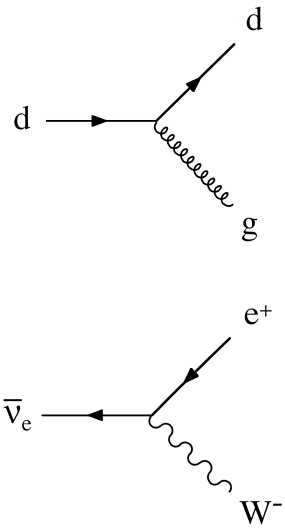


FIGURE 30 Two processes which occur by gluon and W-boson emission within the family of particles (d_r, d_g, d_b, e^+, $\bar{\nu}_e$).

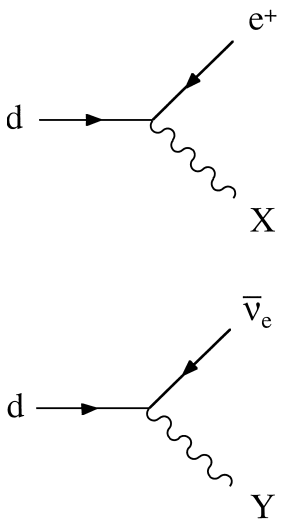


FIGURE 31 Two processes which occur by emission of new particles X and Y, predicted by grand unification schemes, within the family of particles (d_r, d_g, d_b, e^+, $\bar{\nu}_e$).

those shown in Fig. 31. These involve the emission of two new gauge bosons X and Y with electric charges $-4/3$ and $-1/3$, respectively, and masses of order $M_X \approx 10^{15}$ GeV/c^2. At the unification mass all these processes are characterized by a single "grand unified coupling constant" g_U, whose value is such that the analogue of the fine structure constant is

$$\alpha_U \equiv \frac{g_U^2}{4\pi} \approx \frac{1}{42}$$

However, at energies $E \ll M_X$, processes involving the exchange of the X and Y bosons are heavily suppressed because of their large masses, in the same way that $W^{\pm}$-exchange processes are suppressed relative to electromagnetic ones at energies $E \ll M_W$ in the unified electroweak theory. Because of this, processes involving the exchange of X and Y particles are difficult to observe at presently attainable energies.

An attractive feature of the model is that it offers an explanation of one of the longest standing mysteries in physics—the equal magnitudes of the electric charges of the electron and proton. It can be shown that the sum of the electric charges of all the particles in any given family must be zero. This gives $3q_d + e = 0$, where q_d is the charge of the down quark. Hence q_d is determined to be $-e/3$ and the mysterious factor of three is seen to be a consequence of the fact that the quarks have three distinct color states. The charge of the up quark q_u is shown to be $2e/3$ by a similar argument, and the equality of the proton charge

$$q_p = 2q_u + q_d = e$$

and the positron charge follows from the usual quark assignment $p = uud$.

The model also gives a prediction for the value of the weak mixing angle. Since the strong and electroweak interactions, when extrapolated to the unification mass M_X, are characterized by a single coupling constant g_U, the three effective low-energy couplings of the standard model can be expressed in terms of the two parameters, g_U and M_X, by reversing this extrapolation. Consequently, in grand unified theories, one of the three low-energy coupling constants can be predicted, given the values of the other two. It is conventional to convert this result into a prediction of the weak mixing angle θ_W, which leads to the value $\sin^2\theta_W = 0.21$, which is close to, but nonetheless not in complete agreement with, the measured value 0.23. We will return to this shortly.

The most striking prediction of grand unified theories is that the proton is unstable. Within the standard model, the stability of the proton is guaranteed by baryon number conservation, as we saw in Section III.E. However, in grand unified theories quarks can turn into antileptons by processes like those shown in Fig. 31, enabling protons to decay by a variety of processes involving the exchange of X and Y bosons and their antiparticles $\bar{X}$ and $\bar{Y}$. Examples are the decays and $p \to \pi^0 + e^+$ and $p \to \pi^+ + \bar{\nu}_e$. Although unstable, the predicted lifetime of the proton decaying via these modes is very long. In particular it is predicted to be of the order of $\tau = 10^{29}$–10^{30} years in the SU(5) grand unified model. To detect proton decays with lifetimes of this order requires a very large mass of detector material. For example, 300 tons of iron would yield only one proton decay per day if the lifetime was of the order 10^{32} years. Several large detectors of various types have been built and the best experimental result is $\tau > 3 \times 10^{32}$ years[3] which is clearly incompatible with the predicted lifetime. (This value assumes that any decays will occur by the modes predicted by grand unified theories. Without this assumption, the limit is reduced to $\tau > 1.6 \times 10^{25}$.) However, other grand unified models predict longer lifetimes than the simplest model and are compatible with the proton decay data, and also with the measured value of the weak mixing angle, as we shall see in the following.

B. Supersymmetry

The most popular grand unified theories incorporate a new proposed symmetry of nature called *supersymmetry*. According to this, every known elementary particle has a supersymmetric partner, or superpartner, which is like it in all respects except for its spin. In the theory, spin-1/2 fermions, leptons and quarks, have spin-0 superpartners, while spin-1 bosons, like photons, have spin-1/2

TABLE VI Particles and Their Superpartners

Particle	Symbol	Spin	Superparticle	Symbol	Spin
Quark	q	1/2	Squark	$\tilde{q}$	0
Electron	e	1/2	Selectron	$\tilde{e}$	0
Muon	μ	1/2	Smuon	$\tilde{\mu}$	0
Tauon	τ	1/2	Stauon	$\tilde{\tau}$	0
Neutrino	ν	1/2	Neutralino	$\tilde{\nu}$	0
W	$W^{\pm}$	1	Wino	$\tilde{W}^{\pm}$	1/2
Z	Z^0	1	Zino	$\tilde{Z}^0$	1/2
Photon	γ	1	Photino	$\tilde{\gamma}$	1/2
Gluon	g	1	Gluino	$\tilde{g}$	1/2
Higgs	H	0	Higgsino	$\tilde{H}$	1/2

superpartners. The supersymmetric partners of fermions are named by adding a prefix "s" to the name of the fermion, while the superpartners of the bosons are named by adding the suffix "ino" to the root of the normal name. This is illustrated in Table VI, where we list the various particles, together with their superpartners and the corresponding spins.

If supersymmetry were exact, a particle and its superpartner would have exactly the same mass. This is obviously not realized in nature or superparticles would have been detected long ago, so supersymmetry is at best only an approximation. This is assumed to be the case in supersymmetric versions of grand unified theories in which even the lightest supersymmetric particles have masses which are of the same order of magnitude as the $W^{\pm}$ and Z^0 masses. When these supersymmetric particles are taken into account, it can be shown that the extrapolation of coupling constants is modified in such a way that the grand unification mass M_X is increased to a value of order 10^{16} GeV/c^2, while the value of the grand unified coupling constant g_U remains relatively constant. As a consequence of this, the proton lifetime increases to a value of order 10^{32}–10^{33} years, which is consistent with present experimental limits. At the same time, the prediction of the weak mixing angle is slightly modified to yield a value in precise agreement with the experimental value.

Supersymmetry is an important component in even more ambitious schemes to unify gravity with the other forces of nature at an enormous "superunification" mass of order 10^{19} GeV/c^2, called the Planck mass. Such theories often replace the point-like elementary particles we have considered by tiny quantized "strings," in order to reduce the problems encountered in quantizing the gravitational interaction. However, while such superunification schemes have aroused considerable interest, they have had little or no impact on experimental particle physics.

To verify supersymmetry it is of course necessary to detect the superparticles. According to most versions of the theory, these can only be created or destroyed in pairs, so that the decay of a superparticle must yield at least one superparticle in the final state, and the lightest such particle must be stable. There are several candidates for the identity of the lightest superparticle, but most models assume it is the photino. If we accept this, then a simple reaction which could be studied is $e^+ + e^- \rightarrow \tilde{e}^+ + \tilde{e}^-$ followed by the decays $\tilde{e}^{\pm} \rightarrow e^{\pm} + \tilde{\gamma}$, giving an overall reaction

$$e^+ + e^- \rightarrow e^+ + e^- + \tilde{\gamma} + \tilde{\gamma}.$$

Here we have identified the superparticles by placing tildas over the symbols for the corresponding particles, so that $\tilde{e}$ is a selectron and $\tilde{\gamma}$ indicates a photino. Current experiments at hadron colliders have failed to detect any supersymmetric particles and this sets lower limits on their masses, which vary somewhat for different particles, but which are typically on the order of order 50 to 10 GeV/c^2. This is of limited significance since their masses are expected to be of order M_W or more, and the decisive search for supersymmetry will probably take place at the higher-energy machines that will be available in the next few years.

C. Neutrino Oscillations

Neutrinos are usually assumed to have zero mass in the standard model, although the model can be easily extended to accommodate small, nonzero masses. One phenomenon which can occur if neutrinos have nonzero masses is *neutrino mixing*. This is directly analogous to quark mixing. We saw in Section V.C that in weak interactions the u and c quarks couple in linear combinations. In the same way, the neutrino states ν_e and ν_μ, which couple to electrons and muons, respectively, could be linear combinations

$$\nu_e = \nu_1 \cos\alpha + \nu_2 \sin\alpha$$

and

$$\nu_\mu = -\nu_1 \sin\alpha + \nu_2 \cos\alpha$$

of neutrinos ν_1 and ν_2 with masses m_1 and m_2. Here α is a mixing angle analogous to the Cabibbo angle θ_C and like the Cabibbo angle it must be determined from experiment. This can be done in principle by studying the phenomenon of *neutrino oscillation*. When, for example, an electron neutrino is produced with momentum p at time $t = 0$, the ν_1 and ν_2 components will have slightly different energies E_1 and E_2 due to their slightly different masses. In quantum mechanics, their associated waves will therefore have slightly different frequencies, giving rise to phenomena somewhat akin to the "beats" heard when two sound waves of slightly different frequency are

superimposed. As a result of these, one finds that the original beam of electron neutrinos develops a muon neutrino component whose intensity oscillates as it travels through space, while the intensity of the neutrino electron beam itself is correspondingly reduced. Specifically, the probability of finding a muon neutrino is calculated to be

$$P(\nu_e \to \nu_\mu) = \sin^2(2\alpha)\sin^2[(E_2 - E_1)t/2],$$

while the probability of finding an electron neutrino is reduced by a corresponding oscillating factor, so that $P(\nu_e \to \nu_e) = 1 - P(\nu_e \to \nu_\mu)$. Similar effects are predicted if instead we start from muon neutrinos. In both cases the oscillations vanish if the mixing angle α is zero, or if the neutrinos have equal masses, and hence equal energies, as can be seen explicitly. In particular, such oscillations are not possible if the neutrinos both have zero masses. More generally, one can allow mixing between all three types of neutrino ν_e, ν_μ and ν_τ, so that an electron neutrino can oscillate into either a muon or a tauon neutrino or both, and so on. Again the phenomenon is only possible if at least some of the neutrino species have nonzero masses.

Attempts to detect neutrino oscillations rest on the fact that electron neutrinos can produce electrons via reactions like $\nu_e + n \to e^- + p$, but cannot produce muons or tauons; whereas muon neutrinos can produce muons, via reactions like $\nu_\mu + n \to \mu^- + p$, but not electrons or tauons. In addition, the time t is determined by the distance of the neutrino detector from the source of the neutrinos, since their momenta are always much greater than their possible masses, and they travel at approximately the speed of light. Hence, for example, if we start with a beam of muon neutrinos formed in pion decays, the yield of electrons and/or muons observed in a detector should vary with its distance from the source of the neutrinos, if appreciable oscillations occur.

It has long been realized that neutrino oscillations offer a possible solution to the *solar neutrino problem*. Several of the reactions which generate energy in the sun also produce electron neutrinos. These are emitted from the sun and can be detected on earth. The measured neutrino flux from several experiments is found to be a factor of 2–3 times smaller than that predicted by the accepted model of the Sun. This discrepancy constitutes the "solar neutrino problem" and is important because it implies that there is either something wrong with our theoretical understanding of stars like the Sun, which is highly successful in other respects; or with the properties of neutrinos assumed in the standard model.

Neutrino oscillations could account for the solar neutrino problem by converting some of the emitted electron neutrinos into muon or tauon neutrinos, which would not be detected by present experiments. This is an attractive possibility, because it can be shown that such oscillations can be greatly enhanced when the neutrinos traverse long distances in matter, such as from the interior of the sun, where they are produced, to the sun's surface. Thus oscillations could be important in the solar neutrino problem, even though attempts to detect them in accelerator experiments have so far proved inconclusive. However, this is rather circumstantial evidence for neutrino oscillations, and it remains important to detect them more directly.

In 1998, clear evidence for the existence of neutrino oscillations was obtained from observations on *atmospheric neutrinos* by the giant *Super Kamiokande* detector in Japan. When cosmic ray protons collide with atoms in the upper atmosphere, they create many pions, which in turn create neutrinos mainly by the decay sequences

$$\pi^- \to \mu^- + \bar{\nu}_\mu, \qquad \pi^+ \to \mu^+ + \nu_\mu$$

and

$$\mu^- \to e^- + \bar{\nu}_e + \nu_\mu, \qquad \mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu.$$

From this, one sees that one would naively expect to see two muon neutrinos for every electron neutrino detected. (There are small corrections to this prediction due to other, rarer reactions and due to effects of the Earth's magnetic field, which we ignore here for simplicity.) However, the ratio was observed to be about 1.3 to 1 on average, suggesting that the muon neutrinos produced might be oscillating into other species. Clear confirmation for this was found by exploiting the fact that the detector measured the direction of the detected neutrinos to study the azimuthal dependence of the effect. In particular, one can compare the measured flux from neutrinos produced in the atmosphere directly above the detector, which have a short flight path before detection; and those incident from directly below, which have traveled a long way through the earth before detection, and so have had plenty of time to oscillate. Experimentally, it was found that the yield of electron neutrinos from above and below were the same within errors and consistent with expectation for no oscillations. However, while the yield of muon neutrinos from above accorded with the expectation for no significant oscillations, the flux of muon neutrinos from below was a factor of about 2 lower. This is rather clear evidence for muon neutrino oscillations, presumably into tauon neutrinos[5] which, for the neutrino energies concerned, cannot be detected by Super Kamiokande. (Or alternatively some other previously unknown neutrino, which does not have conventional weak interactions, and so would have escaped detection in Z^0 decays (see the end of Section II.E).)

The existence of neutrino oscillations, and by implication nonzero neutrino masses, is now generally accepted on the basis of the previously cited and other evidence.

However, the details, including the values of the neutrino mass differences and the various mixing angles involved, remain to be resolved.

D. Dark Matter

Nonzero neutrino masses have implications for cosmology. Modern cosmology is based on the "big bang model," according to which the universe is composed of particles which were mostly created in the aftermath of the explosion with which the universe began. This model predicts that the present density of neutrinos in the universe should be comparable with the density of photons if the neutrino masses are much smaller than 1 MeV/c^2. Since the photon density is of order 10^9 times the nucleon density, this implies that neutrinos whose masses are more than $10^{-9}\ m_p \sim 1$ eV/c^2 would make a significant, possibly even dominant, contribution to ρ, the energy density of the universe. We define the $\Omega \equiv \rho/\rho_c$ as the density relative to ρ_c, the critical density below which the expansion continues for ever, and above which it will eventually halt and the Universe will then start to contract. Explicity,

$$\rho_c = \frac{3H_0^2}{8\pi G} = O(10^{-24})\,\mathrm{kg\,m^{-3}},$$

where G is Newton's gravitational constant and H_0 is Hubble's constant. Observations of the "luminous" matter in the universe, i.e., objects which emit electromagnetic radiation, suggest that $\Omega_L \approx 0.01$; whereas in the most popular version of the big bang model, called the inflationary big bang model, the density is assumed to be very close to the critical density, so that the relative density $\Omega \equiv \rho/\rho_c = 1$ to a very good approximation. The observations can only be reconciled with the prediction of the inflationary big bang model if in addition there exists a substantial amount of nonluminous "dark matter" which gives a large contribution Ω_D to Ω such that $\Omega = \Omega_L + \Omega_D \approx \Omega_D$.

The existence of very large quantities of dark matter is not in serious doubt and there is evidence from several cosmological sources. However, the nature of dark matter is unknown and it is likely that it has several components, one of which could be neutrinos with nonzero masses. Neutrinos would be examples of "hot dark matter"—particles which are light enough to have been highly relativistic in the early stages of the evolution of the universe. However, if hot dark matter is assumed to be the dominant form of dark matter it would give rise to a very uniform energy distribution in space, and calculations suggest that the observed galaxies could not have formed in the time available since the big bang, if this were so. This problem is avoided if the dark matter consists primarily of "cold dark matter"; that is, of particles which are heavy enough to have been nonrelativistic at an early stage in the universe's evolution. Cold dark matter, like ordinary baryonic matter, will tend to clump together under gravitational attraction, making galaxy formation easier. Hence, the most popular hypothesis at present is that dark matter consists of nonbaryonic cold dark matter in the form of weakly interacting massive particles, or WIMPS. The problem is that there are no known candidates for WIMPS, so new types of particle have to be postulated. One possibility is the lightest particle which appears in supersymmetric theories. In most such theories this is a stable neutralino which interacts only via electroweak interactions. Its mass is not precisely predicted, but is restricted by experiment to be greater than 30 GeV/c^2, since otherwise it would have been detected. If such theories are correct, these particles would be produced in large numbers following the big bang, and could well account for the bulk of dark matter. This would in turn imply that the Earth, on its journey through space, is passing through a sea of WIMPS, and experiments to detect them are currently being developed at several sites around the world.

SEE ALSO THE FOLLOWING ARTICLES

ACCELERATOR PHYSICS AND ENGINEERING • ATOMIC PHYSICS • COLLIDER DETECTORS FOR MULTI-TEV PARTICLES • COSMOLOGY • DENSE MATTER PHYSICS • INTERACTING BOSON MODEL • NEUTRINOS • NUCLEAR PHYSICS • PARTICLE ACCELERATORS • PROTON DECAY • SPACE PLASMA PHYSICS • UNIFIED FIELD THEORIES

BIBLIOGRAPHY

Close, F., Marten, M., and Sutton, C. (1987). "The Particle Explosion," Oxford University Press, Oxford, England.

Griffiths, D. (1987). "Introduction to Elementary Particles," Wiley, New York.

Halzen, F., and Martin, A. D. (1984). "Quarks and Leptons," Wiley, Chichester, England.

Kane, G. L. (1987). "Modern Elementary Particle Physics," Addison-Wesley, Reading, MA.

Martin, B. R., and Shaw, G. (1997). "Particle Physics," 2nd edition, Wiley, Chichester, England.

Particle Data Group web site: http://pdg.lbl.gov.

Perkins, D. H. (2000). "Introduction to High Energy Physics," 4th edition, Cambridge University Press, Cambridge, England.

Rollnick, W. B. (1994). "The Fundamental Particles and their Interactions," Addison-Wesley, Reading, MA.



Proton Decay

David B. Cline
University of California, Los Angeles

I. Introduction
II. Expectations and Theory of Proton Decay
III. Experimental Detectors for Proton Decay
IV. The Next Step in the Search for Nucleon Decay
V. Significance of the Limits on Nucleon Lifetime

GLOSSARY

Antiproton Particle with a –1 charge that will annihilate on contact with a proton and is the antiparticle of the proton.
Baryon number Property or quantum number assigned to the proton or neutron.
Electroweak theory Theory that includes the unification of the weak and electromagnetic interactions at the energy of the intermediate vector boson mass $\sin^2\theta w$—the parameter of the electroweak theory.
Grand unified theory (GUT) Any theory that postulates that the strengths of the weak, electromagnetic, and strong interactions are the same at the unification energy ($\sim 10^{15}$ GeV).
Lepton Electron-like particle with integral charge (e, μ, τ) or zero charge (neutrinos).
Pati–Salam theory Unified interactions theory that postulates that individual quarks decay inside the nucleus.
Quark Fundamental building block of protons and neutrons, which comes with $+\frac{2}{3}$ and $-\frac{1}{3}$.
Supersymmetry Theory that helps explain why there are large differences in the mass scales of ordinary matter and the GUT unification energy ($\sim 10^{15}$ GeV). It postulates the existence of new supersymmetric particles that double the total number of elementary particles. The new particles share unusual spin properties. Supersymmetry is usually expected in theories of superstrings.

THE TERM PROTON DECAY refers to the disintegration of the proton into particles with smaller mass and very likely the same total charge as the proton or bound neutron. This process, if observed, would violate one of the fundamental principles of elementary particle physics and may be linked to the fact that the universe appears to hold more protons than antiprotons.

I. INTRODUCTION

Of all the unsolved problems in science one of the most intriguing is the *stability* of the proton. There is no well-established physical principle that maintains this stability. An equally profound mystery is the apparent excess of matter over antimatter in the universe. Both proton stability and an excess of matter are required for life to

have formed in the universe. There may be a connection between proton disintegration and the matter–antimatter asymmetry in that in the early universe protons were formed more abundantly than antiprotons in a reverse "proton decay" type of process. Thus, the observation of proton instability would have profound implications concerning the nature of the universe.

Protons are extremely stable. So far, the lifetime for disintegration by any possible mechanism into any possible final state has been shown to be greater than 10^{27} years. Thus, protons are the longest-lived particles that we know of. The corresponding limit for the disintegration of the electron is 10^{22} years. Could protons be completely stable? This seems unlikely since such a stability would imply a rigorously conserved "quantum number." One such quantum number is electric charge, and electrons are expected to be entirely stable because there is no way for the electron to disintegrate and conserve the electric charge and energy at the same time. In the case of the proton there are many possible ways for it to decay while conserving electric charge and energy. It could be that the proton carries a conserved property (sometimes called baryon charge or baryon number). One rule in the physics of elementary particle is that every rigorously conserved quantum number is associated with a long-range field or some invariance principle. For example, momentum conservation is a consequence of translational invariance (the laws of physics are independent of position). The long-range force associated with charge conservation is the electromagnetic field. No such long-range force has been identified for baryon numbers, and there are strong restrictions on such forces from the test of the equivalence principle of general relativity. (This principle states the equivalence of gravitational and inertial mass.) Without a rigorously conserved quantum number we are left in a quandry over the apparent stability of the proton. In order to gain further insight it is necessary to understand what protons could decay into if they are unstable. Since angular momentum is a rigorously defined quantum number, the angular momentum of the proton ($\frac{1}{2}$ unit) must be matched by the angular momentum of the decay products. Only three such types of decay products are possible: electron, muon leptons, and neutrinos (ν_e, ν_μ, ν_τ, where the subscript denotes the type of lepton). Thus, the proton must decay into another kind of particle, the lepton. Three such decays that conserve all other quantum numbers are

$$p \rightarrow e^+ + \pi^0$$
$$p \rightarrow \mu^+ + \mu^+ + \mu^-$$
$$p \rightarrow K^+ + \bar{\nu}_\mu.$$

In the first case only one lepton appears in the final state along with a neutral meson; in the second case three leptons appear. These are two of the very large variety of final states that the proton can disintegrate into while maintaining the conservation of all quantum numbers except baryon number and lepton number (an equivalent quantum number carried by leptons such as electrons, muons, and tau leptons). The stability of the proton is likely connected to the difficulty of transforming baryons into leptons due to the great difference between these particles. For example, a proton is thought to be made from three quarks, whereas a lepton is a fundamental particle with no known subparts.

In order to discuss the theoretical possibility of proton decay we must first discuss the ways of distinguishing baryons from leptons. Baryons participate in the strong electromagnetic and weak interactions, whereas leptons participate only in electromagnetic and weak interactions. As we now know, the weak and electromagnetic interactions are unified into one electroweak interaction. If the electroweak and strong interaction were unified (as in grand unified theories, or GUTs) at some very high energy, the distinction between leptons and baryons (at those energies) would disappear.

The prime motivation for believing in a grand unification of the forces comes from the prediction that the strong force coupling strength is decreasing with increased energy. Thus, as shown in Fig. 1 the weak, electromagnetic, and strong forces should have the same strength in the vicinity of 10^{15} GeV center of mass energy. If the forces have the same strength, the major distinction between electrons and quarks should disappear. Hence, quarks and leptons can transform into one another. These baryons and leptons could transform into one another. Many GUTs have been formulated that indicated that the energy scale for the unification is 10^{14}–10^{15} GeV. This energy is vastly higher than anything available in the present universe (the highest-energy cosmic rays reach only an equivalent energy of $\sim 10^5$ GeV). However, these energies were present during the "big bang" and the early universe. These energies can be reached in a virtual state of the proton system if we apply the Heisenberg uncertainty principle:

$$\Delta E \, \Delta t \sim \hbar$$

If $\Delta E \sim 10^{15}$ GeV, then the corresponding Δt is equal to 10^{15} GeV. Plausible theoretical estimates indicate that the proton may live 10^{29}–10^{35} years if the GUTs have any validity. One of the most remarkable predictions of the SU(5) and other GUTs is the value of the weak-electromagnetic mixing angle (the Weinberg angle), $\sin^2 \theta w$. At the energy of the grand unification, $\sin^2\theta = \frac{3}{8}$. However, at very small energy there is a correction factor that brings the value down. The best estimate of $\sin^2 \theta w$ is 0.216 ± 0.006, compared with the world average from

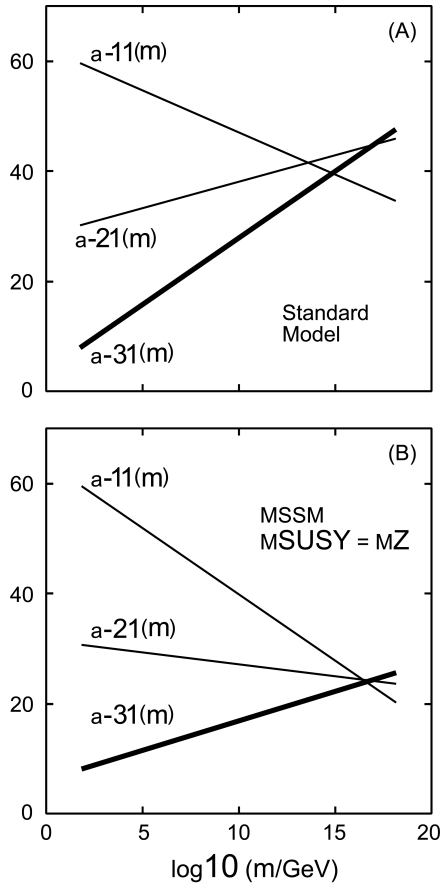


FIGURE 1 (A) The extrapolation of the three running coupling strengths fails to meet at any energy. (B) With the inclusion of low-energy supersymmetric particles, the coupling strengths now unify at the energy of about 10^{16} GeV.

the W and Z masses of 0.221 ± 0.013, a remarkable agreement.

The theory of nucleon decay can be distinguished by the mechanism by which the baryon number of the nucleon is reduced to zero in the final state products. Two general possibilities have been proposed, which we can indicate schematically:

1. Nucleon $\rightarrow$ 3 quarks $\rightarrow$ antilepton + antiquark + quark
2. Nucleon $\rightarrow$ 3 quarks $\rightarrow$ leptons + antileptons + quarks + antiquarks

A classic example of decay type (1) is $p \rightarrow u + u + d \rightarrow e^+ + \bar{d} + d \rightarrow e^+ +$ neutral meson (π^0). An example of Type (2) decay is $n \rightarrow u + u + d \rightarrow \nu + \nu + \mu^- + \bar{u} + d \rightarrow \nu + \nu + \mu^- + \pi^+$. In this case the individual quarks have decayed and the quark–antiquarks have been produced in the final interactions of the system.

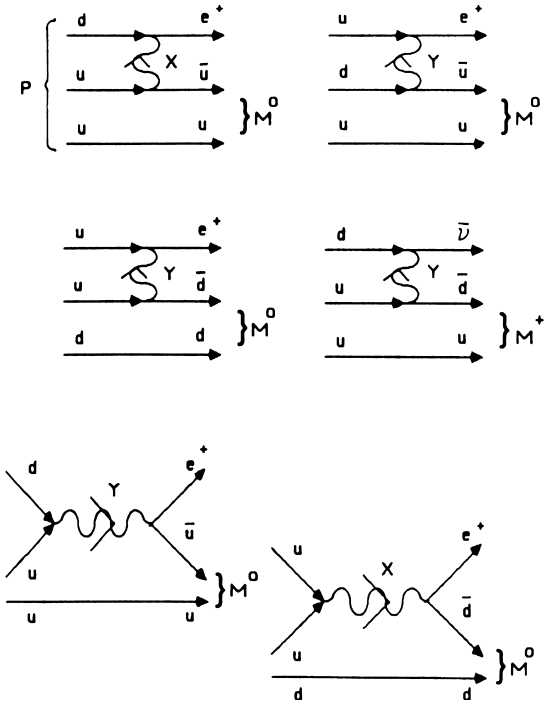


FIGURE 2 Diagrams for proton decay in Modes, of grand unification.

In this final state about one-half of the nucleon energy goes into neutrinos. Decay of Types (1) and (2) follows from a large variety of theoretical models of grand unification of forces.

The two basic kinds of theories for proton decay are related to the mechanism by which the baryonic charge is converted to leptonic charge. Figure 2 shows the fusion type of theory, in which two quarks transform into an antilepton and an antiquark. The SU(5) theory is such an example. Figure 3 shows the other possibility, in which the individual quarks "decay" into leptons and other particles. A theory of this type has been put forward by Pati and Salam and is called [SU(4)]. A third type of proton decay is expected in the theory that combines grand unification and supersymmetry. This is sometimes called a SUSY-GUTS theory. The dominant decay mode is expected to

$$p \rightarrow K^+ + \bar{\nu}\mu$$

Since there is a theoretical prediction for supersymmetry it will be important to search for this decay mode in future experiments. We will not attempt to discuss all of the remaining theoretical possibilities but will only outline the possible nucleon decay lifetime ranges that may be expected from these theories. The search for proton decay within this time limit is the target of several large experimental detectors in the world. We now turn to this search.

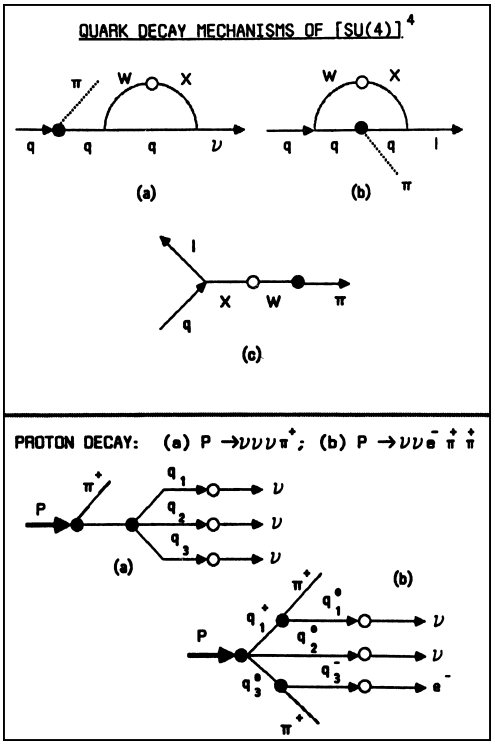


FIGURE 3 Proton decay process invented by J. C. Pati and A. Salam. In this model several kinds of intermediate particles lead to the proton decay (1 = charged lepton).

II. EXPECTATIONS AND THEORY OF PROTON DECAY

The most interesting development in the theory of proton decay is the possibility that the scale of the Grand Unification of Forces could be changed by the supersymmetric GUTs model. In Fig. 1(A), we show that in the normal GUT theory, the coupling constants fail to cross or unify at any energy. However, in the SUSY-GUT model (Fig. 1(B)), there is an energy of unification. The current limits on proton decay have ruled out the simplest GUT model based on the Group SU95). However, if supersymmetry is involved, a model based on the supersymmetric SU(5) is still viable. Recent studies of unification that invoke theories beyond supersymmetry (so-called string theories) do not really provide any direct predictions. In essence, with the simplest GUT theory ruled out, it is really an issue of broadest experimental search for proton decay by the very best methods for the future.

III. EXPERIMENTAL DETECTORS FOR PROTON DECAY

There are two general kinds of experimental techniques by which to search for proton or bored neutron decay:

1. Nucleon disappearance inside the nucleus, causing radioactive changes afterward
2. Active detection of the nucleon decay products in very large detectors

In the first case the lifetime limits that can be achieved appear to be less than $\sim 10^{29}$ yr. It is important that these searches be independent of the kind of decay the proton undergoes. The most sensitive search for nucleon decay involves the detection of the decay products. In this type of experiment very large quantities of matter are studied in an attempt to detect the disintegration of one or more of the nucleons in the detector. Three types of detectors are in use or are being contemplated:

1. Large water detectors with the nucleon decay products observed by the emission of Cherenkov light and detected by photomultipliers (Fig. 4)
2. Large, thin, solid plate stacks with appropriate ionization detectors inserted between the plates
3. Large quantities of liquid argon in which the nucleon decay products are detected by ionization of electrons that are drifted over large distances and collected on wires (Fig. 5)

These detectors can have a sensitivity in the following ranges:

1. Water detectors: 100–30,000 tons
2. Solid plate detectors: 100–10,000 tons
3. Liquid argon: 100–10,000 tons

The masses and other properties of existing or planned detectors are given in Table I. The possible lifetime that corresponds to detecting one nucleon decay in 1 year in an N-ton detector is given by the relation

$$\tau_{\text{p decay}} = N \times 3 \times 10^{29}$$

so for 10^4 tons we could in principle reach a lifetime level of 3×10^{33} year for either protons or neutrons in the detector.

In practice there are two difficulties: (1) The overall detection efficiency is usually much less than 1 (typically $\sim$30%), and (2) there are backgrounds that can mimic the signature for nucleon decay (e.g., cosmic ray neutrino interaction in the detector), requiring more stringent requirements for the nucleon decay candidates. These effects tend to make the lifetimes that can be reached more than an order of magnitude or two below the ideal limits.

Before giving the status of the search for nucleon decay, we first classify the type of decay modes that are being sought. We classify these processes by the type and

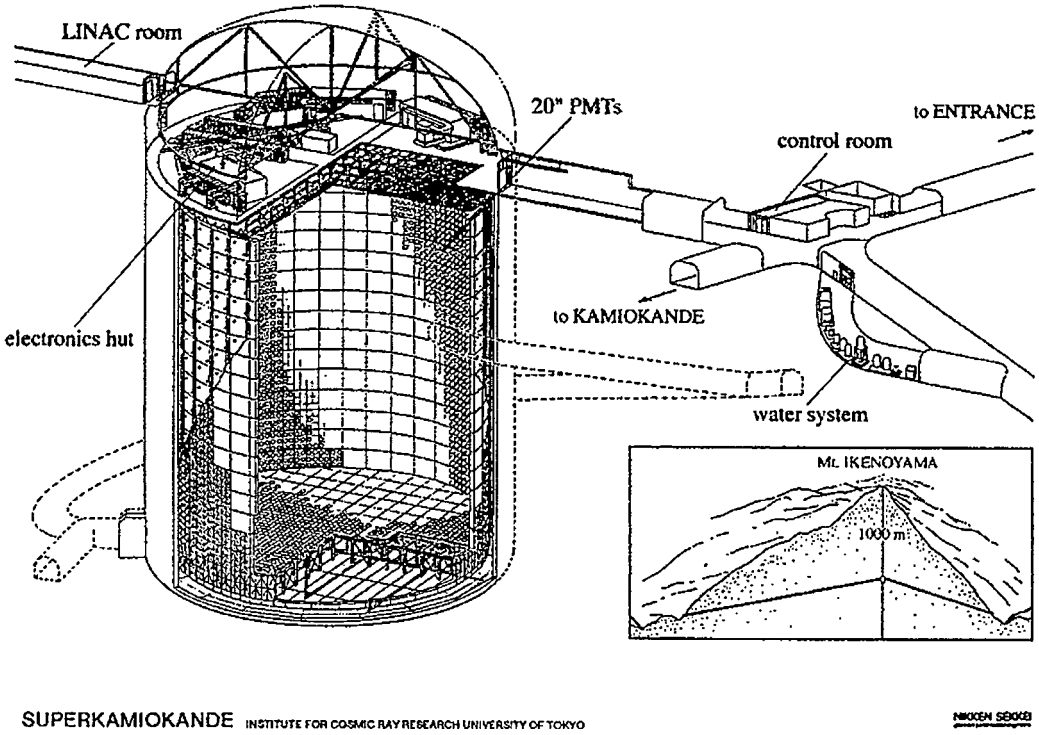


FIGURE 4 Schematic drawing of the 50,000-ton water-detector, Superkamiokande, in the Kamioka Mine in Japan. The walls of the detector are lined with very sensitive photomultiplier.

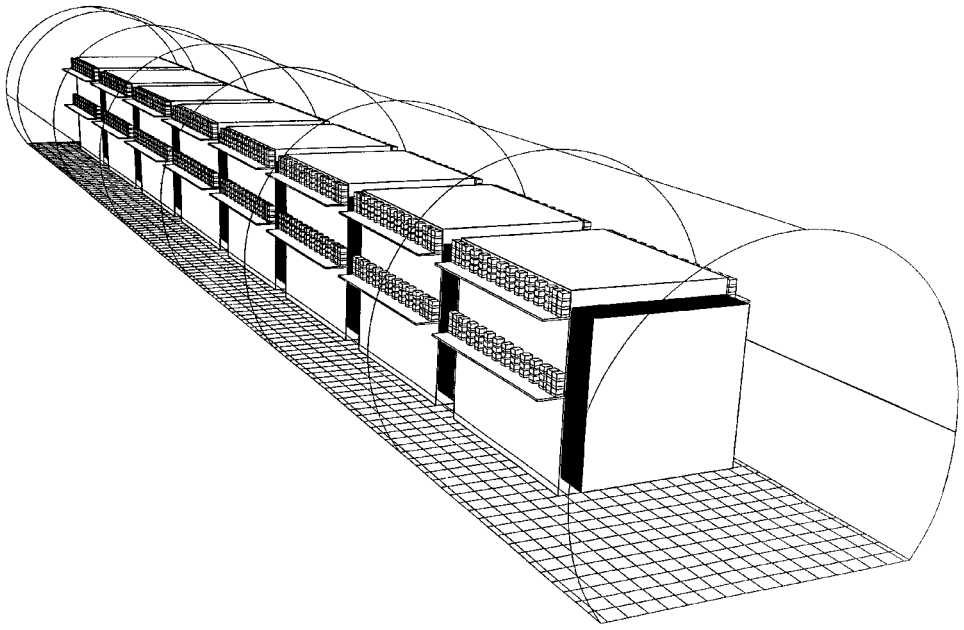


FIGURE 5 Schematic of the liquid drift detector, ICARUS, being constructed at the Gran Sasso Underground Laboratory near Rome, Italy. It comprises a 5000-ton liquid-argon image changer, with excellent spatial and energy resolution, and an iron plate detector, NOE.

TABLE I Properties of Detectors for Current and Future Proton Decay Searches

Experiment, geomagnetic latitude, depth	Total mass, fid mass, year of operation	Particle recognition	Remarks
Souden[a] 48°N 713 m rock	1000 tons Fe 500 tons 1990	$\gtrsim$160 MeV $\pi^{\pm}$ $\gtrsim$160 MeV $\pi^{\pm}$ $\gtrsim$ 50 MeV $\gamma, e^{\pm}$	1.6-mm-thick Fe; uses resistence plastic tube to record signal
ICARUS/NOE[b] 35°N 5000 m.w.e.[c]	5000 tons liquid Ar 4000 tons 1995	$\gtrsim$50 MeV $\pi^{\pm}$ $\gtrsim$50 MeV $\mu^{\pm}$ $\gtrsim$50 MeV $K^{\pm}$ $\gtrsim$5 MeV $\gamma, e^{\pm}$	Excellent identification of different particles in homogeneous liquid Ar with electron drift chamber readout
Super Kamiokanda[d] 26°N 2700 m.w.e.[c]	50,000 tons H_2O 33,000 tons H_2O 1983	$\gtrsim$200 MeV $\pi^{\pm}$ $\gtrsim$160 MeV $\mu^{\pm}$ $\gtrsim$20 MeV $\gamma, e^{\pm}$	11,000 × 20 in. ϕ photomultiplier tubes covering 20% of the surface

[a] Iron plate detector in the Sonda Mine, MN.
[b] Liquid argon detector being constructed for the Gran Sasso near Rome.
[c] Meters of water equivalent.
[d] Water detector in the Kamioka Mine, Gifu Prefecture, west of Tokyo.

number of leptons that are produced in the decay. We define two general decay schemes:

1. Nucleon → (antilepton or lepton) + meson(s)
2. Nucleon → (3 charged leptons or neutrinos) + mesons

As mentioned before, the conservation of angular momentum requires that the nucleon decay products include an odd number of leptons, antileptons, or a mixture of leptons and antileptons. Also electric charge conservation is assumed.

In Table II we give the current status of the limits on decay modes of Type (1). We note that in several experiments possible candidates for decays have been observed, but there has been no completely convincing example of a decay that is not compatible with background.

In Table III we give the early limits that have been obtained on the decays of Type (2). The number of possible decay modes of the nucleon is much larger for this case, because we are dealing with combinations of three leptons or mixtures of leptons and antileptons. Furthermore, many of these decay processes will have neutrinos in the final state, leading to a poorly defined experimental signature.

The present conclusions concerning the search for nucleon decay can be summarized as follows:

1. The decay mode $p \to \pi^0 e^+$ has been searched for, and no candidates in the largest detectors have been observed. The current limit on the τ/B for this decay is 2×10^{33} years.
2. While there are candidates for other proton decay modes, to date all candidates are consistent with the

TABLE II Search for Nuclear Decay into Antileptons and Mesons $\tau/B \times 10^{32}$ Years (Approximate Limits), Where B Is the Branching Fraction into the Specific Mode

Mode (Proton)	Lifetime limit (approximately $\times 10^{32}$)		
	Soudan II[a]	Super Kamiokande[b]	IMB[c]
$e^+\pi^\circ$	0.38	21.0	5.4
$\bar{\nu}K^+$	0.43	6.7	1.8

[a] Soudan II group.
[b] Kamiodande group.
[c] Irvine–Michigan–Brookhaven group.

TABLE III Search for Nucleon Decay into Multileptons and Mesons—Selected Final States

Decay mode	Lifetime limit ($\times 10^{31}$)		
	HPW[a]	IMB[b]	Soudan II[c]
$e^+e^-\nu$	—	28	7.4
$\mu^+\mu^-\nu$	3.1	14	—
$e^+e^-e^+$	—	79	—
$\mu^+\mu^-\mu^+$	5.0	67	5.9
$\mu^+\mu^-\nu\pi^+$	2.0		
$\nu\nu e^-\pi^+\pi^+$	2.0		
$\nu\nu\mu^-\pi^+\pi^+$	2.0		
$\nu e^-e^-\pi^+\pi^+\pi^+$	2.0		
$\nu e^-\mu^-\pi^+\pi^+\pi^+$	2.0		
$\nu\mu^-\mu^-\pi^+\pi^+\pi^+$	2.0		

[a] Harvard–Purdue–Wisconsin group.
[b] Irvine–Michigan–Brookhaven group.
[c] Soudan II group.

expectations of backgrounds due to atmospheric interactions in the detectors.

3. The search for proton decay through SUSY-GUTS is progressing. As stated before, the most important decay mode is expected to be

$$p \rightarrow K^+ + \bar{\nu}_\mu$$

The search for this decay has yielded a world average lifetime limit of 7×10^{32} years. However, this limit depends on a background subtraction and could still be uncertain due to the background.

Future searches for proton decay are likely to concentrate on the detection of this mode and

$$p \rightarrow \pi^0 \mathrm{e}^+$$

since, at present, these are the best motivated by theory.

IV. THE NEXT STEP IN THE SEARCH FOR NUCLEON DECAY

How can the existence of nucleon decay be established, given that the present level of lifetime is likely 10^{32} years or greater? Many of the present detectors are being improved, and this could lead to a definitive observation of nucleon decay in the latter half of the 1990s. However, the current limits on the nucleon lifetime are comparable to the effective rate of neutrino interactions in the detectors, which constitutes the major experimental background. Therefore, it is likely that a new kind of experimental detector is required that gives sufficient information about the event characteristics in order to distinguish fully nucleon decay from neutrino background.

A new detector at the Soudon Mine will use drift chambers and plates and may be capable of providing a more convincing search than previous iron plate detectors. However, the ultimate mass of the detector is likely to be less than 1000 tons, so it may not extend the lifetime limits much beyond that in Tables II and III. Other detectors that use liquid argon and electron drifting are also being developed. These detectors might be constructed in the mass range of 1000 to 10,000 tons and would allow a definitive search for nucleon decay in most channels to 5×10^{33} years. One other possibility is to continue the search for a specific decay such as

$$p \rightarrow \pi^0 \mathrm{e}^+$$

to longer lifetime using a larger water detector (say 50,000 tons). Such large water detectors in mines or under water have been discussed. The Super Kamiokande detector has now operated successfully.

One new detector is worth describing here because it may be useful to extend the ultimate lifetime limits in the proton decay search on earth. The ICARUS/NOE detector being constructed at the Gran Sasso laboratory near Rome would be the first generation of such detectors. This detector will have 5000 tons of liquid argon. However, detectors as large as 30,000 tons might be constructed in the future (Super ICARUS). These detectors could break the 10^{33}-year level and might reach 10^{34}-year lifetimes for some decay modes. The principle of these detectors is to use the electrons from the ionization of the proton decay products to image the event. The electrons are drifted over large distance to a readout system. The resulting event picture should have extremely good resolution, and this could be used to reject neutrino background. The ICARUS detector is well matched to the search for SUSYGUTS proton decay.

If nucleon decay occurs through the Pati–Salam decay schemes (three quark decays), the discovery of nucleon decay will be considerably more difficult because of the likely presence of neutrinos in the final states. The separation of signal and neutrino background will be much more difficult.

To summarize the prospects for establishing the existence of proton decay:

1. Current detectors have reached lifetime limits of $\sim 10^{32}$-years for a large variety of modes and $\sim 2 \times 10^{33}$ years for the $\pi^\circ e^+$ mode.
2. New detectors are being constructed. Three such detectors are
   a. Super Kamiokande: 50,000 tons of water (located in Japan and now operating);
   b. ICARUS/NOE: 5000 tons of liquid argon (located in Italy and being constructed by an Italy/UCLA team);
   c. Soudan II: heavy plate, $\sim$1000 tons of drift chamber (located in Minnesota).

Two of these detectors mentioned in summary could detect proton decay for the cases

$$p \rightarrow \pi^\circ e^+ \qquad \tau \sim 2 \times 10^{34} \text{ years}$$
$$p \rightarrow K^+ + \bar{\nu}_\mu \qquad \tau \sim 10^{34} \text{ years.}$$

To go further, say 5×10^{34} years for a large number of decay modes, another generation of very high resolution and high mass ionization tracking detectors may be needed ($\sim$30,000 tons of liquid argon). However, the background from neutrino interactions in the detector could limit this range.

The search for proton decay is simply one of the most important issues about our universe. For this reason, once started, it is unlikely that physicists will ever give up the search as long as any feasible technique can be found.

Some decays, such as

$$p \rightarrow \nu\nu\nu\pi^+$$

will be difficult to establish even if the lifetime is as short as 10^{31} years.

The ultimate proton decay detector would be on the moon in a lunar scientific station. The neutrino background is greatly reduced on the moon due to the lack of an atmosphere. It may be possible to search for proton decay to lifetimes of 10^{35} years eventually on a lunar base. NASA has sponsored various workshops recently to discuss lunar base experiments. The projected model for a Lunar proton decay detector would use compacted moon dust and very low weight drift chambers.

V. SIGNIFICANCE OF THE LIMITS ON NUCLEON LIFETIME

There are two important consequences of the limit on nucleon decay:

1. The limit on the conservation of baryon decay is increased to approximately 10^{32} years for most decay processes.
2. The minimum energy at which grand unification of the force occurs in the "standard" theories must be $\sim 10^{15}$ GeV.

At present these two conclusions are not in conflict with a large number of possible theoretical models. The lifetime limit for nucleon decay corresponds to one of the largest (or smallest coupling) numbers ever studied in elementary particle physics. Thus, the law of conservation of baryon number is "nearly" perfect. There are still reasons to believe that this is not an exact law of nature, which we summarize as follows:

1. There is no long-range or fundamental principle that ensures exact conservation.
2. The universe is baryon number asymmetric: this may indicate that the law is not exact and was broken in the early universe.
3. The GUTs appear to give a "logical" reason for the long proton lifetime associated with the very high energy at which unification occurs.
4. Supersymmetry is consistent with the current limits and could be tested with the next set of detectors. For these reasons the experimental search for baryon number nonconservation must continue.

SEE ALSO THE FOLLOWING ARTICLES

NEUTRINOS • PARTICLE PHYSICS, ELEMENTARY • UNIFIED FIELD THEORIES

BIBLIOGRAPHY

Allison, W. W. A., *et al*. (1998). *Phys. Lett. B* **427,** 217 (Soudan).

Amaldi, U., *et al*. (1987). *Phys. Rev. D* **36,** 1385.

Arnowitt, R., and Nath, P. (1988). *Phys. Rev. Lett.* **60,** 1817.

Cline, D. (1980). "Experiment to detect proton decay and neutrino oscillations (using cosmic ray neutrino events)," *Publ. Erice Subnuclear* (QCD161:I65) **307**.

Cline, D. B. (1990). "Search for Proton Decay and Supernova Neutrino Bursts with a Lunar Base Neutrino Detector," *AIP Press* in Lunar Base Experiments.

Cline, D. B., and Rubbia, C. (1978). "Proceedings of the Seminar on Proton Stability," University of Wisconsin, Madison, December 8, "*World Scientific*" (Frampton, ed.), Last Workshop on Grand Unification (1990).

Dienes, K. (1997). *Phys. Rep.* **287,** 447.

Georgi, H., and Glashow, S. (1974). *Phys. Rev. Lett.* **32,** 438.

Langacker, P. (1981). *Phys. Rep.* **4,** 274–277.

Marciano, W. J., and Sirlin, A. (1980). *Phys. Rev.* **D22,** 2695.

More, A. (1984). "An Experimental Search for PatiSalam Decay Modes of the Proton," doctoral dissertation, University of Wisconsin, Madison. (HPW)

Pati, J. C., and Salam, A. (1973). *Phys. Rev.* **D8,** 1240.

Pati, J. C., and Salam, A. (1974). *Phys. Rev.* **D10,** 275.

Shiozawa, M., *et al.* (1998). *Phys. Rev. Lett.* **81,** 533 (Super Kamiokande).



Unified Field Theories

Frank Wilczek
Massachusetts Institute of Technology

I. Symmetry in the Standard Model
II. Unification using Extended Symmetry
III. Quantitative Unification
IV. Succeeding with Supersymmetry
V. Unification with Gravity?

GLOSSARY

Effective, or running, coupling The strength of a fundamental interaction is specified by its coupling constant. Thus, the strength of an electron's interaction with photons is specified by the value of its charge, conventionally denoted e. It can be determined, for example, by measuring the force between electrons. A very important result of quantum field theory is that the coupling strength depends on the distance at which it is measured, due to the screening or antiscreening effect of quantum fluctuations (virtual particles). We say that there is an effective coupling, whose magnitude varies, or runs, with distance. Alternatively, and more practically, we can consider the probability that an energetic electron will radiate a photon. Here we find that the probability changes with energy. Thus we define an effective coupling that varies, or runs, as a function of energy. The running of the coupling plays a particularly important role in quantum chromodynamics, or QCD, where it is a very large effect. In that context the decrease of the effective coupling at short distances, or large energies, is called *asymptotic freedom*.

Field A dynamical entity filling all of space–time. In an older terminology, it might be called an *ether*. Classical fields take on different value as functions of time and space; the state of the field at time t is the totality of its values, over all space, at that time. Thus, for example, the classical electric field is a dynamical entity whose state–history is specified by a set of vectors, one for each point of space–time. Note that a field exists even at points where its value vanishes. This is no more mysterious than the idea that a bank account must have some balance, even though the value of that balance might happen, on some specific occasion, to be zero.

Gauge field In order for the laws of physics to appear the same to observers executing more general—accelerated—motions, we must allow for the possibility that gravitational fields appear in their description. This is perhaps best epitomized in Einstein's elevator thought-experiment: a scientist in a closed elevator-laboratory could equally well describe her world by postulating that it is accelerating, or that it is subject to a gravitational field. These descriptions are equally valid and, in general relativity, mathematically equivalent: this is Einstein's celebrated equivalence principle. Turning the argument around, by postulating the appropriate extended symmetry (general covariance) one can deduce the existence and main properties of the gravitational field.

The local symmetries central to the other important modern field theories, which describe the strong, electromagnetic, and weak interactions, are of a less intuitive nature than general covariance, which relates to motion in space–time. They have to do with transformations of one type of particle into another, or of the phases of quantum-mechanical wave functions. But their implications are similar in scope. By postulating appropriate extended forms of symmetry, one can deduce the existence and main properties of the color gluon, photon, and intermediate vector boson (W and Z) fields.

Local, or gauge, symmetry An extended form of symmetry, characteristic of important field theories in physics. A local symmetry, also often called a gauge symmetry, is one whose transformations can be made independently at all points of space-time. The prototype for this innovation is Einstein's theory of gravitation, general relativity. While Lorentz symmetry contemplates a single uniform velocity for the entire Universe, the more sweeping principle of general covariance considers all possible motions. In other words, it allows the velocity to be chosen independently at each point of space–time. To fully elucidate the concept of local symmetry we must consider it together with that of a gauge field.

Planck scale The conventional measure of the coupling strength of gravity is Newton's constant G_N. From this quantity, together with the speed of light c and Planck's constant $\hbar$, we can construct a complete set of physical units. In these units the fundamental scale of length is $(G_N\hbar/c^3)^{\frac{1}{2}} \sim 10^{-33}$ cm and the fundamental unit of energy is $(\hbar c^5/G_N)^{\frac{1}{2}} \sim 10^{19}$ GeV. The latter is often called the Planck scale. For physics, the Planck scale is the scale at which the running coupling for gravity, which is tiny for practically accessible energies $\lesssim 10^4$ GeV, becomes comparable to that for the other fundamental interactions.

Quantum field The entities described by fields in quantum theory are so different in character from their classical counterparts as to deserve an additional definition. Quantum fields are still dynamical entities that fill all space–time, but their state is defined much more abstractly than for classical fields. It consists in the quantum–mechanical amplitude for each possible configuration of the corresponding classical field. This abstract definition does not immediately convey the physical significance of the quantum field concept, which emerges only after mathematical analysis. Roughly speaking, quantum fields create and destroy particles (or "wavicles"—for these particles obey the rules of quantum mechanics). For example, the quantum electromagnetic field creates and destroys photons, while the quantum electron field creates electrons and destroys anti-electrons (positrons).

Spontaneously broken symmetry There are many important cases in physics where a symmetry of the equations governing a system is not exhibited in their stable solutions. Thus the equations governing the interactions of electrons in a ferromagnet are unchanged by rotation, but to minimize their total energy the electrons' spins must align, and in doing so collectively choose a preferred direction. Spontaneous symmetry breaking has been a powerful tool in constructing unified field theories, since it allows us to consider the possibility that the laws of physics are in a profound sense simpler than what we perceive superficially.

Supersymmetry Supersymmetry is a new kind of symmetry, which extends the Lorentz symmetry of special relativity. It postulates the existence of additional purely quantum dimensions. When a particle takes a step into one of the quantum directions, its position in ordinary space-time does not change, but it changes its spin and quantum statistics. For example, a spin-0 boson will transform into a spin-1/2 fermion. Supersymmetry transformations mix ordinary and quantum dimensions. In order for such transforms to be symmetries of physical law, there must be particles of different spin and statistics with closely related physical properties. Although at present the evidence is far from conclusive, there are serious reasons to believe that (spontaneously broken) supersymmetry is a feature of fundamental physical law.

Symmetry We say that physical laws have symmetry if there are transformations that might, as a matter of pure logic, have changed the form of the laws, but in fact do not. This generalizes and makes precise the notion of an object possessing symmetry. Thus a circle possesses rotational symmetry, because if it is rotated around its center it does not change. Lorentz symmetry (also called Galilean invariance), which lies at the heart of the special theory of relativity, is the statement that motion by a constant velocity is a symmetry of physical law. Logically, the laws of physics might have taken a different form for an observer moving (together with all her measuring apparatus) at constant velocity—but in fact they do not.

THE PHYSICAL WORLD presents us with a wondrous diversity of phenomena. One might have anticipated, *a priori*, that as knowledge progresses, and more wide-ranging and precise observations are gathered, any adequate summary of this knowledge would require a proliferation of independent concepts and contingent facts, and become more complex. Remarkably, the trend of physics,

especially over the last 150 years, has been quite the contrary. It has proved possible to formulate a small number of fundamental laws that in principle govern not only the behavior but also the form and structure of matter. These laws specify the existence of a large number of quantum fields, and specific rules for how they interact with one another. The laws are characterized by a high degree of symmetry. The symmetry transformations relate many of the quantum fields to one another. From this perspective the different fields should be viewed as different aspects of a single entity, like the different facets of a single crystal.

Unified field theory's program is to identify a unique conceptual structure of this kind that successfully describes observed reality. The ideal of a comprehensive, fully unified theory has been partially, but not completely, realized. And there are concrete, promising ideas for extending its scope.

I. SYMMETRY IN THE STANDARD MODEL

Experimental discoveries starting in the waning years of the 19th century revealed that the classical subject-matter of physics, namely, forces of gravity and electromagnetism, was seriously incomplete as a description of Nature. Additional interactions of a fundamentally new kind were found to be at work in atomic nuclei. After many ingenious investigations physicists developed a working description of nuclear physics, which gave them a crude but useful description of the phenomena. This description required two new interactions. One is called the *strong interaction*: it is the most powerful force in Nature, but operates only over distances less than about 10^{-13} cm (a typical nuclear size, not coincidentally), and it is the basic force responsible for binding the protons and neutrons in nuclei together. In addition, there is a very peculiar so-called *weak interaction*. The weak interaction is both short-range and feeble, as its name implies, but it has the remarkable ability, not possessed by other interactions, to change the character of particles. It is responsible for β-radioactivity, and for the nuclear burning processes that generate energy in stars. A typical process described by the weak interaction is the decay of free neutrons into protons, electrons, and antineutrinos.

Despite their relatively late appearance in history, the strong and weak interactions are every bit as necessary as gravity and electromagnetism to the description of Nature. Gravity already received its classic formulation in the 17th century in Newton's work, subsequently profoundly modified and deepened in Einstein's general relativity; and electromagnetism was beautifully summarized in Maxwell's equations of 1864. For several decades following their discovery, there was no similarly worthy formulation of the basic equations of the strong or weak interactions. To find such equations was universally regarded as a primary problem of theoretical physics, but progress was fitful and uncertain. Then, in a brief period culminating around 1970 for the weak interaction and in 1973 for the strong interaction, the problem was solved, and the modern Standard Model took shape.

The Standard Model has a tight mathematical structure, and is rich in interesting details. Substantial, intellectually challenging books have been written about various small pieces of it. The story I wish to tell here can be understood as the unfolding of three qualitative ideas: *dynamical symmetry, asymptotic freedom*, and *spontaneous symmetry breaking*. These ideas form the conceptual center of the Standard Model; remarkably, as we shall see, they also suggest quite concretely how to go beyond it.

The first central concept of the Standard Model, dynamical symmetry, is a precise mathematical principle usually called local symmetry or gauge invariance. In calling it dynamical symmetry, I wish to emphasize the amazing and beautiful idea that *the mathematical concept of symmetry is embodied in specific, tangible physical forms*. In 1949 E. P. Wigner, who received the 1963 Nobel Prize for developing the consequences of symmetry in physics, foresaw that "in the future we may well derive the laws of nature and try to test their validity by means of the laws of invariance rather than to try to derive the laws of invariance from what we believe to be the laws of nature." In the intervening years, this vision has been brilliantly vindicated, as I shall now explain.

The traditional concept of a symmetry of physical law is that there are transformations that (logically) might have changed the form of the law, but in fact do not. For example, Lorentz invariance, which lies at the heart of the special theory of relativity, is the statement that motion by a constant velocity is a symmetry of physical law. Logically, the laws of physics might have been changed by this transformation—but in fact they are not.

The great innovation of dynamical symmetry is to connect the very existence, and detailed properties, of specific fields and particles with an extended concept of symmetry. The prototype for this innovation is Einstein's theory of gravitation, general relativity. While Lorentz invariance applies to motion with a constant velocity, the more sweeping principle of general covariance applies to all possible motions. But in order for the laws of physics to appear the same to observers executing more general—accelerated—motions, we must allow for the possibility that gravitational fields appear in their description. This is perhaps best epitomized in the famous Einstein elevator thought-experiment: a scientist in a closed elevator could

equally well describe her world by postulating that it is accelerating, or that it is subject to a gravitational field. These descriptions are equally valid and, in general relativity, mathematically equivalent—the equivalence principle. Turning the argument around, by postulating the appropriate extended symmetry (general covariance) one can *deduce* the existence and main properties of the gravitational field!

The symmetries that are embodied in the Standard Model, unfortunately, are of a less intuitive nature than Galilean invariance or general covariance, which relate to motion in space–time. The symmetries of the Standard Model concern transformations of one type of particle into another, or of the phases of quantum–mechanical wave functions. But the outcome is very similar: by postulating appropriate extended forms of symmetry, one can deduce the existence and main properties of fields, and the particles they create. Indeed, the first product of this way of thinking is a new way of regarding Maxwell's electrodynamics, due to Hermann Weyl and Fritz London. They showed, shortly after the birth of modern quantum mechanics in the mid-1920s, that the existence and main properties of the electromagnetic field, including the force it mediates and its embodiment in photons, could be derived as a consequence of a postulated symmetry of the equations of quantum mechanics (specifically, under space–time–dependent changes in the phases of charged fields). Thus they accomplished for electromagnetism what Einstein had accomplished for gravity. Indeed, even before the advent of modern quantum theory, Weyl had developed a related but different theory of how electromagnetism might be derived from a dynamical symmetry, in which he postulated symmetry under arbitrary changes in size. This was a brilliant and fruitful, though quite erroneous, theory.

According to the Weyl-London dynamical symmetry principle there is a property—(electric) charge—of particles, to which certain special particles—photons—respond in a particular, completely predictable way. All the observed phenomena of electromagnetism—what Dirac called "all of chemistry and most of physics"—can be shown to follow from this setup, of photons responding to electric charge.

In the Standard Model description of the strong interaction, we modify the dynamical symmetry of electromagnetism by allowing three different kinds of charge. These charges are commonly referred to as colors—say red, white, and blue—although of course they have nothing to do with color in the ordinary sense. The most straightforward generalization would be to have three "color photons," which respond to these color charges just as ordinary photons respond to electric charge. But the generalization that actually describes the strong interaction is more subtle and profound. It uses ideas that C. N. Yang and R. Mills had developed, in a somewhat different context, in 1954. In this version, one postulates an extended dynamical symmetry, which does not merely act on the different colors separately but also allows for the possibility of one changing into another. To allow this wider symmetry one must postulate additional fields, and one finds not just three but eight photon-like particles, called the color gluons.

The mathematical theory of eight color gluons responding to and changing three color charges, which I have crudely sketched earlier, is quite precise, and when properly formulated its equations display a high degree of symmetry. The theory is known as quantum chromodynamics or QCD, and its mathematical symmetry is denoted SU(3) (for three colors).

A most important difference between the new theory and electrodynamics is that the color gluons themselves carry color charge, unlike the photon, which is electrically neutral. Ordinary light beams—beams of photons—pass freely through one another. Color gluons interact powerfully, so that if one could make beams of gluons, they might bounce off each other, or explode on impact, or do something complicated (it has never been carefully worked out), but certainly they would not peacefully pass through one another.

Similarly, in the Standard Model description of the weak interaction we postulate two additional kinds of charge, say green and purple. There are three color gluons in this case, embodying the symmetry SU(2).

Finally, one needs to include electromagnetism itself. This is the theory of just one type of charge, and is denoted U(1). (There is an important complication here, due to spontaneous symmetry breaking, as mentioned in the caption to Fig. 1.)

Now these symmetries of $SU(3) \times SU(2) \times U(1)$ specify how many "photons" of various types there are, and how they will interact with given charges. To complete the description of the Standard Model, one needs only to say what the color charges of the various fundamental particles are. This is done in Fig. 1.

Since the Standard Model turns out to be such a simple conceptual generalization of electrodynamics, you might wonder why it was so hard to discover. Specifically, you might wonder why there was such a long gap—more than a scientific generation—between the Yang–Mills work and the emergence of the Standard Model. The main reason was that in each application, to the strong interaction and to the weak interaction, a significant additional idea is required.

In the case of the strong interaction, the additional idea is asymptotic freedom, discovered by D. Gross and F. Wilczek, and independently by H. D. Politzer, in 1973. Very rapidly (within a few weeks) after this discovery its

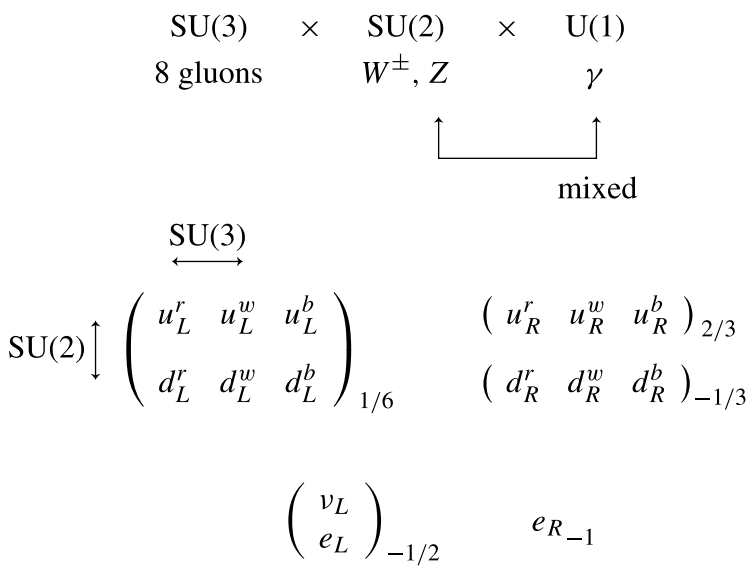


FIGURE 1 The core of the Standard Model: the gauge groups and the quantum numbers of quarks and leptons. There are three gauge groups, and five separate fermion multiplets (one of which, e_R, is a singlet). Implicit in this figure are the universal gauge couplings—exchanges of vector bosons—responsible for the classic phenomenology of the strong, weak, and electromagnetic interactions. The triadic replication of quark and leptons, and the Higgs field whose couplings and condensation are responsible for SU(2) × U(1) breaking and for fermion masses and mixings, are not indicated. SU(3) acts horizontally within each multiplet, while SU(2) acts vertically. The U(1) that appears directly in the Standard Model actually does not quite respond to electric charge, but to a closely related quantity known as hypercharge. Spontaneous symmetry breaking mixes up the SU(2) and U(1), so that the physical photon combines elements of each. The numbers to the right of each multiplet denote the hypercharge of particles in that multiplet.

application to a Yang–Mills theory of quarks and gluons, i.e., QCD in its modern form, was suggested by several authors, including D. Gross and F. Wilczek; S. Weinberg; and H. Fritzsch, M. Gell-Mann, and H. Leutwyler.

Taken at face value, the equations of Yang–Mills theory, like the equations of electromagnetism that they generalize, suggest an inverse-square law force: that is, a force falling off like one divided by the square of the distance. This is a fine starting point for the description of electromagnetic phenomena, but appears hopelessly wrong for the strong interaction.

On the one hand, our failure experimentally to find free single quarks requires that the force between quarks must forbid their distant separation. In particular, this force cannot fall to zero at large distances. On the other hand, profound experiments by J. Friedman, H. Kendall, and R. Taylor in the late 1960s demonstrated that the force between quarks becomes weak when the quarks are close together. From both perspectives, it is necessary to have a different force law—stronger at long distances and weaker at short distances—from what the Yang–Mills theory seems to suggest.

But amazingly enough the theory, when its consequences are calculated more accurately, is found to behave in just this way! The naive force law was derived by solving the equations of the classical field theory. It implicitly ignores the effect of quantum fluctuations. A more accurate calculation must take into account the effect of these quantum fluctuations, or "virtual particles." They have the effect of making the vacuum—what we ordinarily think of as empty space—into a dynamical medium, capable of rearranging itself in response to the presence of charges, thereby modifying the force law. It is not entirely a trivial matter to calculate the effect of this, but it can be done. Upon doing so, one finds that the force law is modified in such a way that it becomes weaker at short distances, as desired. The mutual interactions among the gluons, which I mentioned before as something with no analog in conventional electrodynamics, is crucial to the effect. A large number of precise experiments have been performed to test for asymptotic freedom, with the favorable results displayed in Fig. 2.

In the case of the weak interaction, the additional idea is spontaneous symmetry breaking. This idea has a long and complicated history. Its essential nature was perhaps first clearly recognized by P. Anderson in 1961 in connection with the theory of superconductivity; later it was presented in an especially transparent way in influential work by P. Higgs in 1964. The idea of adapting Yang–Mills theory to describe the weak interaction, including the crucial

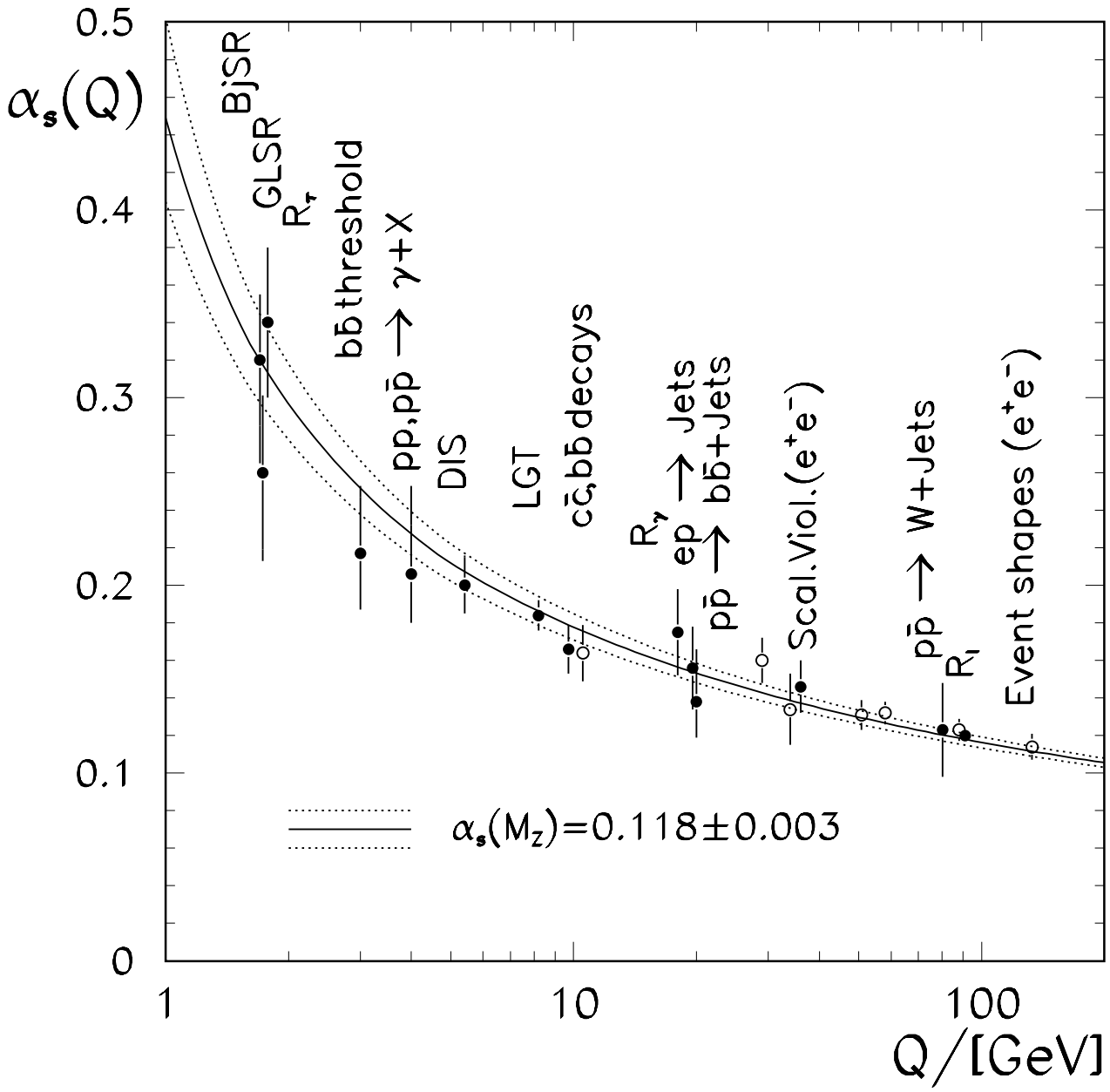


FIGURE 2 A recent compilation of tests of QCD and asymptotic freedom. Results are presented in the form of determinations of the effective coupling $\alpha_s(Q)$ as a function of the characteristic typical energy–momentum scale involved in the process being measured. Clearly the evidence for QCD in general, and for the decrease of effective coupling with increasing energy–momentum scale (asymptotic freedom) in particular, is overwhelming.

"mixing" feature (see the caption to Fig. 1) was proposed by S. Glashow in 1961; it was combined with the idea of spontaneous symmetry breaking by S. Weinberg in 1967 and independently by A. Salam in 1968. At this point important mathematical issues necessary for an adequate formulation of the theory still remained. They were finally settled by G. 't Hooft, building on work of M. Veltman, in 1970.

As its name suggests, the idea of spontaneous symmetry breaking is that the most symmetrical solution of the equations governing a physical situation may not be stable, so that the symmetry breaks "spontaneously." A standard example is provided by a ferromagnet: although the basic equations governing the iron oxide are completely symmetric under rotations, a lump of the stuff will have a specific preferred direction along which the atomic

spins tend to align, which breaks the symmetry. Similarly, it can easily happen that the only stable solutions of a theory with a large Yang–Mills–type symmetry display less symmetry. Indeed, it is absolutely necessary to exploit this possibility in order to use the Yang–Mills theories.

For future use, let me take note of one crucial effect of spontaneous symmetry breaking. It gives mass to some of the gluons—in this context, to the W and Z bosons. This has the physical effect of making the force they mediate appear much feebler, and of shorter range, than would otherwise be the case. These effects are an important part of what is required to make the basic Yang–Mills theory look like the observed weak interaction.

II. UNIFICATION USING EXTENDED SYMMETRY

While little doubt can remain that the standard model is essentially correct, a glance at Fig. 1 is enough to reveal that it is not a complete or final theory. The fermions fall apart into five lopsided pieces with peculiar hypercharge assignments; this pattern needs to be explained. Also the separate gauge symmetries, which as I mentioned are mathematically similar, are almost begging to be unified.

Given that the strong interactions are governed by transformations among three colors, and the weak by transformations between two others, what could be more natural than to embed both theories into a larger theory of transformations among all five colors?

Georgi and Glashow originated this line of thought in 1974, and showed how it could be used to bring considerable order into the jumble of Fig. 1. As shown in Fig. 3, the five scattered SU(3) × SU(2) × U(1) multiplets get organized into just two representations of SU(5).

One might have anticipated that it would be necessary to have six colors to accommodate the Standard Model—three for the strong interaction SU(3), two for the weak SU(2), and finally an additional one for the (hyper)charge of U(1). That it is possible to get by with five is a deep and beautiful fact, which all by itself should greatly encourage us to think that these unification ideas are on the right track. Since it is so basic and significant, and its essence can be understood with some simple arithmetic, I will explain it a little more fully. (A really adequate explanation involves some significant mathematics; I will just be supplying what I hope are some reasonable-sounding words to describe what emerges from mathematical analysis.)

Let us revert back a moment to the SU(3) theory, and address a question that may have occurred to you: Why are there exactly eight color gluons? Nine would be easier to understand; there would be one gluon corresponding to changing any unit of color into any other (including itself), for $3 \times 3 = 9$ altogether. This is close to being right; however, it turns out that when you add one unit of each color charge (red plus white plus blue) together the whole result cancels, so from the nine potential gluon degrees of freedom we must throw out one combination, namely, the combination that would have coupled to R + W + B, leaving eight as advertised. Similarly in the case of SU(2) we must throw out the G + P combination, giving $2 \times 2 - 1 = 3$ gluons in that case.

When you expand to SU(5), treating all five colors on an equal footing, it is no longer the case that these combinations cancel completely: R + W + B + G + P does, of course, but the opposite (so-called orthogonal) combination $\frac{1}{3}$(R + W + B) − $\frac{1}{2}$(G + P) does not. The gluon that couples to this charge gives an extra U(1), which comes for free, so to speak, with the five-color scheme. It is tied to a very specific pattern of charges for particles, completely specified in terms of their strong and weak colors. It is truly remarkable that these charges that follow from unification agree precisely with the observed (hyper)charges of the quarks and leptons in the Standard Model. Suddenly those peculiar little numbers of Fig. 1 appear in an entirely

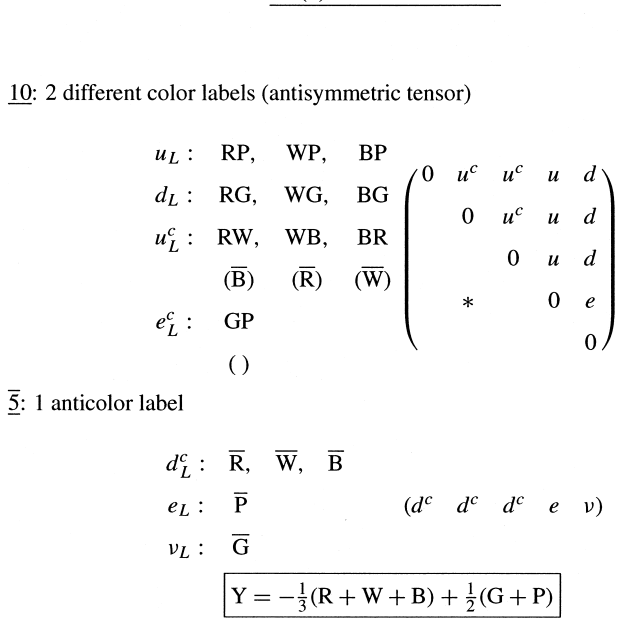


FIGURE 3 Organization of the fermions in one family in SU(5) multiplets. Only two multiplets are required. In passing from this form of displaying the gauge quantum numbers to the form familiar in the Standard Model, one uses the bleaching rules R + W + B = 0 and G + P = 0 for SU(3) and SU(2) color charges (in antisymmetric combinations). Hypercharge quantum numbers are identified using the formula in the box, which reflects that within the larger structure SU(5) one only has the combined bleaching rule R + W + B + G + P = 0. The economy of this figure, compared to Fig. 1, is evident.

different light. The ugly ducklings of the Standard Model have matured into elegant swans.

To summarize, *the structure of the Standard Model, with the particle assignments gleaned from decades of experimental effort and theoretical interpretation, can be reproduced perfectly by a simple abstract set of rules for manipulating symmetrical symbols*.

III. QUANTITATIVE UNIFICATION

We have now seen that the simplest unification scheme is extremely successful at the level of *classification*; but major questions arise when we carefully consider the underlying dynamics.

A first question concerns the implication of added dynamical symmetry, that there should be additional types of gluons mediating additional types of interactions, beyond those in the Standard Model. In particular there should be gluons that change strong into weak color charges, e.g., R into G. One can easily deduce the physical implications of these interactions. They appear, at first sight, catastrophic. Among other things the additional interactions are capable of causing protons to decay, whereas we know from experiment that protons live at least 10^{33} years on the average, and there is no experimental evidence for any instability at all. Fortunately, however, there is a simple way out of this problem, indeed one that has a precedent in the Standard Model. We have discussed that a big part of what makes the weak interaction weak (that is, feeble) is spontaneous symmetry breaking, which made its gauge particles, the W and Z bosons, heavy. If the large SU(5) or other symmetry is spontaneously broken, generating much larger masses for the dangerous new gauge particles, we can hope to suppress the unwanted interactions.

Another question concerns the implications of added symmetry within the Standard Model itself. Part of the power of dynamical symmetry is that it determines the interactions of the gauge bosons uniquely, once an overall coupling strength is specified. Thus if SU(5) or some higher symmetry were exact, then the fundamental strengths of the different color-changing interactions would have to be equal, as would the (properly normalized) hypercharge coupling strength. In reality the coupling strengths of the gauge bosons in $SU(3) \times SU(2) \times U(1)$ are observed not to be equal, but rather to follow the pattern $g_3 \gg g_2 > g_1$.

Fortunately, our experience with QCD and asymptotic freedom has taught us that coupling strengths "run"—their apparent value depends upon the distance or energy scale at which they are measured. For our present purpose of understanding the disparity among the observed couplings, it is just what the doctor ordered. As was first pointed out by Georgi, Quinn, and Weinberg in 1974, if a dynamical symmetry such as SU(5) is spontaneously broken at some very short distance scale, then we will only find the complete symmetry accurately reflected in processes at that scale or smaller. Thus we should not expect that the effective couplings measured at much larger distances, such as are actually measured at practical accelerators, will be equal; rather they will all have been affected to a greater or lesser extent by the quantum corrections due to virtual particles, as we have previously discussed in connection with QCD and asymptotic freedom. Indeed the pattern $g_3 \gg g_2 > g_1$ is just what we should expect, since the (inverse) asymptotic freedom effect, that effective couplings grow at large distances, is due to the mutual interactions among gluons. This is more significant the more gluons there are, so it affects g_3 much more than g_2, and g_2 somewhat more than g_1.

It is quite poetic, how each of the major conceptual players in the Standard Model—dynamic symmetry, asymptotic freedom, spontaneous symmetry breaking—is brought back to replay its role on a greatly expanded scale, in these proposals for the next level of unification.

Marvelously, this particular circle of ambitious ideas about unification has concrete, testable quantitative consequences.

Specifically, the running of the couplings gives us a truly quantitative handle on this physics, for the following reason. To fix the relevant aspects of the unified theory, one basically needs only to fix two parameters: the scale at which the couplings unite (that is, the scale of spontaneous symmetry breaking) and their common value when they unite. Given these, one calculates three outputs: the three *a priori* independent couplings for the gauge groups $SU(3) \times SU(2) \times U(1)$ of the Standard Model. Since there are more outputs than inputs, the framework is eminently falsifiable. It is startling, how close it comes to working (Fig. 4).

On closer examination difficulties appear. Accurate modern measurements of the couplings show a small but definite discrepancy between the couplings, as appears in Fig. 4. Also, although the predicted unification scale is quite large—$M_{\text{unif}} \sim 10^{15}$ GeV, corresponding to distances as small as 10^{-28} cm—and quite effective in alleviating the threat of proton decay—it is not quite large enough. Heroic dedicated experiments to search for proton decay did not find it, and they currently exclude the minimal SU(5) prediction by about two orders of magnitude.

Given the boldness and scope of the extrapolations involved, perhaps we should not have hoped for more. There are several perfectly plausible bits of physics that could upset the calculation, such as the existence of particles with masses much higher than the electroweak but much smaller than the unification scale. As virtual particles

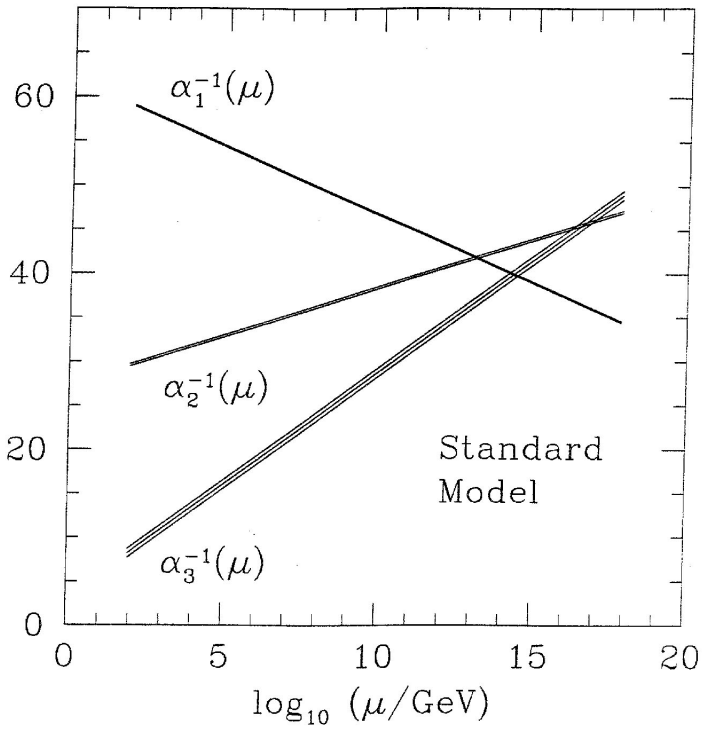


FIGURE 4 Evolution of Standard Model effective (inverse) couplings toward small space–time distances, or large energy–momentum scales. Notice that the physical behavior assumed for this figure is the direct continuation of Fig. 2, and has the same conceptual basis. The error bars on the experimental values at low energies are reflected in the thickness of the lines. Note the logarithmic scale. The qualitative aspect of these results is extremely encouraging for unification and for extrapolation of the principles of quantum field theory, but there is a definite small discrepancy with recent precision experiments.

these would affect the running of the couplings, and yet one certainly cannot exclude their existence on direct experimental grounds. If we just add particles in some haphazard way things will only get worse: minimal SU(5) nearly works, so the generic perturbation from it will be deleterious. Even if some *ad hoc* prescription could be made to work, that would be a disappointing outcome from what appeared to be a rare and precious, deep but logically straightforward, clue regarding physics well beyond the Standard Model.

IV. SUCCEEDING WITH SUPERSYMMETRY

Fortunately, there is a theoretical idea, attractive in many other ways, that seems to point a way out from this impasse: supersymmetry. Although there were several partial anticipations in the literature, perhaps the decisive early work on the subject was that of Wess and Zumino in 1974.

Supersymmetry is a symmetry that extends the Poincaré symmetry of special relativity. (There is also a general relativistic version—supergravity.) In a supersymmetric theory one has not only transformations among particle states with different energy–momentum but also between particle states of different *spin*. Thus spin-0 particles can be put in multiplets together with spin-$\frac{1}{2}$ particles, or spin-$\frac{1}{2}$ with spin-1, and so forth. Supersymmetry is certainly not a symmetry in Nature: for example, there is certainly no bosonic particle with the mass and charge of the electron. Nevertheless there are many reasons to be interested in supersymmetry.

Most important for our purposes, supersymmetry can help us to understand the vast disparity between weak and unified symmetry breaking scales, known as the gauge hierarchy problem. It raises several distinct difficulties, including the following. In calculating radiative corrections to the $(\text{mass})^2$ of the Higgs particle from diagrams of the type shown in Fig. 5, one finds an infinite, and also large, contribution. By this I mean that the divergence is quadratic in the ultraviolet cutoff. No ordinary symmetry will make its coefficient vanish. If we imagine that the unification scale provides the cutoff, we find that the radiative correction to the $(\text{mass})^2$ is much larger than the final value we want. As a formal matter one can simply cancel the radiative correction against a large bare contribution

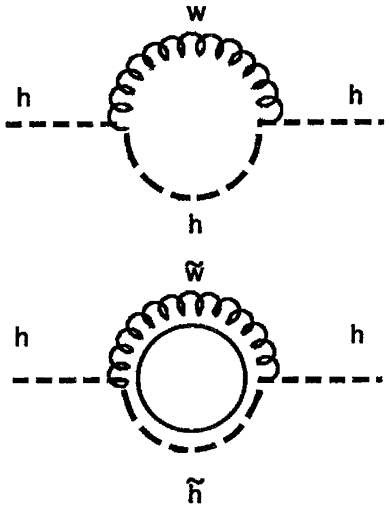


FIGURE 5 Contributions to the Higgs field self-energy. These graphs give contributions to the Higgs field self-energy that separately are formally quadratically divergent, but when both are included the divergence is removed. In models with broken supersymmetry a finite residual piece remains. If one is to obtain an adequately small finite contribution to the self-energy, the mass difference between Standard Model particles and their superpartners cannot be too great. This—and essentially only this—motivates the inclusion of virtual superpartner contributions in Fig. 6 beginning at relatively low scales.

of the opposite sign, but in the absence of some deeper motivating principle this seems to be a horribly ugly procedure.

In a supersymmetric theory there will be, for any given set of virtual particles circulating in the loop, another graph with their supersymmetric partners circulating. We will be assured adequate cancellation if and only if supersymmetric partners are not too far split in mass—in the present context, if the splitting is not much greater than the weak scale.

This important argument suggests the relevance of "low-energy" supersymmetry. Specifically, it suggests that we should expect to see the supersymmetric partners of known particles with masses not much beyond 1 TeV—that is, accessible to future accelerators based on known technology and economics.

The effect of low-energy supersymmetry on the running of the couplings was first considered by S. Dimopoulos, S. Raby, and F. Wilczek in 1981. One might have feared that such a huge expansion of the theory, which essentially doubles the spectrum, would utterly destroy the approximate success of the minimal SU(5) calculation. This does not occur, however. Since supersymmetry is a space–time rather than an internal symmetry, it does not affect the group-theoretic structure of the calculation. Thus to a first approximation the absolute rate at which the couplings run with momentum is affected, but not the relative rates. The main effect is that the supersymmetric partners of the color gluons, the gluinos, weaken the asymptotic freedom of the strong interaction. They tend to make its effective coupling decrease and approach the others more slowly. Thus their merger requires a longer lever arm, and the scale at which the couplings meet increases by an order of magnitude or so, to about 10^{16} GeV. This increase in unification scale significantly reduces the predicted rate for proton decay through exchange of the dangerous color-changing gauge bosons, so that it no longer conflicts with existing experimental limits.

Upon more careful examination, there is another effect of low-energy supersymmetry on the running of the couplings, which although quantitatively small has become of prime interest. There is a small but precisely calculable change in the calculation due to some of the additional particles (to be specific, although perhaps not informative, it comes from the additional Higgs particles and Higgsinos.) From Fig. 6 you see that the change is a most welcome one. The minimal implementation of supersymmetric unification puts the unification of couplings calculation into excellent agreement with experiment.

Since the running of the couplings with scale is logarithmic, the unification of couplings calculation is not terribly

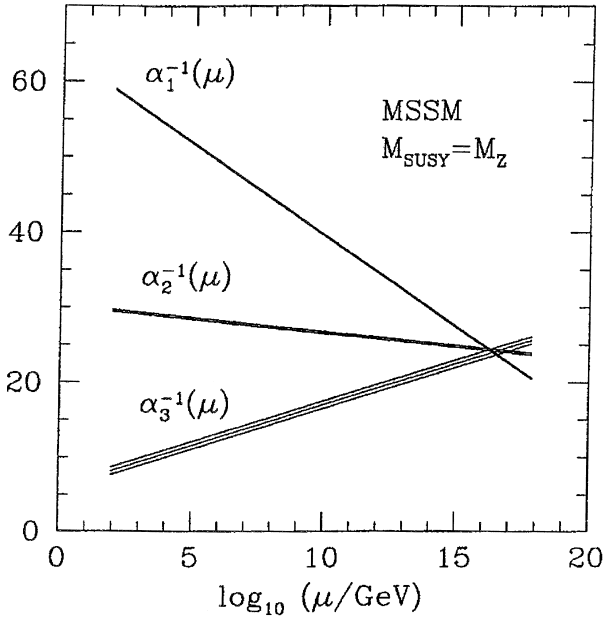


FIGURE 6 Evolution of the effective (inverse) couplings in the minimal extension of the Standard Model, to include supersymmetry. The concepts and qualitative behaviors are only slightly modified from Fig. 4 (a highly nontrivial fact!) but the quantitative result is changed, and comes into adequate agreement with experiment. I would like to emphasize that results along these lines were published well before the difference between Figs. 4 and 6 could be resolved experimentally, and in that sense one has already derived a successful *prediction* from supersymmetry.

sensitive to the exact scale at which supersymmetry is broken, say between 100 GeV and 10 TeV. Many ambiguities, both theoretical and experimental, result if we try to push the calculation beyond the accuracy displayed in Fig. 6.

If we accept all this at face value, we can draw three profound conclusions:

- The great principles of quantum field theory, including the pillars of dynamical symmetry, asymptotic freedom, and spontaneous symmetry breaking on which the Standard Model rests, remain valid far beyond the domain in which they were originally inferred.
- The dynamic symmetry of the world extends well beyond the SU(3) × SU(2) × U(1) of the Standard Model, at least to its consolidation in SU(5).
- Supersymmetry has, through the effect of its virtual particles, already been discovered.

During the next decade, as the Large Hadron Collider (LHC) comes into operation, this last item will be either triumphantly confirmed or proven illusory.

V. UNIFICATION WITH GRAVITY?

We can attempt to extend the calculation of Fig. 6 further, to include gravity. The couplings of the Standard Model gauge fields run only logarithmically with energy, due to vacuum polarization effects. But the graviton couples directly to energy, and so its effective strength grows with energy even classically, and much faster. The effective gravitational coupling is exceedingly feeble, compared to the other interactions, at accessible energies. But it becomes comparable to them at $Q \sim 10^{18}$ GeV, near the Planck scale. Thus, the unification of couplings calculation, boldly extended to include gravity, works reasonably well. This is quite a remarkable result, since the physical ingredients entering into the calculation are so disparate.

If we take this seriously, it goes a long way toward explaining a profound fact about physics that otherwise appears quite mysterious: gravity, at a fundamental level, is extraordinarily feeble. As a quantitative measure of this feebleness, we might take the ratio of the gravitational attraction to the electrical repulsion between protons—it is about 10^{-38}. In our calculation this ratio emerges dynamically, as the ratio of exponentials of inverses of the observed Theory of Matter coupling constants. No spectacularly small quantities are involved. The big ratio of mass scales arises, basically, because the strong coupling α_3 at the unification scale is about 1/25, and the couplings run only logarithmically. Therefore quite a long run is required before one reaches the scale where α_3 approaches unity, protons are assembled, and ordinary life begins.

Putting it another way, when confronted with the tiny value of the proton mass (approximately 1 GeV, as an energy) in units of the Planck scale (10^{19} GeV), it has been traditional to ask, "Why is the Planck scale so large?" From the point of view emerging here, the better question is, "Why is the proton so light?" And the answer is that the QCD effective coupling is just a little on the small side at the Planck scale, but it runs so slowly that it become truly strong only at much lower energies, capable of binding quarks into protons. Since the mass of the proton predominantly arises from this binding energy, through $m = E/c^2$, the big mismatch of scales has been demystified.

SEE ALSO THE FOLLOWING ARTICLES

ATOMIC PHYSICS • COSMOLOGY • ELECTROMAGNETICS • GRAVITATIONAL WAVE PHYSICS • GROUP THEORY, APPLIED • MECHANICS, CLASSICAL • PARTICLE PHYSICS, ELEMENTARY • QUANTUM MECHANICS • RELATIVITY, GENERAL • RELATIVITY, SPECIAL

BIBLIOGRAPHY

Cheng, T. -P., and Li, L. -F. (1984). "Gauge Theory of Elementary Particle Physics," Oxford University Press, New York.

Dimopoulos, S., Raby, S., and Wilczek, F. (1991). "Unification of couplings," *Physics Today* **44-10,** 23–33.

Feynman, R. (1985). "QED: The Strange Theory of Light and Matter," Princeton University Press, Princeton, NJ.

Weinberg, S. (1995, 1996, 2000). "Quantum Theory of Fields—I: Foundations; II: Modern Applications; III: Supersymmetry," Cambridge University Press, New York.

Weyl, H. (1949). "Philosophy of Mathematics and Natural Science," Princeton University Press, Princeton, NJ.

Wilczek, F. (1999). Quantum field theory. *Rev. Mod. Phys.* **71,** S85–S95; also published in "More Things in Heaven and Earth—A Celebration of Physics at the Millennium" (B. Bederson, ed.), Springer-Verlag, New York.