The Laws of Physics and Fundamental Particles

by Peter Rowlands

THE LAWS OF PHYSICS AND FUNDAMENTAL PARTICLES

by PETER ROWLANDS

Published by Newart Visual Aids Ltd.,
Runcorn, Cheshire.

Abstract

The laws of classical mechanics, electromagnetics, quantum mechanics and thermodynamics and the basic structures of fundamental particles are derived from an arrangement of the physical parameters space, time, mass and charge as a group of order 4.

(1) Forces and Particles

The two aims of fundamental physics are to explain the laws governing the forces of nature and to account for the existence of the fundamental particles. It is generally believed that these aims are closely related and progress has been made by describing both forces and particles in terms of unifying symmetries. Symmetry appears to be the key to unification, but the most fundamental symmetry is not that between the forces or the particles, for these symmetries do not explain the description of both forces and particles in terms of the apparently arbitrary parameters space, time, mass and charge. In fact, there is evidence for an entirely new symmetry between these parameters which is responsible for all the individual symmetries between forces and particles and which is the unifying principle for all the laws of physics.

(2) The Laws of Physics

In physics, inferences are made from observations of a particular kind. The observations are numerical measurements whose significance only becomes apparent when they are applied to physical laws which have a specifically mathematical form. The components of these mathematical expressions are described as physical parameters and all physical measurements are made with the object of supplying particular numerical values for these parameters.

Now, certain physical laws have long been recognised as more fundamental than all the others. Classical mechanics, electromagnetics and quantum mechanics, which between them account for nearly the whole of physics, are founded on only a few general principles. It is one of the main aims of contemporary physics to unify these fundamental laws in a comprehensive system.

In the past, attention has been focused directly on the laws themselves and on their mathematical form; unification has been assumed to result in expressions of increasing mathematical complexity, introduced arbitrarily, as in the case of general relativity. However, all the really fundamental laws are, intrinsically, statements about the properties and relationships of a few equally fundamental parameters, namely space, time, mass and charge. These parameters have distinct characteristics, which are certainly not arbitrary since they are used to provide consistently successful results in physical theories, yet it is the practice of physicists to regard them as elementary components of physical laws whose origin is inexplicable. In fact, however, they are neither elementary nor inexplicable and they are much more significant in principle than the laws of physics, for every single law of physics, whatever its mathematical form, can be derived entirely from a structural grouping of only these four basic parameters.

Theoretical physics has failed to find a true unification because it has merely assumed the existence ad hoc of space, time, mass and charge and has not discovered why we use such parameters to make measurements. Because this question has remained unanswered, physicists have attempted a fundamental unification of physics without knowledge of the fundamentals. The laws of physics are only secondary properties of this fundamental structural unit of all physics and knowledge based only on them is necessarily incomplete.

(3) The Physical Parameters

The fundamental parameters of classical mechanics, electromagnetics and quantum mechanics are space, time, mass and charge. The laws of strong and weak nuclear interactions would seem to require two other parameters, of a similar nature to charge, but we can instead redefine our concept of charge to refer to the source of any nongravitational interaction. We can say that there are three types of charge or three "dimensions" of charge.

Now, although we think of physical parameters as objects of measurement, it is a remarkable fact that the only method of direct measurement is through the parameter space. Measurement of space involves an element of enumeration or division into a finite number of discrete components; it is a counting process, based on the series of natural numbers or integers, and, although we can choose the grain size of space to be indefinitely large or small (because it is not a conserved quantity), we cannot alter the method of its measurement. It is, of course, possible also to describe the measurement of time as a counting process; thus we measure time by the number of repetitions of a regular periodic event such as the oscillations of a steel spring or a quartz crystal or the revolution of the earth; however, all these measurements of time are really inferences from direct measurements of space; time is only measurable under the special conditions which prevail when a particular force is assumed to cause an object to traverse a certain distance with complete regularity; we measure distances and make assumptions about the forces acting to convert them into fixed time intervals. Measurements of mass and charge are also, ultimately, measurements of space; we have even less knowledge of these quantities than we have of time; their presence is only inferred when they interact to produce forces, which are, of course, only known through measurements of space and time.

It may seem obvious, in view of its unique capacity for direct measurement, to describe space as a real parameter, that is one to which we assign real number values in physical equations. However, it is probably less obvious that other parameters are not necessarily real in this sense. Thus, if time were imaginary, we could easily explain the appearance of the term $(ict)^2$ in the 4-vector equation

$$r^2 = x^2 + y^2 + z^2 + (ict)^2$$

and also discover why time is always measured through the effects of force and acceleration (as t^2) and never through the effects of motion (as t). We could also find a similar distinction between mass and charge. Thus masses are always positive whereas charges are either positive or negative; the force between identical masses is attractive (positive), whereas that between identical charges is repulsive (negative) (the signs are reversed by convention); these facts could be explained simply by assigning only real values to mass and only imaginary values to charge. The imaginary nature of charge would explain why every particle has an antiparticle which has an identical mass but which behaves in the opposite sense in strong, weak and electromagnetic interactions; in mathematical terms $+i$ and $-i$ are indistinguishable and so there must be both positive and negative solutions to all equations involving them.

Thus we postulate that the parameters have the following properties:

space	real
time	imaginary
mass	real
charge	imaginary

This real-imaginary division does not, of course, specify all the properties of the parameters which we can discover in this direct way. Thus mass and charge have a property in common. According to two of the most fundamental laws of physics, mass and charge are conserved within any physical system. Space and time, of course, are not. It is possible to make any number of measurements of space and time which are each separate and individual operations. It would, therefore, be meaningless to discuss the conservation of space and time. On the other hand, we cannot make separate and individual determinations of mass and charge. Even though we can assume the existence of individual masses and charges, we can only determine their effects, and even their presence, by their interactions with each other; we can only know the total effect of a mass or charge by its mutual effect on all other masses or charges. It seems that not only are mass and charge opposite to space and time in terms of conservation, but also in all related properties.

The table of parameters now becomes:

space	real	nonconserved
time	imaginary	nonconserved
mass	real	conserved
charge	imaginary	conserved

It begins to look as if, to complete the symmetry, there might be a property which is common to space and charge, and an opposite property which is common to time and mass, and indeed there is. Space is three-dimensional and denumerable (which means that it is measured by a counting process) and so also is charge, for there are three types of charge and each exists in individual units. Time and mass are both one-dimensional and also non-denumerable, for time is a continuum measurable only when a periodically discontinuous function of distance, and mass has a continuum of values determined only through the forces which represent the interaction of all masses. There are no real point-masses in physical systems.

Now, it would seem likely that the properties of dimensionality and denumerability are related and that the presence of one necessarily implies the presence of the other. It is significant that the dimensions of space (and those of charge in forces) are combined as squares according to the Pythagorean equation. This equation seems to offer, for space, an equally valid alternative method of combination of individual integral measurements to that of the direct addition used in the ordinary measurement of distances, and it is unique in this respect, for the Pythagorean integers are a unique series; it also introduces the unique directional or vector properties of space. For charge, it offers a combination of imaginary numbers as real values. The Pythagorean equation requires the existence of independent systems of measurement, or dimensions, for both space and charge. Charge is an imaginary quantity and the minimum number of independent imaginary numbers in the simplest possible system (the quaternions) is three, and so there must be at least three dimensions of charge, and, by analogy, three dimensions of space.

(4) A Group of Order 4

Assigning the arbitrary symbols +a, +b, +c to the properties of space, the complete table of the properties of the parameters now becomes:

space	real	nonconserved	denumerable (3-D)	+a	+b	+c
time	imaginary	nonconserved	nondenumerable	-a	+b	-c
mass	real	conserved	nondenumerable	+a	-b	-c
charge	imaginary	conserved	denumerable (3-D)	-a	-b	+c

It may now be apparent that space, time, mass and charge form a noncyclic group of order 4 (D2), with the multiplication table as set out below:

*	space	time	mass	charge
space	space	time	mass	charge
time	time	space	charge	mass
mass	mass	charge	space	time
charge	charge	mass	time	space

The multiplication rule for this group would be:

$$+a \ast +a = -a \ast -a = +a$$
$$+a \ast -a = -a \ast +a = -a$$

and similarly for b and c. Here, space is the identity element, though the symbols could be rearranged to assign this property to any of the other parameters; each element is its own inverse; and the multiplication rule identifies the group as Abelian. D2 is the simplest group with dihedral symmetry; the only simpler groups are cyclic and, except for C2, do not preserve the interchangeability of the elements.

The group accommodates all the known properties of the physical parameters and is the fundamental physical fact on which all the laws of physics are ultimately based; they can now be shown to be essentially secondary effects or particular cases derived from it. The laws of physics suggest how information from the parameters must be organized to fulfil their individual properties; they arise because information from the parameter group is both exclusive and governed by the absolute symmetry which exists between its elements. The descriptions of parameters as "real", "conserved" or "denumerable" are the immediately identifiable aspects of a physically complex but mathematically simple symmetric system.

We can be certain that such an extraordinary and unexpected symmetry between the physical parameters is not accidental, and that it must be related to the fact that space is the only parameter used in direct measurement. Without going into extensive speculation, it is possible to see this as an example of the fact that universal reality does not have the characteristics or properties of any particular part of it. It is neither entirely measurable nor unmeasurable, discontinuous or continuous, real or imaginary. Physical parameters are assumed to be universally applicable. To ensure this we must create a system in which every characteristic of one parameter is countered by the existence of some opposite characteristic in another. The actual characteristics of the parameters are then determined by the fact that we want to create a universal system of measurement. Space has all the properties required of a parameter of measurement - it is real, can be applied separately in individual situations (not conserved, or of varying grain size), and denumerable. The other parameters exist to ensure that the universal application of this parameter of measurement does not impose its characteristics on reality as a whole.

(5) Classical Mechanics

There are only three really fundamental principles of classical mechanics: the definition of force and the statements that the vector sum of the forces in any system and the divergence of the gravitational field due to a distribution of masses are both zero. These are, respectively, Newton's second and third laws of motion and law of gravitation. All other laws and principles, including conservation of energy, momentum and angular momentum, which respectively express the translation symmetry of time and the translational and rotational symmetry of space in dynamical systems, are derived from or contained within these.

Essential to classical mechanics is the distinction between mass and charge as required by the parameter group. Mass and charge have the common property of conservation; this means, also, that, unlike measurements of space and time, we can only determine the measurable effects of any mass or charge by its mutual effect on all other masses or charges. Mass and charge, however, are different in all other respects. Charges are denumerable and so exist as localized units; they can be positive or negative individually and so it is possible to have a zero local charge. Masses, on the other hand, are not denumerable; they represent a true continuum of interaction,which means that it is impossible to define a system excluding mass. (This is why systems with negative, and consequently zero, masses are never found in nature.) Classical mechanics defines the laws of those systems with zero charge.

Let us, therefore, imagine a system with a single mass (or infinitesimal element of a mass distribution) and no charge. We know that the effect of a mass is determined by that of all other masses; we also know that the conservation of space and time is not intrinsic to the system, so that we may include within the system a varying space and time of which we have independent knowledge. Since the information provided by the parameter group is, by definition, exclusive, we have here the full description of the system. Now, since we must know the effect of all other masses on the mass in our system to know that we have a system at all and since we can only know the variation of space and time, then it follows that the effect of all other masses on the mass in our system is directly expressed by the known variation of space and time.

The characteristics of each parameter impose particular restrictions on this variation of space and time. The effect of other masses on the mass of our system cannot be arbitrary even though individual variations of space and time are; the only way to avoid this difficulty is to suppose that the effect makes one parameter vary with respect to the other. In fact, because space must remain vector additive (via the Pythagorean equation), for direct measurement we make space (x) vary with respect to time (t) by defining a new quantity, velocity (dx/dt), whose differential form reflects the continuous nature of the interaction of masses. However, since we must have a real quantity to express the directly available information concerning the interaction of masses, then we cannot use an imaginary differential such as velocity. Imaginary information is not directly available and can only be recovered if it is related to real information. We, therefore, have to define a real differential, acceleration (d^2x/dt^2) which is at once directly accessible and which retains the vector properties of space and the continuum properties of time. This, then, is the quantity which describes the effect of the mass continuum (or, indeed, any other external factor) on the isolated mass of our system.

However, it is still necessary to ensure that the system conserves mass. Space and time are still the only variable parameters, that is ones of which we have direct knowledge. Since, as nonconserved quantities, they are intrinsically arbitrary, they give no information about the actual mass (m) of the system and do not specify that this quantity is conserved; acceleration is independent of mass because we have no direct knowledge of mass. However, the conservation of mass must be specified at all times in any system which we care to define. In effect, the mass m must be associated at all times with the quantity which we measure as acceleration. Mathematically, this is accomplished by defining a quantity which is their product, force (md^2x/dt^2) (a formulation which also ensures that mass remains an additive quantity). This definition is, of course, Newton's second law of motion.

Of course, we cannot have direct knowledge of a force because we have no direct knowledge of mass, and we could not define such a quantity to exist if this knowledge was not available in some form. A system cannot be posited to contain *information* that cannot be discovered. However, we *can* define that the total of such a quantity within a system is zero, and such a definition would ensure that the mass of the system is conserved. This, indeed, is Newton's third law of motion - to every action there is an equal and opposite reaction - and, for the interaction of masses, it is an expression of their mutual attraction. There is an equal and opposite force between the mass of the system and all external masses; this interaction is attractive because the acceleration is in the same direction as the space vector. Newton's second and third laws of motion relate directly to the translational and rotational symmetries of space and time which are responsible for the conservation laws of energy, momentum and angular momentum.

Now, if we had a system containing several masses, there would be a series of interactions, producing an acceleration in each mass which would depend, not on its own value, but on that of all the others. The acceleration produced in a mass of any value at any point in space would be described as the gravitational field (g) at that point. Since such accelerations are additive vectorially, then this total gravitational field would be the vector sum of all the gracitational fields (g_n) which would result if each mass was imagined to be alone in empty space and responsible for its component of gravitational field. However, if we had a single isolated mass in empty space, we would be unable to detect any variation of its gravitational field with respect to the coordinates of space, because, if any such variation existed, then we would be able to detect the presence of mass independently of other masses. Again, we would be positing the existence of information that could not be discovered (acceleration must be independent of mass yet inconceivable without it), and again we can avoid this difficulty by equating the information to zero. In mathematical terms, we can express this by the equation

$$\nabla . g_n = 0.$$

In the absence of other masses, g_n, which defines the nature of the mass, is unknown as a parameter with any relation to physical measurement. We cannot discover the position, or even the presence, of this mass in space. In classical mechanics, the vector g_n is expressed in terms of a scalar potential ϕ_n, where

$$g_n = - \nabla \phi_n .$$

Thus

$$\nabla^2 \phi_n = 0$$

and, summing all the individual potentials for a distribution of masses in space, we arrive at Laplace's equation

$$\nabla^2 \phi = - \nabla . g = 0.$$

On integration and application of the principle of mutual interaction of masses, this equation yields Newton's inverse square law of gravitation.

(6) Electromagnetics

Since mass is a universal component of physical systems, a system containing charges must also contain masses. The addition of charges to a system adds no new direct information but it does add a new requirement; we must know how the charges in any system interact with all others; and we can only know of this effect through the concepts of force and acceleration. The interaction of charges must, therefore, be analagous to that of masses and must be so defined as to include the effects of the mass of the system. We thus define an electric field $\underline{E}$ and electric (scalar) potential V by analogy with the gravitational terms $\underline{g}$ and ϕ, the electric field at any point being the force which would be exerted on a unit charge at that point. By analogy with gravitation, we can say that, for a system of localized stationary charges

$$\nabla . \underline{E} = - \nabla^2 V = 0$$

and hence the law of force between individual charges would be inverse square.

By this definition, we unavoidably create a situation in which masses and charges become numercially related. The result of this may be discovered if we examine the group-multiplication properties of each of the parameters. In terms of group division, we invariably find that

$$\text{mass/charge} = \text{space/time}$$

or its inverse (since each element of the group is its own inverse). It seems that, in order to preserve the complete symmetry of the group, we must introduce a fixed relation between space and time whenever we have one between mass and charge. Thus, we introduce the term c (the velocity of light) and the 4-vector system for the space-time manifold

$$r^2 = x^2 + y^2 + z^2 + (ict)^2$$

as the symmetrical complement to the full quaternion (1, i, j, k) representation of the sum of the forces, in absolute units and over unit distance, due to mass (m) and electromagnetic (e), strong (s) and weak (w) charges

$$F = m^2 + (ie)^2 + (js)^2 + (kw)^2$$

The constancy of c is, of course, a fundamental component of its definition; it depends on the constancy of the value of a unit of charge; if this were increased, for example, then c would decrease.

In physical terms, the 4-vector system acts as an expression of the finite time of interaction on our knowledge of a charge's location, and leads to a new equation for the variation of electric potential for a system of non-stationary charges in free space:

$$\Box^2 V = \left(- \nabla^2 + \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) V = 0$$

In fact, the 4-vector system and the inverse-square law for static charges are, together with the laws of force derived from classical mechanics, all the principles we need to be able to derive by standard arguments all the laws of electromagnetics: namely Maxwell's equations, the Lorentz transformations and the equivalence of mass and energy.

(7) Special Relativity and Conservation of Energy

The Lorentz transformations are derived by defining a velocity of light which is constant in all frames of reference. The symmetry of space and time requires only that c is constant in empty space; there is no fundamental reason for assuming that it is invariant. Experimental evidence, including a series of observations of the light from double stars, shows that the velocity of light is independent of the velocity of its source, but it may yet be dependent on that of the observer. (This possibility cannot be ruled out until there is a direct comparison of measurements taken at different speeds; indirect tests, such as the Michelson-Morley experiment, can be explained without assuming the c invariance.) In the latter case, the behaviour of fundamental particles, which originate as energy states of charge-induced fields, would be identical to that described by the Lorentz transformations with invariant c; particles would still be "relativistic" in the mathematical sense.

According to some expositions, special relativity is distinguished from Newtonian mechanics by its recognition of the fact that the velocity of light is finite. However, Newtonian mechanics does not say that the velocity of light is infinite; only that the velocity of propagation of gravity is infinite. Special relativity is usually described in terms of a kinematical argument in which the time of events is determined by the speed of light signals. The concepts of length and time are described in operational terms. However, as we have seen, the properties of electromagnetic systems are independent of these kinematic considerations; the Lorentz transormations can be derived, if required, simply from the known symmetry of space and time.

Certainly, if we take an operational definition of length and time, we may actually be using light signals to determine space-time coordinates in dynamical systems. However, we must make a distinction between those systems in which there is an intrinsic 4-vector relation between space and time and those in which this relation occurs accidentally because we have measured space and time through an intrinsic 4-vector (typically, an electromagnetic) system with the additional assumption that c is invariant; the kinematical arguments of special relativity are a description of such accidental effects, since they do not derive from direct considerations of 4-vector systems. Thus, planetary positions may be determined by using an electromagnetic method of measurement and hence an electromagnetic space-time, but this does not mean that such a space-time is intrinsic to the dynamics of planetary orbits. The differences between the operational and absolute space-times may produce apparent effects which would otherwise be accepted as evidence of non-Newtonian behaviour in gravitational systems.

The derivation of the 4-vector system from first principles means that the strictly kinematical postulates of special relaticity are unnecessary. There is no evidence, in any case, that the Lorentz transformations apply to anything other than systems containing charges; all the experiments adduced as support of special relativity rely on an application of Maxwell's equations. There is, in fact, a fundamental reason for believing that they do not apply to the gravitational interaction. Charges are by definition localized units, their interactions taking place over a finite time, but masses are a true continuum of interaction and the gravitational force must, therefore, be instantaneous. This could be deduced from the property of group multiplication; for systems in which charge is zero, the space/time ratio must be infinite. This is the Newtonian position and it is not affected by whether the velocity of light, which is relevant only to electromagnetic-type systems, is finite or even invariant.

A corollary of this property of gravitational systems is the fact that the mass associated with gravitational energy, as defined by E/c^2, must be zero. In cases where there is no intrinsic connection between space and time, there can be none between mass and energy. In fact, the fixed relation between mass and energy, which is represented by the equation $E = mc^2$, is really the result of a fixed relation between the fundamental parameters mass and time (since we associate conservation of energy with symmetry under time translation); such connections would be expected between every pair of parameters within the group, once the relationship between any single pair (mass and charge in the case of electromagnetism) had been established. Certainly, given such a connection, we can establish that the correct mathematical form of the relationship is given by $E = mc^2$, but it is especially significant that this equation is not derived by special relativity, but rather defined. Special relativity does not prove, by any compelling mathematical argument that there must be such a relationship between mass and energy, only that any such relationship must be of this particular form, if it exists. There is every reason to believe that it exists for all electromagnetic-type systems, including all systems involving fundamental particles, but we do not know that it exists for all systems.

If we accept this argument it will be necessary to ensure that the interconversion of mass and energy is strictly defined to apply only to electromagnetic and other 4-vector systems and not to those involving gravitational potential. This should be possible since both types of system are individually conservative (gravitational and electromagnetic systems are frequently coupled in physical processes but their total energies are not really subject to interconversion) and if it were established we could then remove both the nonlinearity of the gravitational field required by general relativity and the related concept of the infinite gravitational collapse to a black hole of a body of size less than the Schwarzschild radius. Such concepts have always implied that energy and mass conservation could be violated in black holes whereas the relation $E=mc^2$ was originally introduced to preserve both.

(8) Gravitation and Inertia

With the derivation of the laws of classical mechanics from first principles there is no problem in identifying gravitational with inertial mass. The properties of inertia and gravitation are both fundamental aspects of the original definition of mass as a parameter and it is not necessary to look for phenomenological explanations such as those relating gravitation to the "inertial forces" produced by accelerating frames of reference. The "weak" principle of equivalence, equating gravitational and inertial effects on a purely local scale, is, of course, an obvious consequence of the identification of the gravitational and inertial masses but it is only of phenomenological significance if we also accept the general principle of relativity that all physical phenomena obey the same laws for all observers whether in inertial or in accelerating frames. In fact the success of the restricted principle that all the laws of physics are covariant with respect to transormations between inertial frames or are the same for all inertial observers need not necessarily lead us to any generalised version, for there is a fundamental reason why these particular frames or observers should be distinguished from all others. This is that the absolute velocity of inertial frames has no physical meaning because it is an imaginary quantity, whereas the absolute acceleration of noninertial frames has an exact physical meaning because it is real. There is now no mystery behind the fact that, on the Newtonian explanation velocity through absolute space cannot be detected whereas acceleration can,and there is, consequently, no need to devise a "physical" explanation.

It is also important to note that the special or restricted principle of relativity can be maintained using either the Galilean or the Lorentz transformation. We can apply the former to electromagnetic systems assuming c varies with the velocity of the observer or the latter assuming c is invariant. We have no a priori reason for assuming, as in special relativity that c is invariant; several experiments have shown that c is independent of the velocity of its source but no experiment, not even one of the Michelson-Morley, has really tested whether c is dependent on the velocity of the observer, which is, on logical grounds, the more probable assumption. Of course, even if experiments do show that the measured velocity of light depends on the state of motion of the observer, we could always arbitrarily define c as invariant and derive apparent length contractions and time dilations using the Lorentz transformations; this would mean that we could still derive all the relevant information concerning electromagnetic systems via special relativity but it would also mean that the length contractions and time dilations would not be real effects to be transferred to dynamical systems where the velocity of light was not directly involved. (On the other hand, superluminal motion, as observed in a number of compact extragalactic radio sources, would perhaps be a real effect and not simply an "illusion" based on the fact that time dilation within the source would be greater than on earth.) In practice, it is often convenient to retain special relativity for describing electromagnetic systems, and this technique has been applied successfully in the study of fundamental particles, but there is no evidence that the theory is needed to "correct" Newtonian dynamics by redefining the measurement of length and time, while at the same time, if we assume an invariant c, we can derive the Lorentz transformations for electromagnetic systems without recourse to a kinematical argument concerning simultaneity. We can, in fact, define two types of special relativity - one based on dynamical arguments and applied to all systems, and the other, based on the mathematical assumption of invariant c combined with the Lorentz transformations derived by symmetry and applied to 4-vector systems; only the latter is relevant to the behaviour of fundamental particles and electromagnetic radiation.

The combination of the "weak" principle of equivalence (meaning the local dynamical equivalence of a gravitational field and acceleration) and the technique of special relativity leads to interesting predictions concerning the interaction of gravitational fields and electromagnetic radiation, for we can thus derive the observed gravitational redshift, deflection of light and radar time delay without introducing any new physical principle.(1) The gravitational redshift can, in fact, be derived very simply from conservation of energy, assuming $E = mc^2$, but in the language of special relativity, this is equivalent to the effect of the slowing down of a standard clock in a gravitational field; this time charge and a similar change in the length of a standard measuring rod contribute equally to the change of refractive index of space in the vicinity of a gravitational field, producing the full values for the deflection of light and radar time delay. A "Newtonian" calculation proposed as long ago as 1801 by Soldner and deriving half the value for the deflection of light is irelevant because light is an electromagnetic system, not a system of particles subject to the ordinary rules of Newtonian dynamics.

Thus far we have had no need to introduce a theory of gravity based on phenomenological considerations to explain any of its effects; the inverse square law of force, the identity of gravitational and inertial mass, the special principle of relativity, the preferred role of inertial reference frames and all the known interactions with electromagnetic radiation are all logical consequences of the initial properties of the group of physical parameters. There is no a priori logical reason to accept the general principle of relativity and the identification of gravity and inertial forces or to accept any preferred role for the velocity of light or the principle of equivalence. In particular, there is no reason for deriving the General Theory of Relativity as a logical necessity; the only justification for such a theory would lie in its ability to predict results beyond those of the Newtonian system. Even so, the General Theory, with its curved space-time of uniform refractive index, would be more of a mathematical device than a physically justifiable theory, for, without a preferred role for c, it would be more logical to assume a flat space-time with a variable refractive index. Atkinson (2) and Cornish (3) have interpreted the equations of motion of Einstein's theory with respect to a flat space-time, while Dicke (4) has used a variable scalar refractive index to account for the classic general relativistic effects.

Since the existence of black holes and gravitational waves has not yet been convincingly demonstrated, there is only one prediction of the General Theory which is apparently not duplicated in the Newtonian system. This is the precession of a planetary orbit centred on a massive spherically symmetric gravitational field. The equation of motion of the planet is assumed to be a geodesic in a Riemannian space-time whose curvature is attributed to the presence of the solar mass; the perihelion precession is derived by substituting the equation of an ellipse into an expression derived from the Schwarzschild solution of the field equations. The theory has claimed success in accounting for the orbits of Mercury, Venus, Earth and the asteroid Icarus.

Setting aside Dicke's arguments for a measurable effect due to the solar quadrupole moment (5) (6), it seems that the case for non-Newtonian behaviour in planetary orbits - or nonlinearity of the Sun's gravitational field - depends on the employment of the geodesic equation in the calculation of perihelion precession. This is equivalent to the assumption that gravitational systems must be described by a 4-vector space-time. The gravitational field thus becomes nonlinear if we accept the identity of inertial forces and gravitation.

Nevertheless, we have already advanced arguments to the effect that gravitation is not described by a 4-vector space-time and it is possible to derive an entirely different explanation for the inertial forces which are responsible for the perihelion precession. This is that the space-time actually measured in astronomical observations is operational (4-vector) and not identical to the absolute space-time of classical mechanics. This operational space-time depends on the velocity of light, which is affected by the presence of the gravitational field of the Sun. The result - with time and distance changes of order GM/c^2r - is an apparent rotation of the coordinate system, which expresses itself in the planetary orbits as a perihelion precession. To retain the laws of classical mechanics in the noninertial frame created by the use of this 4-vector space-time, we assume that the rotation, which will be equivalent to the deflection of a light ray, is caused by the presence of fictitious inertial forces. We, therefore, simply add an inertial force term, derived from the relativistic expression for the deflection of light ray, to the classical vis viva integral, to obtain the accepted general relativistic expression which predicts the perihelion precession:

$$\left(\frac{dr}{dt}\right)^2 + \left(1-\frac{2GM}{rc^2}\right) r^2 \left(\frac{d\phi}{dt}\right)^2 = \frac{2GM}{r} - E .$$

If this explanation is correct, we have no criterion for deciding on experimental grounds whether the gravitational field is linear or nonlinear. The evidence so far acquired suggests that we can accept Newtonian mechanics in gravitational systems, special relativity in 4-vector systems and general relativity to first order in gravitational systems complicated by inertial forces. General relativity will presumably be valid to higher finite orders in systems which we cannot as yet investigate experimentally, but it will certainly fail at the level where it predicts singularities due to field nonlinearity. (Black holes of the Laplace type will, of course, remain a theoretical possibility.) General relativity thus becomes not so much a physical theory of gravity, as in the traditional presentation, as a mathematical theory of inertial forces. We retain the option of gravitational field nonlinearity but see no reason, experimental or theoretical, for introducing it.

General relativity as a gravitational theory is founded on the relation between gravitation and inertia and, in particular, the idea of Mach that all the inertial properties of matter can be attributed to the gravitational attractions of distant stars and galaxies. However, not only have we no experimental evidence for an extra gravitational force due to the stars or galaxies in addition to the static interaction, but we also have no knowledge of whether general relativity would actually produce the effect that Mach predicted. For all its mathematical elegance, traditional general relativity contains many ad hoc features - the preferred role of c, the arbitrary assumption of the Newtonian approximation, the choice of a Riemannian space-time - and it leaves many things, such as the uniformity and isotropy of the universe, unexplained. In an alternative approach to a nonlinear theory, we could describe gravity, like the electromagnetic force, in terms of a flat four-dimensional space-time and develop equations analagous to Maxwell's equations to take account of the supposed inertial forces. This would introduce a nonlinearity into the gravitational field which would be equivalent to applying $E=mc^2$ to gravitational potential, as discussed in the previous section. The effects would presumably be similar to those predicted by general relativity, including the existence of black holes and gravitational waves; the differences would result from replacing the curved space-time with a flat space-time of variable refractive index. However, in this case, the analogy between gravity and electromagnetism, abandoned by general relativity, would be retained, and the approximation to the inverse square law, which is an arbitrary assumption in general relativity, would have a more logical foundation.

Nevertheless, solid evidence for any non-Newtonian behaviour in gravitational systems is virtually nonexistent, while the purely Newtonian explanation of gravity has a far more logical foundation than was hitherto believed. Black holes and quantum gravity present formidable theoretical difficulties for post-Newtonian theories, the presence of singularities and infinities usually indicating that a theory is defective. It may also be significant that all attempts to combine electromagnetism and gravity in a single geometrical theory have been totally unsuccessful. Such theories were once thought to be required to explain effects which we now know are not post-Newtonian and there is a considerable probability that the Newtonian theory of gravity, which is the only one with a completely logical foundation, is also the only true description of the phenomenon.

(9) Particles

The laws which we have derived for systems involving charges are those which we know are valid for the electromagnetic interaction. However, our definition of charge also includes the sources of strong and weak interactions. These forces appear to be governed by laws of the form

$$F = \frac{g^2}{r^2} \exp \left(\frac{-mcr}{\hbar} \right)$$

where g^2 is a coupling constant analagous to the electromagnetic term $e^2/4\pi\varepsilon_0$, and m is the mass of the particle which transmits the energy of interaction. However, if this mass were zero, as we should expect it to be for complete symmetry with the electromagnetic force, then the strong and weak forces would be exactly inverse square. Such deviations from ideal behaviour in charge systems seem to be generally associated with the existence of particles of definite mass. It seems that a complete understanding of these forces cannot be accomplished without some consideration of the theory of particles.

The properties of the fundamental particles are as much a consequence of the symmetrical arrangement of the physical parameters as are the laws of physics. In this case they arise from the application of the quaternion system to the parameter charge. The quaternion system suggests that, ideally, the charges for the three interactions are equal in value; we could thus postulate the existence of a fundamental charge Q which is responsible for all of the interactions but which assumes the particular characteristics of either the weak, strong or electromagnetic charge depending on the quaternion operator with which it is associated. The problem, however, is that the imaginary quaternions (i, j, k) are not physically distinguishable and so cannot be used to label particular interactions or to indicate how two different charge systems would interact, whereas, by analogy with space, any two charge systems should produce a unique and identifiable resultant force.

The solution to the problem involves the creation of the system of "coloured" quarks. Quarks combine only in threes and are often assumed to possess fractional charges, but it is more likely that, as in the model of Han and Nambu (7), they possess either unit or zero charge. Quarks may be considered as particles which would, ideally, contain one unit, either positive or negative, of each of the electromagnetic, strong and weak charges (± ie ± js ± kw). However, in many cases, some of the individual units of e, s or w are missing. The "colour" concept and the combination of quarks only in threes is introduced to ensure that the assignment of quaternion operators to e, s and w in any individual quark is never known because it is always one of three possibilities, whereas we do know the charge structure of the total combination. The missing charges arise in some of the coloured quarks because it is necessary to prevent any of e, s or w being associated with more than one quaternion operator in such a combination. (It is possible, also, that the exigencies of this system actually limit the dimensions of charge to the minimum of three required by the quaternion representation.)

The sign of the charges on the quarks and the number of missing charges determines the nature of each of the heavy particles known as baryons and mesons, and it is not difficult to work out their charge structures from those of the quarks: baryons, which are combinations of three quarks, invariably have a + s and ±w component; mesons, which are combinations of quark and antiquark usually have neither; both sets of particles may or may not have a ± e component. The charge structures of the leptons may be discovered from the decays of these heavier particles: thus muons have ± e ± w, electrons ± e ∓ w, and neutrinos ± w. Since e, s and w are not interconvertible, it becomes obvious why baryon number and lepton number are conserved: a baryon can only decay into another particle with +s and ± w components, while a lepton can only arise from a particle with ± w component but no + s component.

The different physical manifestations of the strong, weak and electromagnetic forces are also among the effects of the quark system. A particularly important reason for this is their association with particles of different masses. The number of missing charges in any particle can be found easily from its quark structure, and it so happens that each missing charge in the particle is replaced by a mass of m_e/α where m_e is the mass of an electron and α is the fine-structure constant. (The size of the particle determines that the energy provided by this mass is equivalent to that of the missing charge.) This is true whether the charges are e, s or w and is a direct indication of their fundamental equality. The actual mass of the particle is thus a direct result of the broken symmetry in its charge structure. Ideally, the quarks which compose the particle should be unit charges not associated with definite masses. The breaking of this symmetry, which is an unavoidable consequence of the nature of the parameter charge, is thus invariably associated with the introduction of particles with definite mass.

(10) CPT Symmetry

The theory of particles presented in the above outline originates in the particular symmetry of space and charge within the group of order 4, but the details of the theory cannot be worked out without reference also to a more general symmetry. Thus, the CPT theorem states that the laws of physics must remain unchanged when space reflection and time inversion is followed by an exchange of particle and antiparticle (charge conjugation). (The reflection property of space is an obvious consequence of its description as a vector; time, on the other hand, is unidirectional because it is a continuum; however, because it is imaginary, it has two possible mathematical representations, positive and negative, and events should be symmetrical under an imagined reversal of time.) In other words, we know that a complete system involving all four parameters has no identifiable properties. Thus it has no identifiable sign preference; a complete sign reversal in the components (except mass, which is always positive) has no effect on the system. But space and time, or space and charge, alone, are not sufficient to describe a system and so space inversion does not necessarily hold at the same moment as time reversal or charge conjugation. The exception, in which the laws of physics do not remain unchanged under the partial symmetry of space inversion and charge conjugation, exists, for reasons which we shall discover in the theory of particle physics, in the weak interaction.

(11) The Quarks

The three nongravitational interactions fulfil many of the conditions which would be expected of the effects of a three-dimensional imaginary charge-like parameter. However, there are certain difficulties in identifying their sources directly with the imaginary components of the quaternion system. In principle, the identification of an individual interaction is specific in any given situation, but the identification of an individual quaternion is not. We always know whether an interaction is strong, weak or electromagnetic, but we do not know any way of distinguishing individually between i, j and k. Thus, we would perhaps expect all the units to be of equal value, but in fact we find very different values for each of the interactions; and we would expect interactions expressed in terms of indistinguishable quaternions to be also indistinguishable in type, whereas we find that each of the interactions has some pronounced properties.

Theoretical physicists are now working towards a model in which the three interactions are fundamentally equal in value and if we assume that the individual charges responsible for each interaction would be numerically identical if measured under ideal conditions we will automatically remove one of the main difficulties of the quaternion representation. However, it is still necessary to explain how the arbitrary operators i, j, k are related to the definite and known properties of electromagnetic, strong and weak interactions. Thus, if we can assume numerically identical fundamental units of electromagnetic (e), strong (s) and weak (w) charge, it does not follow that their quaternion representations are, say, specifically ie, js, kw, because we do not know that the quaternion representations in another system must be identical. The physical manifestations of e, s and w must represent to some extent the way in which the quaternion values of different systems are combined.

It may be possible to derive information about charge by analogy with space. Thus, even though spatial dimensions are intrinsically arbitrary, their resultant is not; in fact, it is always possible to combine two space systems in a unique way. Furthermore, this resultant itself has the form of a dimension and can be expressed as such. We would, therefore, expect to find a unique combination for any two charge systems; in this case, we may label the individual units which form the unique combinations as e, s and w. Thus, in any combination of two systems, the relative quaternion values assigned to e, s and w are established; only the individual quaternion values within a system are undefined. It is important to note, however, that e, s and w do not represent particular types of force as such; they are simply the resultant values of charge combinations, the parts that are always combined together. Their physical properties result from the arrangements necessary to ensure this.

Now, we have three well-defined systems of charge e, s and w, which have both positive and negative values, and any complete unit of charge would be a combination of one unit of each of these. There are eight such combinations and we may assign to each one the name of one of the u, d, s, c quarks or their respective antiquarks:

Charge Combination	Quark or Antiquark
+ e + s + w	u
− e + s + w	d
− e + s − w	s
+ e + s − w	c
− e − s − w	$\bar{u}$
+ e − s − w	$\bar{d}$
+ e − s + w	$\bar{s}$
− e − s + w	$\bar{c}$

These are the only types of unit charge combination which are physically possible (though this does not mean that other quarks are not possible.) Assuming that they are correctly assigned to their respective quarks, it is now necessary to assign quaternion values to them in such a way that no particular i, j, k is assigned to any of e, s, w but that it is nevertheless possible to combine the e, s, w of any two systems in a unique way, i.e. with one quaternion assigned to both e's, another quaternion assigned to both s's and the remaining quaternion to both w's.

Let us consider the assignment of quaternions to the u quark, +e +s +w. If we assumed that, at any one time, we had a choice of any of three such representations, then we would not be specifying a unique quaternion value for any particular charge. These are two obvious ways in which this could be done:

(a) (i) +ie +js +kw
(ii) +ie +ks +jw
(iii) +ke +is +jw
(b) (i) +ie +js +kw
(ii) +ke +is +jw
(iii) +je +ks +iw

The actual labels i, j, k are, of course, arbitrary and there are numerous other assignments which would be essentially identical to either (a) or (b). We could also distinguish between the individual representations (i) - (iii) by an arbitrary use of the colour labels blue, green and red.

These groups of "coloured" representations do not, of course, convey any more physical information than any single assignment of quaternions to e, s, w, such as (a) (i). However, if we were to consider them as units of combination, then we would find that, if we assigned a unit value to ie in (a) (i) and (a) (ii), then we could not also assign a unit value to ke in (a) (iii), because, in combination, all charges of the same type must be associated with the same quaternion value. In other words, if we could find a system in which the representation (a) (i) with unit ie was indistinguishable from the representation (a) (iii) with zero ke, then we could have a system in which the combination property of e's was unique but not related to the assignment of a specific quaternion value.

In fact, such a system does exist for the (a)-type representation, involving the u, d, s and c quarks. However, this depends on two conditions: the first is that the division between states and antistates in the various combinations of e, s, w must be made by assigning a positive unit of one type of charge to all states and a negative unit to all antistates; the second is that at least one type of charge must be allowed to change sign, when appropriate, by assuming a simultaneous sign reversal in other parameters such as space and time. The first property is assigned, arbitrarily, to s, and the second property to w; for the first time e, s and w are physically distinguishable.

In principle the assignment of such properties means only that we can separately identify the three components, e, s and w, in combinations. There is no physical meaning, as such, in classifying the combinations of e, s, w as quarks or antiquarks on the basis of the sign of the s component; it simply means that we can identify one of three interactions. There is not even any special meaning in deciding that the e and s components conserve their sign with respect to space and time, whereas the w component may or may not. Here we are simply keeping open an option which we know must exist. There is no general rule which states that any symmetry other than CPT must exist; any or all of the three interactions may break the C, P or T symmetry.

With these two conditions specified, the components of the system based on the (a) - type representation are given as follows:

		Blue	Green	Red
u	+e	1 i	1 i	0 k
	+s	0 j	1 k	0 i
	+w	0 k	0 j	1 j
d	-e	0 i	0 j	1 i
	+s	0 j	1 k	0 j
	+w	0 k	0 i	1 k
s	-e	0 i	0 j	1 i
	+s	0 j	1 k	0 k
	-w	0 k	0 i	1 j
c	+e	1 i	1 i	0 k
	+s	0 j	1 k	0 i
	-w	0 k	0 j	1 j

		Blue	Green	Red
u	+e	1 i	1 i	0 k
	+s	1 k	0 j	0 i
	+w	1 j	0 k	0 j
d	-e	0 i	0 j	1 i
	+s	1 k	0 k	0 j
	+w	1 j	0 i	0 k
s	-e	0 i	0 j	1 i
	+s	1 k	0 k	0 k
	-w	1 j	0 i	0 j
c	+e	1 i	1 i	0 k
	+s	1 k	0 j	0 i
	-w	1 j	0 k	0 j

The two tables give alternative representations. The different quaternion arrangements for each quark are specified by an arbitrary colour label, although, for convenience, the colours of individual quarks will be assumed to be those actually assigned in these tables. The actual quaternion labels are, of course, also arbitrary though the arrangements are not, except in the cases where the charges are all assigned a zero value; any other combination table found for the (a) - type representation must be physically identical to one of these. Whatever quaternion labels are assigned to e, s, w, their respective number of unit charges is based only on the properties which we have assumed to make them identifiable.

With the values of 1 or 0 as assigned to the unit charges in either table, it is possible to show that any combination of three quarks or three antiquarks, which contains one of each colour, has the same number of units of e, s, w, whichever colours are assigned to the individual quarks or antiquarks. With the antiquarks found simply by reversing all the signs in the tables, any combination of quark-antiquark of the same colour-anticolour must also have the same number of units of e, s, w, whatever the colour.

This system of coloured quarks and antiquarks thus presents us with a unique way of combining the charge-components of different systems without specifying their quaternion values. The terms quark and colour now take on their conventional meanings in particle theory. The quarks are almost identical to those of the Han-Nambu theory, with the s component directly equated to the baryon number. Here, however, there is an additional term in w, and in some cases this requires a sign reversal to preserve the colour invariance. We assume, in the first instance, that the symmetry broken by this sign reversal is that of space reflection or parity (P) and that the symmetry of the combination CP is conserved since there are reasons for believing the breaking of the time-reversal symmetry to be inherently less probable. The system, as defined, is exclusive to the u, d, s, c quarks and their respective antiquarks (unless heavier quarks exist with identical charge structures). This does not mean that other types of quark cannot exist in other systems. At least two more (b and t) are believed to be required by experimental evidence. Alternative possibilities include the creation of some kind of system in which more than one kind of charge breaks the C, P or T symmetry or quark systems based on variations of the main representation, for example:

		Blue	Green	Red
b?	$-e$	0 i	0 k	1 i
	$+s$	1 j	0 i	0 j
	$\pm w$	1 k	0 j	1 k

		Blue	Green	Red
t?	$+e$	1 i	1 i	0 k
	$+s$	1 j	0 j	0 i
	$\pm w$	0 k	1 k	0 j

(12) Mesons, Baryons and Leptons

Mesons and baryons may now be constructed from the u, d, s quarks and $\bar{u}$, $\bar{d}$, $\bar{s}$ antiquarks in the conventional way:

particle	quark combination	charge structure	typical decays
Meson octet (spin 0)			
π^+	$u\bar{d}$	$+e$	$\mu\nu$
π^-	$d\bar{u}$	$-e$	$\mu\nu$
π^0	$u\bar{u}$	0	$\gamma\gamma$; γe^+e^-
η	$d\bar{d}$	0	$\gamma\gamma$; $\pi^0\gamma\gamma$; 3π; $\pi^+\pi^-\gamma$
K^+	$u\bar{s}$	$+e+0$ or $+e+2w$	$\mu\nu$; $\pi^\pm\pi^0$; 3π;
K^-	$s\bar{u}$	$-e+0$ or $-e-2w$	$\mu^\pm\pi^0\nu$; $e^\pm\pi^0\nu$
K^0	$d\bar{s}$	0 or $+2w$	3π; $\pi^+\pi^-$; $\pi^0\pi^0$
$\bar{K}^0$	$s\bar{d}$	0 or $-2w$	$\pi^\pm e^\mp\nu$; $\pi^\pm\mu^\mp\nu$
Meson singlet			
η'	$s\bar{s}$	0	
Baryon octet (spin 1/2)			
n	udd	$+s+w$	$pe^-\nu$
p	uud	$+e+s+w$	——
Λ	uds	$+s\pm w$	$p\pi^-$; $n\pi^0$
Σ^-	dds	$-e+s\pm w$	$n\pi^-$
Σ^0	uds	$+s+w$	$\Lambda\gamma$
Σ^+	uus	$+e+s\pm w$	$p\pi^0$; $n\pi^+$
Ξ^-	dss	$-e+s\pm w$	$\Lambda\pi^-$
Ξ^0	uss	$+s\pm w$	$\Lambda\pi^0$
Baryon singlet			
Λ	uds	$+s\pm w$	
Baryon decuplet (spin 3/2)			
Δ^-	ddd	$-e+s+w$	
Δ^0	udd	$+s+w$	
Δ^+	uud	$+e+s+w$	$N\pi$
Δ^{++}	uuu	$+2e+s+w$	
Σ^{*-}	dds	$-e+s\pm w$	
Σ^{*0}	uds	$+s\pm w$	$\Lambda\pi$; $\Sigma\pi$
Σ^{*+}	uus	$+e+s\pm w$	
Ξ^{*-}	dss	$-e+s\pm w$	$\Xi\pi$
Ξ^{*0}	uss	$+s\pm w$	
Ω^-	sss	$-e+s-w$	$\Xi^0\pi^-$; $\Xi^-\pi^0$; ΛK^-

The charge structures of the individual particles explain many significant facts. Thus it is obvious that π^0, which has zero charge structure, must be its own antiparticle, whereas $\bar{K}^0$ or K^0, which may have a charge structure $\pm 2w$, is not. The $\pm$ sign of the w component of most particles suggest that some symmetry must be broken by the weak force if the colour invariance of the particles is to be retained. This is especially true of the K particles. The K^+ and K^- particles can be distinguished by their electrical charges, but the K^0 and $\bar{K}^0$ particles can only be distinguished by the average sign of their w components. To preserve colour invariance we must assume that the weak interaction cannot distinguish between different signs of w (which may be why the neutrino, with charge structure $\pm w$ only, has no rest mass and only one direction of spin) and so the neutral K particle is generally considered to oscillate between the K^0 and $\bar{K}^0$ states via the weak interaction. When the K^0 and $\bar{K}^0$ states are in phase (the K_1^0 decay mode), the charge structure is $+2w-2w$ or 0 and the particle decays to $\pi^+ + \pi^-$, which is space-reflection symmetric, without violating the combined CP or T invariance. But when K^0 and $\bar{K}^0$ are out of phase(the K_2^0 decay), the charge structure is nonzero and so the only remaining symmetry, CP or T invariance, must be violated in order to allow the particle to decay to $\pi^+ + \pi^-$ with zero charge structure. (It is probable, in fact, that the weak charge component of the K particles is always effectively zero, like that of the other mesons, and that this is maintained by the initial breaking of the time symmetry; the effective maintenance of zero weak charge structure without violation of parity is characteristic of $\pm 2w$ particles and is responsible for their classification as bosons.)

Also, it is clear why Ω^- is the only member of the baryon decuplet which can decay only by a weak interaction; for it is the only member of the series which always has a negative weak unit of charge. Strangeness, which is related to the number of s quarks in the particles, is also related to the average number of negative weak units of charge (or apparently negative weak units of charge) in the members of a multiplet. Thus the s quark has strangeness -1 (average $-w/3$), while Δ, Σ, Ξ, Ω have strangeness 0, -1, -2, -3 (average 0, $-w/3$, $-2w/3$, $-w$) respectively. Any decay of Ω^- will reduce this number and so violate the conservation of strangeness which is characteristic of the strong interaction (since the strong interaction involves only the s component of the charge structure). Thus, the decay of Ω^- to $\Xi^0 + \pi^-$ only balances the charge structures on each side of the equation if Ξ^0 takes the $-w$ value.

Again, the decay of Λ to $p + e^- + \bar{\nu}_e$ is slower than the same decay of the neutron, because Λ has charge structure $+s\pm w$ (strangeness -1) and only the Λ particle with structure $+s+w$, equivalent to a neutron with strangeness 0, may decay via this mode. Conventionally, this is attributed to the existence of two types of weak current; in fact, the strangeness-conserving weak current is the one which retains the sign of w, while the strangeness-charging weak current is the one which reverses the sign.

Strangeness is closely related to charm, which is equally an expression of the average number of weak units of charge in a particle. Strangeness and charm are distinguished by being associated with different signs of electromagnetic charge and are also given opposite signs by convention, but the s and c quarks are a natural pairing because they have the same values of w and because the weak decay of a charmed quark is predominantly via a strange quark mode. The c quark decays via the weak interaction to s but not $\bar{s}$ when the D^+ meson ($c\bar{d}$; charge structure $+e+0$ or $+e-2w$) decays to $K^- + \pi^+ + \pi^+$ ($+e+0$ or $+e-2w$) but not $K^+ + \pi^+ + \pi^-$ ($+e+0$ or $+e+2w$).

The particle decays can all be derived successfully from their charge structures (more successfully in the case of weak components than for any model involving fractional charges) and from these decays we can also determine the charge structures of leptons: muons have $\pm e \pm w$, electrons $\pm e \mp w$, and neutrinos $\pm w$.

The conservation of baryon number and lepton number are now obvious consequences of the separate conservation of e, s and w charges, but a detailed study of the neutrinos reveals why muon-lepton number and electron-lepton number are also individually conserved. The various neutrinos and antineutrinos may be specified as follows:

	helicity	sign of w	lepton number
ν_μ	negative	negative	+ 1
$\bar{\nu}_\mu$	positive	positive	- 1
ν_e	negative	positive	+ 1
$\bar{\nu}_e$	positive	negative	- 1

Hence $\bar{\nu}_e$ and ν_μ are not antiparticles of each other because they each have the same sign of w, while muons are not really heavy electrons and cannot decay to electrons by

$$\mu^- \rightarrow e^- + \gamma$$

because they have the opposite sign of w.

Neutrinos, unlike other leptons, are only distinguished from their antiparticles by the sign of the symmetry - breaking w component. The weak interaction cannot distinguish between + and - signs of w and so neutrinos and antineutrinos are only distinguishable from each other by their helicity or spin orientation with respect to the direction of motion (negative helicity being equivalent to left-handed spin). Weak interactions involving neutrinos are invariant under the CP inversion in which a neutrino of negative helicity becomes an antineutrino of positive helicity. Invariance under charge conjugation alone would imply the existence of neutrinos and antineutrinos of the same helicity, while parity invariance alone would have the same effect because a parity change would reverse the helicity without converting a neutrino to an antineutrino. Time-reversal, on the other hand, would preserve helicity while reversing the direction of the particle's spin and motion. The fixed helicity of neutrinos is thus equivalent to invariance under CP inversion or time reversal and its negative value for ν_μ and ν_e is presumably derived from the negative e component of both electrons and muons determining the initial values for their helicities. Assuming the convention of special relativity as applied to particles, it is possible to argue that neutrinos can only preserve their helicity and hence the CP invariance in all reference frames if their direction of motion is independent of the velocity of the observer, which means that they must be travelling at the speed of light and thus have zero rest mass.

Neutrinos are distinguished in their behaviour from the neutral K particles, which are likewise separated from their antiparticles only by their weak charge structure, because they are point-charges without internal structure, whereas K particles are composite mesons with definite masses; there is even a mass splitting of about 10^{-11} MeV between the K_1^0 and K_2^0 decay states, due to the weak charge component. The neutral K particles, therefore, do not have fixed helicities and their decays may conserve parity or space-reflection symmetry and violate the CP invariance.

The leptons are usually grouped with quarks into families which are particularly relevant in the unified gauge theories. These may be illustrated, with the relevant charge structures in the following table:

Leptons	Charge Structures	Quarks	Charge Structures
ν_e e^-	+w -e+w	u d	+e+s+w -e+s+w
ν_μ μ^-	-w -e-w	c s	+e+s-w -e+s-w
? ν_τ ? τ^-	± w -e ± w	t b	+e+s ± w -e+s ± w

The members of any one family all have the same sign of w, which explains such interconversions as u ⟶ d and c ⟶ s in weak interactions.

(13) The Three Nongravitational Interactions

The charge structures of mesons and baryons are also responsible for fundamental differences between the three nongravitational forces. Thus it is clear that the electromagnetic term e is different from the strong and weak terms, s and w, in being neither required to be present in a baryon state nor absent in a meson state. It appears as an independent term, attached to the particle only in a superficial way. In effect, the particles are grouped into states of the same strangeness, which are identical in every respect except for the electromagnetic component. These multiplets, such as Δ^-, Δ^0, Δ^+, Δ^{++} or Σ^{*-}, Σ^{*0}, Σ^{*+}, are defined to have a multiplicity, M, given by the number of component states. In any collection of such component states, the electromagnetic charge is not fixed to any one state but oscillates between all the available states.

There are two notable aspects of this property of electromagnetic charge. The first is that the electromagnetic force is not bound to any particle in the same way as the other two forces; this means that the electromagnetic interaction is, according to the standard theory, infinite in range, whereas the range of the other two interactions depends on the masses of the particles exchanged. Thus the electromagnetic interaction becomes the expected long-range resultant force produced by a combination of all three types of charge, and its presence also indicates the presence of strong and weak interactions. The resultant electromagnetic charge is directly associated with the position of the particle defined by its mass.

The second aspect is that the strong interaction has the property of isospin conservation or charge independence; that is it is unaffected by the presence or otherwise of the electromagnetic charge. This property is obviously not possessed by the electromagnetic charge and, since it is additionally dependent on the conservation of strangeness, it must also be violated to some extent by the weak interaction.

The strong and weak interactions are explained by a modification of the field theory used for the electromagnetic interaction. There are only two types of quark combination: baryons and mesons; but baryons may be combined with mesons to produce other baryons. Strong charges are always units confined to baryons. If they are to interact, then baryons must be capable of absorbing energy, in a form available to strong charges, which has been emitted by other baryons; this is, of course, analagous to the absorption and emission of virtual photons in the electromagnetic interaction. Since baryons may emit and absorb mesons, energy is available to be transferred in this form; however, the nonzero rest mass of the meson necessarily reduces the range of the interaction to that described by the Yukawa potential, while, according to quantum field theory, the spin 0 state of the meson makes the force between like particles apparently attractive as opposed to the repulsive force between like particles when the spin 1 photon is involved. Also, mesons themselves are affected by the strong force because they spend part of their time as virtual baryon-antibaryon combinations.

The strong interaction, however, is generally closely associated with colour; in colour-anticolour combinations, such as mesons, the s component disappears; the absence of colour here involves the absence of the s component. It is possible, therefore, to associate the absorption or emission of virtual mesons with an interchanging of colour in the component quarks of a baryon and to postulate that, because the baryon structure of three confined quarks is defined to be colour-invariant, the property of colour is actually responsible for the confinement of quarks, within baryons. According to the theory of quantum chromodynamics, the strong force between baryons thus becomes a result of a colour force between quarks which is repulsive for identical particles in the same way as the electromagnetic force because it is brought about by the exchange of the massless spin 1 bosons called gluons. Quarks may be considered as permanently bound by the nature of the colour system rather than by any force, though some sort of colour force may actually exist; the forces between quarks are sometimes thought to be constant with distance and experiments show quarks to be *loosely* bound inside hadrons. Quark confinement may not need a "physical" explanation.

The weak interaction is, like the electromagnetic interaction, independent of colour, but it is not similarly independent of the particle of origin. It is associated not only with mesons and baryons, but also with leptons, which come into existence as decay products of the heavier particles, while its range is such that it is assumed to be carried by a very heavy spin 1 boson (W or Z). This boson must be a virtual combination of heavier particles and antiparticles to confine the weak charge in the same way as the exchange of the meson confines the strong charge. (It is interesting to note that such processes as the weak decay of π^+ involve the initial creation of a virtual nucleon-antinucleon pair, via the strong interaction, before the weak transition to $\mu^+ + \nu_\mu$.)

As we have seen, the violation of individual C, P or T symmetry is characteristic of the weak interaction and is introduced to preserve the colour-invariance of the quark-system. Most weak interactions involve the neutrino and, hence, are not space- reflection symmetric, because the neutrino spin is invariably left- handed with respect to the direction of motion. However, the decay of K_2^0 to $\pi^+ + \pi^-$ does not involve the neutrino and, because the product $\pi^+ + \pi^-$ *is* space-reflection symmetric, the interaction is apparently unique in violating the time-symmetry instead. This type of "superweak" interaction seems to occur only where a ±2w component is involved and the conservation of parity or space-reflection symmetry is presumably related to the fact that K and all other particles with 0 or ±2w are bosons with symmetric wave functions, whereas particles with ± w are fermions with antisymmetric wave functions.

(14) The Structure of Leptons

We have established the charge structures of leptons by examining the typical decays of mesons and baryons, but, while we have explained all the heavier particles in terms of the coloured quark system and the definition of unit charge, we have yet to find any fundamental reason for the existence of this apparently independent system of lighter particles. There is no reason to believe that there are any particles which cannot be derived from combinations of quarks and antiquarks, but the two obvious types of combination which preserve colour invariance, the mesons and baryons, produce no charge structures which are identical with those of any individual leptons. However, the charge structures of the four flavours (u, d, s, c) of red quark in the first quark table are identical to the charge structures of the four leptons ν_e, e^-, μ^- and ν_μ ; this is certainly significant.

To find an explanation for this coincidence, we must undertake a more detailed examination of the weak interaction, the force in which all leptons are involved. The weak interaction has always been found to involve fermions - particles with a weak charge component of ±w (not ±2w which is equivalent to 0) - and every weak interaction so far known can be expressed as an interaction between *four* fermions. Weak interactions also result in changes of flavour wherever quarks are involved, in the same way as strong interactions result in changes in quark colour. Flavour changes between quarks, like colour changes occur in pairs. Thus, in the decay

$$\Lambda \longrightarrow \pi^- + p$$

which we may write as

$$\underset{\mathrm{uds}}{\Lambda} + \underset{\mathrm{uud}}{\mathrm{p}} \qquad \underset{\mathrm{udu}}{\mathrm{p}} + \underset{\mathrm{udd}}{\mathrm{n}}$$

by replacing π^- with a virtual $\bar{\mathrm{p}}$ + n combination and transferring $\bar{\mathrm{p}}$ to the left hand side of the equation, an s quark in Λ changes to a u quark in p at the same time as a u quark in p changes to a d quark in n. The pairing ensures the conservation of e and w charges; in a single u$\longleftrightarrow$d or s$\longleftrightarrow$c transition, an e charge appears or disappears, whereas a u/d$\longleftrightarrow$s/c transition results in a change of sign of w; a double transition leaves the total e and w components unchanged. Thus weak interactions between baryons always involve four particles.

Leptons are also paired off in weak interactions, muons and electrons being always accompanied by their appropriate neutrinos. Let us, therefore, suppose that weak interactions involving leptons also take place through exchange in quark flavour. We may guess that the quarks associated with e^- and $\bar{\nu}_e$ are, respectively, d and $\bar{\mathrm{u}}$, which have the same charge structures; thus, in an interaction such as neutron decay, which we may write as

$$\underset{\mathrm{udd}}{\mathrm{n}} + \underset{\bar{\mathrm{u}}\bar{\mathrm{u}}\bar{\mathrm{d}}}{\bar{\mathrm{p}}} \qquad \underset{\mathrm{d??}}{e^-} + \underset{\mathrm{??}\bar{\mathrm{u}}}{\bar{\nu}_e}$$

we have the u of n converted into d and the $\bar{\mathrm{d}}$ of $\bar{\mathrm{p}}$ converted into $\bar{\mathrm{u}}$. Now, the charge structures of e^- and $\bar{\nu}_e$ would not be altered if we identified the unspecified quarks as d$\bar{\mathrm{d}}$ and u$\bar{\mathrm{u}}$, and it may be that the complete equation for neutron decay should be written

$$\underset{\mathrm{udd}}{\mathrm{n}} + \underset{\bar{\mathrm{u}}\bar{\mathrm{u}}\bar{\mathrm{d}}}{\bar{\mathrm{p}}} \qquad \underset{\mathrm{dd}\bar{\mathrm{d}}}{e^-} + \underset{\mathrm{uu}\bar{\mathrm{u}}}{\bar{\nu}_e}$$

This would mean supposing that, in a weak interaction, we could reverse the signs of charge in one quark from each particle to produce a pair of forbidden quark-antiquark combinations; leptons, if so structured, would be virtually quarks or antiquarks which could be regarded as dimensionless point charges with no internal structure. They would, of course, have to retain the $\pm$w component, since only the weak interaction would allow this symmetry-breaking charge conjugation, and they would presumably have zero s component to avoid the complications associated with colour, which would otherwise prevent the formation of such combinations - leptons could be considered "colourless" if colour changes were impossible.

Assuming this proposed structure of leptons to be correct, the reason for the families of associated quarks and leptons, introduced into the unified gauge theories, becomes clear, for the muon and muon-neutrino are virtually equivalent to s and c quarks in the same way as e^- and ν_e are equivalent to d and u. Thus the decay of π^+ via the virtual p + $\bar{\mathrm{n}}$ pair may be represented as follows:

$$\pi^+ \longrightarrow \underset{\mathrm{uud}}{\mathrm{p}} + \underset{\bar{\mathrm{u}}\bar{\mathrm{d}}\bar{\mathrm{d}}}{\bar{\mathrm{n}}} \longrightarrow \underset{\mathrm{u}\bar{\mathrm{u}}\mathrm{c}}{\mu^+} + \underset{\bar{\mathrm{s}}\bar{\mathrm{d}}\mathrm{d}}{\nu_\mu}$$

The charges on each lepton are thus determined by those on the associated quark of the family; the quark and antiquark which make up the remainder of the particle are a meson-type combination of zero charge, whose components depend on those of the particles actually involved in the interaction.

The weak interaction is generally supposed to take place through the exchange of a W^+ or W^- boson, which causes the change of quark flavour (though there is also a form of weak scattering via a neutral Z boson which is equivalent to $W^+ + W^-$). Thus, in the interaction

$$n + \nu_e \longrightarrow p + e^-$$

the neutron decays to a proton with the emission of a W^- particle which is then absorbed by the neutrino to convert it into an electron. It is expected that decays such as

$$W^- \longrightarrow e^- + \bar{\nu}_e$$

and

$$W^+ \longrightarrow \mu^+ + \nu_\mu$$

will take place. The $W^\pm$ boson is thus equivalent to a combination of $\mu^+ + \nu_\mu$ or $e^- + \bar{\nu}_e$ and, therefore, to a combination of three quarks and three antiquarks. The many alternative combinations for each W particle, provided by the various quark-antiquark meson-type combinations of zero charge which can be permutated with the fixed quark associated with each lepton, may explain the heavy mass associated with the W particle of not less than 39.8 times that of the proton.

(15) Baryon Masses

In general, we can express the rest masses of fundamental particles as close approximations to multiples or half-multiples of the term m_e/α, where m_e is the mass of the electron and α is the fine structure constant, $e^2/4\pi\varepsilon_0 \hbar c$. Particularly significant is the fact that the mass of the pion, the particle generally associated with the strong interaction, is almost exactly equal to $2m_e/\alpha$, and thus the Compton wavelength for the pion $(\hbar/m_\pi c)$ is the same as half the classical radius of the electron $(e^2/8\pi\varepsilon_0 . m_e c^2)$ or the distance which would separate an electron-positron pair whose electromagnetic field energy was equal to its original total rest energy $(2m_e c^2)$.

Now, the electron-position interaction may be said to represent the unconfined electromagnetic force because it is independent of the strong charge but the electron itself has a finite classical radius because the electromagnetic charge is confined to the particles determined by the requirements of the quark system. Although electrons are considered as point charges, they are created as decay products of baryons and mesons and may have a baryon-like structure of three quark/antiquark components; they are, in particular, the only ultimate decay products of the pion itself, other than massless neutrinos with only weak charges. The distance required to "create" an electromagnetic charge must therefore be of the same order as that for a strong charge. The latter is, for theoretical reasons, restricted to the interiors of hadrons and thus determined by the range of the interaction or the Compton wavelength of the exchange meson. Thus if $2m_e$ is the mass equivalent of the electromagnetic energy of the electron-position pair, then $2m_e/\alpha$ must be of the order expected for the mass equivalent of a pair of strong charges. If all charges are really equal, this energy may also be a limiting value for fundamental charges; the energy of a pair of strong charges (2hc/r) may be the minimum required to confine two wave functions at a wavelength equal to the distance of separation (r). The actual mass of the pion thus suggests that the rest masses of fundamental particles are, in some sense, a direct consequence of their charge structures.

We can see how this determines the masses of the baryon octet and decuplet. Here, the rest mass is a term which compensates for the missing units of charge in the members of a multiplet. We may express this by the formula

$$\text{mass of particle} = \frac{n_o M_o}{M} \frac{me}{\alpha}$$

where n_o is the total number of zero charges in the components of a multiplet of multiplicity K, and M_o is the highest multiplicity in the octet or decuplet. Thus, in the baryon decuplet, we have the multiplet Σ^*, with M=3; M_o, which is the multiplicity of the Δ particles, is 4; and n_o represents the total number of zeros, derived from the quark tables, in the combinations dds, uds and uus. This is either 15, 17 or 19, and so, for the ground state,

$$\text{Mass of } \Sigma^* \text{ particle} = \frac{15 \times 4}{3} \frac{m_e}{\alpha} = 20 \frac{m_e}{\alpha}$$

which is in reasonably close agreement with the measured value, 1385 MeV.

The formula arises from the charge-independence of the strong interaction. In effect, the Δ particle represents four simultaneous states of average mass $(n_o/4)(m_e/\alpha)$ and the Σ * particle represents the excited states of all of these with average mass $(n_o/3)(m_e/\alpha)$; similarly, the Ξ*particle represents the excited states of Σ *, and the Ω particle the excited states of Ξ *.

The derivation of the individual baryon masses can now be set out as follows:

	particle	quark structure	n_o	M_o	M	predicted mass (m_e/α)	measured mass (m_e/α)
Octet	N(n,p)	udd, uud	$\underline{9}$, 11,13	3	2	13.5	13.4
	Λ	uds	$\underline{5}$, 7	3	1	15	15.9
	Σ	dds, uds, uus	15,$\underline{17}$,19	3	3	17	17
	Ξ	dss, sss	11,$\underline{13}$	3	2	19.5	19
Decuplet	Δ	ddd, udd, uud, uuu	$\underline{20}$,22,24	4	4	20	18
	Σ *	dds, uds, uus	$\underline{15}$,17,19	4	3	20	20
	Ξ *	dss, uss	$\underline{11}$,13	4	2	22	22
	Ω	sss	$\underline{6}$	4	1	24	24

The indicated values of n_o are those used in deriving the predicted masses and, except in the case of Σ and Ξ, are those of the ground state. (It is perhaps significant that the ground state values here are already occupied by Σ * and Ξ *.) The higher values may contribute towards the masses of some of the observed baryon resonances such as Δ (1690) and Ξ * (1820).

The mass formula gives good agreement for all the members of the decuplet except the Δ particles. However, this apparent discrepancy is easily explained, for it is possible to derive the constant mass difference between the successive multiplets in the decuplet by a different argument, which accounts for the observed mass of the Δ particle. Thus, it is usual to see the baryon octet and decuplet as simply representing excited states of a single type of particle, and all the particles in a multiplet as being virtually interchangeable. The decuplet, for instance, shows the excitation from Δ to Σ * to Ξ * to Ω as occurring due to the transition of one d quark to an s quark at each level. Using the particular convention assumed in the quark tables above, this requires a transition in the w component of one red quark from +1 to -1. This would require a net loss of 2 charges, compensated for by a mass increase of $2m_e/\alpha$, which is the mass difference between particles of different multiplets given by the familiar Gell-Mann-Okubo formula. The extra mass of $2m_e/\alpha$, predicted for the Δ particle on the basis of zero charge components, does not, however, disappear. The apparently competing requirements for the particle's mass are satisfied simultaneously by giving the Δ particle a kinetic energy equivalent to the missing mass, which explains its rapid decay. The full width of the Δ state at half maximum of energy is ~120 MeV, which is of order $2m_e/\alpha$ and considerably greater than that for any other member of the decuplet.

The position is somewhat more complicated with the components of the baryon octet, partly because the mass of the Λ particle may be affected by the existence of singlet Λ states, and partly because we have two terms, Λ and Σ^o, with the same quark structure, though presumably with different masses. If we had a particle of mass $15.5\ m_e/\alpha$, we would have the expected mass difference of $2m_e/\alpha$ between N and Λ, and between Σ and Ξ. The additional mass splitting of ~$1.5\ m_e/\alpha$ between Λ and Σ could be accounted for by an investigation of the colours of the individual quarks necessary to produce the observed masses. Thus to maintain N and Λ at the ground state it is necessary that they should not contain a red u quark - that is one with a zero e component. However to produce Σ and Ξ particles of the observed masses, it is necessary to specify that the uds term or the uss term must contain a red quark. In other words, Σ is distinguished from Λ by the colour of its u quark, and to convert N into Σ requires, not only the conversion of a d quark to an s quark, with mass increase of $2m_e/\alpha$, but also the conversion of a blue or green u quark to a red u quark, with loss of 1x3x $\frac{1}{2}$ charges in N and a further mass increase of 1.5 m_e/α. (No such colour changes are required in the baryon decuplet.)

The observed masses of the various baryons are thus in approximate agreement with the predicted masses. It is probable that more accurate values may be obtained by defining further conditions which must be fulfilled in the constitution of particles. Thus, the Gell-Mann-Okubo mass formula, based on SU (3) symmetry, predicts the relation

$$\frac{1}{2}(m_N + m_\Xi) = \frac{3}{4} m_\Lambda + \frac{1}{4} m_\Sigma$$

between the masses of the baryon octet. This is compatible with the above theory if $m_\Lambda - m_N$ is taken to be the 2.5 m_e/α required by the experimental values rather than $2m_e/\alpha$. This may be possible if Λ is assumed to be mixed with a singlet state, and it may be that it is also possible to determine the degree of mixing from the requirements of SU(3) symmetry.

Masses or apparent masses for the three lightest quarks (u, d, s) may be derived, in the usual way, from the masses of the baryons; the relative heaviness of the strange quark is responsible for the apparent symmetry-breaking effects of the so-called "semi-strong" force. The masses of heavier quarks seem to require the unified theories of electromagnetic, strong and weak interactions for full explanation.

(16) Meson Masses and Meson Decay

The meson octet does not represent the regular progression of excited states from the lowest member which we observe in the baryon octet and decuplet. Each multiplet must be considered to be virtually independent. Furthermore, the octet is definitely part of a nonet, with the η' signlet invariably mixing with the η state; consequently the latter must be considered as one component of a doublet.

Since we consider the meson multiplets to be independent of each other and of the octet, we cannot use the mass formula which we defined for baryons. Instead we assume that

$$\text{mass of meson} = n_o \frac{m_e}{\alpha}$$

where n_o is the total number of zero charges in the components of the multiplet. The possible values of n_o for the various meson multiplets are given below:

π^-, π^o, π^+	2, 6, 8, 10, 12, 14, 16
K^o, K^+ $K^-, \bar{K}^o$	3, 5, 7, 9, 11
η, η'	4, 6, 8, 10, 12

If we choose n_o for the ground state of π, then it may be that the values of $n_o = 7$ for K and $n_o = 8$ for η are determined by the condition, derived from the Gell-Mann-Okubo mass formula, that

$$m_K^2 = \frac{1}{4} m_\pi^2 + \frac{3}{4} m_\eta^2 .$$

These values of n_o give a good correlation with the observed masses of the members of the meson octet.

Meson decay is unique to the individual particle and so removes the degeneracy represented by the multiplets. Thus the total mass of the decay products of a meson of multiplicity M is not greater than (n_o/M) (m_e/α). However, as long as the total mass of the decay products is not greater than that of the original meson, the n_o in the formula may be any value which can be accommodated within the charge structure of the meson. Thus, the K particle, which has five possible values of n_o/M, has also five possible decays which yield products of different total mass. These are, respectively:

decay	n_o/M	mass of decay products (m_e/α)
$\mu\nu$	1.5	1.5
$\pi e\nu$	2.5	2
$\pi\mu\nu$	3.5	3.5
2π	4.5	4
3π	5.5	6

It is possible that there is a correlation between the mass of the decay products and n_o/M, and that the apparent discrepancies of order 0.5 m_e/α are the result of some change of quark colour, or sign of w, which is necessary to produce the higher values of n_o. With no such change required in the $\mu\nu$ and $\pi\mu\nu$ decay modes, the mass of $\mu\nu$, and hence of μ (which originates only as a decay product of K or π), would be fixed at 1.5 m_e/α. This large mass is presumably responsible for the particle being unstable against decay to the lighter electron.

The π and η mesons do not possess so many decay modes, but it is possible to interpret their typical decays according to the same pattern:

particle	decay	n_o/M	mass of decay products (m_e/α)
π	$\mu\nu$	2	1.5
η	3π	6	6

(17) Resonances

If there is sufficient evidence that the rest masses of ground-state mesons and baryons can be attributed mainly to the missing charges in their quark structures, the exact masses are a more complicated problem in which a number of possibly conflicting conditions are resolved by the application of the quantum mechanical uncertainty principle relating energy to lifetime. Another complication arises from the fact that particles are not really explicitly defined objects, few particles being inherently stable and every particle spending at least part of its lifetime as a virtual combination of other particles. This instability is, indeed, to be expected from our definition of particles as the direct result of the attempt to impose the unit of charge on nature; we know the unit of charge only through various particle-representations each of which must be assumed to have its own probability of occurring and lifetime, but each is really a representation not an externally existing object; none gives the exact description of the unit of charge which we must assume actually exists, and any representation in terms of mass or energy must be further limited by the terms of the uncertainty principle.

The essentially arbitrary and transient nature of the fundamental particles is exemplified by the multitude of short-lived resonance states so far discovered. There is no qualitative difference between the resonances and the regular baryons and mesons; the only distinguishing feature in most cases is their relative lifetime before decay. The existence of so many possible particle-states suggests that the concept of "particle" is not really fundamental and that it does indeed exist only in so far as it is a means of defining a unit charge. At the same time, it is certainly possible to find definite conditions for the formations of baryons and mesons with particular masses and it is likely that these conditions extend also to the resonances.

Since baryons and mesons are assumed to interact via the strong interaction in which virtual mesons are successively emitted and absorbed, and since many of the excited states of these particles decay principally to less excited states of the same particle with the emission of a π meson, we may assume that at least some of the resonances may be derived by combining ground-state baryons and mesons with π mesons of various masses. The main series of meson resonances are the octets with π mesons of mass 765, 962, 1070, 1235 and 1320 MeV respectively. These show an approximate correlation with the particles which would result from a combination of one π meson of mass 6, 8, 10, 12 or 14 m_e/α with the π, K and η mesons of respective mass 6, 7 and 8 m_e/α. Similarly, the octets of baryon resonances of spin 1/2, 3/2, 5/2 and 7/2 may be derived approximately from combinations of the ground-state baryons with π mesons of respective mass 6, 8, 10 and 12 m_e/α. These correlations may suggest the structure of the respective meson and baryon resonances but others are possible; since there are many available meson masses and there is also the possibility of mixing between states with identical quantum numbers, the suggestion is at present no more than a speculation.

(18) Quantum Mechanics

We have seen that charges do not exist as isolated units, but are found only in particles associated with definite masses; forces measured between these particles are not necessarily those which would exist between isolated unit charges. The quaternion representation suggests that, ideally, the strong, weak and electromagnetic forces would all be of equal magnitude, all of inverse square law form, and all transmitted by massless particles at the speed of light. Quantization of energy would follow logically from this situation. Thus, the energy of interaction of two charges of strength Q, separated by a distance r, is given by Q^2/r. This is also the energy due to the repulsion of the self-charge for a sphere of radius r uniformly charged over its surface. This energy travels at velocity c, so the time period of the interaction is r/c and the frequency c/r. Thus, energy is related to frequency by a constant Q^2/c which actually becomes equal to Planck's constant if Q^2 is taken close to the value for the strong interaction (s^2/r is of order $2hc/r$ which may be the expected energy for the wave functions of two particles confined within the distance r).

The fundamental equivalence of the three forces is now a generally accepted principle. The unification of the weak and electromagnetic interactions has already been accomplished and it is known that their apparent differences are due to a process called spontaneous symmetry breaking which results from the particles involved in the interaction assuming various masses. Theoretical physicists now believe that the same process may be responsible for the differences between these forces and the strong interaction. The strong force only appears to be stronger because it breaks fewer symmetries. Apparently, the strength of each of the forces is not constant but depends on the energy with which particles interact. As this energy increases, the weak force grows stronger and so does the electromagnetic force, though at a lesser rate, while the strong force grows weaker. At a sufficiently high energy (of order 10^{15} GeV) all the forces would have the same strength.

It seems, therefore, that the apparent difference in forces is due to particles, which are otherwise composed of units of the three charges, possessing definite masses. Quantum mechanics originates in this situation. Particles have an energy relating to their units of charge which is quantized; certain arrangements of the units are necessary to fulfil the requirements of the quaternion representation; however, to maintain symmetry between different particles it is necessary to assume masses which compensate for the missing charges and for the energy and momentum of these masses to be quantized in the same way as if the charges were actually present, with $E=\hbar\omega$ and $p=\hbar k$. Since the mass of a particle is the quantity associated with the definition of its position, we would expect a particle of this kind to have the properties associated with a wave packet, a free particle being associated with a travelling wave whose simplest form of x, t dependence is given by the wave function

$$\psi(x,t) = \exp i\,(kx-\omega t).$$

The equations of quantum mechanics are derived by assuming only that this information may be applied to the classical relation $E=p^2/2m$ and there is no obvious reason to doubt the validity of the Schrödinger formulation.

However, quantum mechanics and the related statistical concepts best represented by Heisenberg's uncertainty principle are simply the result of the classical effect of retardation in the nongravitational interactions between unit charges. The Schrödinger equation and the extended relativistic quantum mechanics of Dirac are of phenomenological origin, being based on the experimental discovery of electrons and protons as real particles. The full quantum theory of fields, in which such "particles" are replaced by quanta of excitation of a matter field, is really a more natural development. The rest mass of a fundamental particle originates as field energy because it represents the energy of missing charges; field energy is wavelike and the wavelike properties of particles result in the so-called "first quantization" of matter; however, the retardation effect produced by the finite velocity of light also necessitates an actual quantization of the field energy or a "second" quantization of matter.

Quantum field theory has been most extensively applied in quantum electrodynamics or the theory of the interactions of electrons and photons. This theory assumes that electrons are point charges without physical extension and introduces the renormalization of mass and charge to prevent the self-energy of the electron, as calculated from perturbation theory, becoming infinite due to emission and reabsorption of virtual photons; usually this means the assumption of an infinitely negative "bare mass" which is subtracted from the infinite positive mass-energy of the virtual photons to give the finite observed mass of the electron. However, it is also possible to derive a finite value for the observed mass assuming a zero bare mass and using a nonperturbative calculation of the self-energy in Landau gauge. (8) The infinities or divergencies which have to be removed by renormalization are, thus, introduced along with perturbation theory rather than with the initial assumption of point charges. If leptons are, indeed, virtually quark-like, then we would expect them to behave like unit point charges with zero bare mass.

(19) Fermions and Bosons

The Dirac theory of the fermion introduces the 4-vector space-time symmetry into the Schrödinger formulation of quantum mechanics. Fermions take part in weak interactions and have a total weak charge component of ±w. Now, the weak charge is allowed effectively to change sign to preserve the quark colour invariance, and this is achieved, for fermions, by a violation of parity conservation or symmetry under a change of sign of space coordinates, the CP invariance or time-reversal symmetry being preserved. (The parity of individual fermions is not observable but is defined to be positive.) This means that, in order to introduce the imaginary space-time symmetry into the quantum mechanics of the fermion, it is more convenient to introduce space with both positive and negative solutions, rather than time, with its unique solution, as the imaginary component of the relationship. Space thus becomes a time-like parameter for fermions. In the Dirac theory, there is a new momentum operator $p_o = i\hbar \partial / \partial x_o$, where $x_o = ct$, which is introduced by symmetry with $p = i\hbar \partial/\partial x$, the momentum operator for the space coordinates. Thus, if the free particle equation

$$\{p_o - (m^2c^2+p_1^2 + p_2^2 + p_3^2)^{\frac{1}{2}}\} \Psi = 0$$

is multiplied by the operator $\{p_o + (m^2c^2+p_1^2+p_2^2+p_3^2)^{\frac{1}{2}}\}$ we may derive the symmetrical equation

$$(p_o - m^2c^2 - p_1^2 - p_2^2 - p_3^2) \Psi = 0$$

which introduces a complete range of negative energy values corresponding to those of positive energy. To find an equation linear in p_o, we set down

$$(p_o - a_1 p_1 - a_2 p_2 - a_3 p_3 - b) \Psi = 0$$

which is the same as the previous equation when multiplied by the operator $(p_o + a_1p_1 + a_2p_2 + a_3p_3 + b)$ provided that

$$a_1^2 = 1 \qquad b^2 = m^2c^2$$

$$a_1a_2 + a_2a_1 = 0 \qquad a_1b+b\,a_1 = 0 \qquad \text{etc.}$$

The new degrees of freedom represented by the introduction of the terms a and b describe the particle's spin and result in the half-integral spin characteristic of all fermions. It is significant that spin is derived by adapting the space part of the wave equation to be symmetrical with the time component. This is no more than a mathematical device for it makes mass-energy look like an imaginary quantity (with both positive and negative solutions) and charge a real quantity (with its unique solution) in the mass-charge symmetry which automatically follows from that of space and time. (Imaginary space is not "natural" and requires the mathematical complications produced by the introduction of spinors.) The true situation is restored in the Dirac theory by the apparently ad hoc posutlate of the existence of antimatter; this makes the apparent negative mass-energy states become actual positive mass-energy states with the opposite sign of charge.

Now, the total wave function of a fermion is the product of the orbital and spin wave functions; the symmetry of these depends only on that of space reflection and time reversal, charge conjugation being excluded automatically in the Dirac formulation. (Charge conjugation is an ad hoc feature of the Dirac theory, not appearing in the equations.) The wave functions of fermions are time-reversal symmetric though not, as we have seen, symmetric to a reversal in space coordinates. The total wave functions are, therefore, antisymmetric, and it is possible to link this with the half-integral spin because both are the result of space coordinates being associated with the symmetry breaking of the weak charge. Particles with antisymmetric wave functions can be shown to obey the Pauli exclusion principle that no two particles can be in the same quantum energy state. In fact, the original Dirac theory invoked this principle to avoid the negative energy states predicted by the relativistic wave equation, by supposing that all the negative energy levels were already occupied and that fermions with positive energy were thus prevented from making downward transitions from the ground state. In quantum field theory, Dirac's relativistic theory of the fermion is reinterpreted as a theory of interacting fermion-antifermion and photon fields; assuming CPT symmetry, fermions and antifermions emerge simultaneously with positive energy only and there is no need to posutlate an unobservable sea of negative energy states.

Bosons have no single $\pm w$ component of weak charge and so are not antisymmetric to a reversal in the sign of spatial coordinates. Particles with $\pm 2w$, like the K mesons, behave like those with zero weak charge; there is a mass splitting between K_1^o and K_2^o due to the 2w weak component, though there is no mass splitting for particles with $\pm w$; the decay of K_2^o consequently breaks the time-reversal symmetry rather than that of space reflection. The status of space and time is not, therefore, reversed when the imaginary space-time symmetry is introduced into the otherwise asymmetric wave equation for bosons. The modification in the resulting Klein-Gordon equation is made in this instance to the time part of the wave equation and describes particles with integral spin. Space remains the real and time the imaginary parameter, mass-energy remains positive and the antiparticles are simply a result of CPT symmetry. The ± sign of charge is effectively introduced into the Dirac equation for fermions with the prediction of negative energy states, but it does not occur in any form in the equation for bosons; this equation describes particles with a single sign of charge. Time becomes space-like and unidirectional (by symmetry with positive mass) and the time-reversal symmetry is thus also preserved for particles in a single charge state. Bosons, therefore, have symmetric wave functions and are not subject to the Pauli exclusion principle.

(20) Thermodynamics

The exact definitions of the physical parameters, as specified by their symmetry group, are relevant to all branches of physics. Thermodynamics provides an example of how, through quantum mechanics, they can be used to predict general conceptual laws even for the behaviour of such complex large-scale physical systems as solids, liquids and gases whose exact properties are unknown.

The laws of thermodynamics introduce a phenomenological term heat, defined in terms of conservation of energy by the first law, which refers to the transfer to or from a system of a quantity of kinetic energy which includes a component of purely random motion. Classically, such random motion is of unknown origin or merely expresses our lack of knowledge of the many variables which determine the behaviour of the system. However, it is possible to advance a qualitative explanation of the experimentally established second and third laws of thermodynamics on the assumption that the random motion of the particles of all matter is essentially of quantum mechanical origin and is ultimately a consequence of the localisation property of charge.

In any system involving ultimate particles acted on by nongravitational forces we have energy states which are expressions of the uncertainty of locating masses in space and time; energy available to the system is necessarily always partly absorbed into these translation, vibrational and rotational modes of random motion, which derive from the statistical or uncertain element in quantum mechanics. In other words, some of the energy of the particles which make up solids, liquids and gases is always occupied with the fundamental uncertainty of the spatial and temporal coordinates of the masses. Such notions have no place in classical mechanics: from this point of view, since classical notions of energy and work are based on the actual positioning of masses in space and time, this means that energy in such a form has become partially unavailable to do work (i.e. move a force a certain distance). Since some such uncertainty is fundamental to all quantum mechanical situations, then it is clear that, in any interaction between energy and a quantum mechanical system, a part of the energy which was originally available to do work must become unavailable. Thus, when energy is supplied to a quantum mechanical system it must be partially converted into that required to determine the coordinates of the system, that is it must be partially absorbed into random modes of motion. Assuming that information which has been lost cannot be recovered, it follows that any energy which becomes unavailable must remain unavailable, even though it may be transferred from one system to another.

Now, if heat is exchanged between two systems, then an energy exchange must take place at the quantum mechanical level, because the heat energy always includes kinetic energy due to random motion; some fundamental quantum mechanical change must have taken place in both systems and some energy, formerly available to do macroscopic work, must have been transferred to quantum mechanical uncertainty states. Thus, if heat is involved in a physical change, then it is obvious that there will be an increase in the total amount of unavailable energy as a result of the change, whereas if the change is accomplished so that no heat enters or leaves a particular system, then it is equally obvious that there will be no increase or decrease in the amount of unavailable energy within the system. These conditions constitute the second law of thermodynamics. It is expressed in mathematical terms by the definition of a new quantity, entropy, which can be shown to be a direct measure of the thermal disorder in a system and which always increases for any physical process in which a quantity of energy becomes unavailable for work.

One further result is obtained by defining a phenomenological quantity, temperature, which is a direct expression of the kinetic energy of a system. At absolute zero of temperature, when the kinetic energy of the system disappears, we would also have zero entropy. Since such a situation is quantum mechanically impossible, it is therefore impossible to reduce any large-scale physical system to a temperature of absolute zero. This is the third law of thermodynamics.

(21) The Direction of Time

Time is, like mass, a one-dimensional continuum and is, therefore, unidirectional, though, unlike mass, it is an imaginary parameter and has the symmetrical properties associated with the existence of its mathematically indistinguishable positive and negative values. This unidirectionality ensures that time symmetry cannot be equated with reversibility in dynamical processes. In fact, the properties of a dynamical system are, on the whole, irreversible because they depend on both the laws of motion and the initial conditions of measurement. The former are insensitive to a reversal in the sign of the time coordinate, because they are expressed in terms of acceleration, and hence are reversible; while the latter, which are expressed in terms of velocity, are sensitive to a reversal in the sign of time and are irreversible. The second law of thermodynamics allows us to identify the direction of time as determined by a sequence of irreversible dynamic events. The increasing disorder in the energy distribution of the universe predicted by this law is irreversible to the extent that the direction of time is irreversible.

Now, according to the Dirac theory of matter, the existence of antimatter is predicted by the negative energy solutions of the relativistic wave equation. In fact, the antistates, which are required to explain the physical meaning of the two mathematically indistinguishable signs of imaginary charge, are introduced into the theory along with the additional imaginary space-time symmetry of relativity. According to quantum theory, of course, the negative energy states are equivalent to negative time states. Thus a prevalence of positive energy states over negative energy states or matter over antimatter. By a further consideration, the conservation of energy ensures that negative energy states can only be created at the same time as positive energy states; antimatter can only be produced at the same time as matter. Disordered radiation cannot produce ordered matter without also producing antimatter which can recreate the disorder; the negative energy states are, in effect, equivalent to positive entropy states. Of course, "negative" energy has no physical significance in itself and is merely a mathematical device of the Dirac theory, for, on the principle of CPT invariance, antimatter actually describes the absence of negative energy, and the true mass-energy associated with either particles or antiparticles is always positive.

(22) Spontaneous Symmetry Breaking

In principle, all nongravitational forces should be identical; in real systems, however, differences arise to set each force apart from the others in magnitude and in characteristic properties. This is now generally attributed to the process of spontaneous symmetry breaking, which allows forces to remain symmetric in principle, while an asymmetry is introduced by the quantum or energy state of the system. In the weak interaction, this results in the weak field acquiring the three intermediate vector bosons W^+, W^- and Z; in the unified theory of weak and electromagnetic interactions, the differences between the forces are attributed to the original mass differences between the intermediate bosons and the photon, and these exchange particles are grouped into a single family. The electromagnetic part of the unified gauge symmetry remains unbroken because the photon is massless. The masslessness of the photon, of course, also ensures that quantum electrodynamics is renormalizable, but, with the addition of the neutral Z particle to the theoretical scheme, the combined theory also becomes renormalizable.

New theories suggest that even the strong interaction is separated from the weak and electromagnetic only by the spontaneous breaking of a symmetry. It is supposed that, before the symmetry breaking, all the three forces were of the same strength and all the quarks and leptons were massless; there were three generations of fermions (because there were three forces?), each including two quark flavours and two leptons; the spontaneous symmetry breaking then imparted masses to all the fermions except for one neutrino in each generation. Such theories have been successful in accounting for the masses of the heavier quarks, which cannot be derived from the masses of the lighter mesons and baryons, and they have also been able to accommodate the new heavier leptons.

The unified theories are a natural result of a quark theory based on unit charge. Spontaneous symmetry breaking may be expected to result from situations in which the masses of particles result from the absence of charges which should otherwise be present to maintain symmetry. The unified theories also assume a close relationship between quarks and leptons with a mechanism for interconversion within families, and we have seen that each known lepton has an identical charge structure with (and may even be identical to) one particular quark of the appropriate flavour.

(23) The Equalization of the Four Forces

The figure of 10^{15}GeV for the energy at which the three nongravitational forces are expected to be equal is obviously of some special significance and it may be reasonable to speculate that it is related to the size of the gravitational interaction. If the value of the electromagnetic charge increases with increasing energy of interaction, then, according to the property of group multiplication, c must decrease from the value measured by this interaction. Let us suppose, for a fundamental charge Q, that the expression

$$Q^2 = hc$$

is needed to explain the quantization of energy, and that $U = 10^{15}$GeV is the energy equivalent of two particles of mass m whose gravitational force is exactly equal to the force between two fundamental charges over the same distance.

$$\begin{aligned} \text{i.e.} \quad Gm^2 &= Q^2 \\ \text{Now} \quad U &= 2mc^2 = \frac{2mQ^4}{h^2} \\ \text{and so} \quad Q^2 &= \frac{GU^2h^4}{4Q^8} \\ Q^{10} &= \frac{GU^2h^4}{4} \end{aligned}$$

This gives a velue for Q^2 of 2.2×10^{-27} Nm^2 which must be close to the predicted value since it lies between those found for the electromagnetic and strong interactions.

We can thus postulate the existence of a fundamental unit of mass equivalent to the fundamental unit of charge. If dimensional analysis is applied to reduce the fundamental constants h, c and G to unity, we can also derive a fundamental unit of length, $(Gh/c^3)^{\frac{1}{2}}$ or R/2, where R is the Schwarzschild radius $(2Gm/c^2)$ for mass m, and a related unit of time, R/2c. The fundamental unit of length would be the size of particle required to maintain a fundamental charge at 10^{15}GeV or with mass m. Replacing h by $\hbar$ produces units of length and time which are of the order at which the supposed effects of quantum gravity become significant. The existence of such fundamental units as Q, m, R/2 and R/2c indicates that, for systems involving charges, there is a fixed relationship between each of the members of the group of order 4. This relationship emerges from the equations relating mass and energy, introduced by the 4-vector system, and energy and time, introduced by quantum mechanics.

(24) The Fundamental Constants

It is unlikely that the value of any of the fundamental constants is arbitrary. The measured values for the strong, weak and electromagnetic charges and their relationships to the fundamental charge will presumably emerge from the details of the unified gauge theory of the three interactions. It may be significant that the strong charge appears to be of order $(2hc)^{\frac{1}{2}}$, where c is the measured value of the velocity of light; the strong charge may determine the value of this latter constant (to equate the field energy between two confined charges to the energy required to confine the particles' wave functions). The measured value for the velocity of light determines the "bare charge" of the electron, $(\hbar c)^{\frac{1}{2}}$ (which is related to the Compton wavelength $\hbar c/m_e$), and the Coulomb charge e results from renormalization. There is a possibility that the Coulomb charge is equal to the hypothetical bare charge $(\hbar c_o)^{\frac{1}{2}}$ at that velocity of light at which all four forces are equal; the value of c_o depends critically on the energy U; for $U \sim 3 \times 10^{15}$ GeV, $c/c_o \sim 1/\alpha$ and $e^2/4\pi\varepsilon_o = hc_o$, but for $U \sim 10^{15}$ GeV, $c/c_o \sim 2/3\alpha$ and for $U \sim 0.5 \times 10^{15}$ GeV, $c/c_o \sim 1/2\alpha$. The fine structure constant α may well determine the c/c_o ratio and, because it results from renormalization, it is also related to the ratio of strong and electromagnetic charges $(s^2/e^2 = 4\pi/\alpha)$. The value of the weak charge is unknown but $e^2/w^2 \geqslant 1/4\pi\alpha$ and $s^2/w^2 \geqslant 1/\alpha^2$; if $s^2/w^2 = 1/\alpha^2$, then the ratio of strong to weak field energies could be comparable to mc^2/mc_o^2. The exact reasons for these relationships between c, c_o, e, s and w are not yet established but they seem to depend only on the value of the fine structure constant.

If the fine structure constant is derived either from renormalization or from an independent derivation of s/e or m_p/m_e the remaining problem will be the actual mass of the electron or its charge to mass ratio. There have been several speculations concerning this ratio and the associated classical radius. Some of these are linked with cosmological phenomena such as the Hubble redshift. One such speculation may be made by returning to the question of inertial forces. We described inertia as an intrinsic property of mass without need of physical explanation, the preferred role of inertial frames in classical mechanics being a consequence of the imaginary nature of time, but it is nevertheless often convenient to assume the existence of fictitious inertial forces to preserve the form of Newton's laws of motion in the noninertial frames resulting from states of absolute rotation. We may suppose that inertial forces are the result of choosing an electromagnetic system of measurement to describe the movements of bodies governed by gravitational forces. The effects of these may well be similar to those predicted by relativistic-type theories if we could remove from the latter the supposed nonlinear effects of gravity.

The true space-time of classical mechanics is such that it describes an inertial frame of reference. However, applying a 4-vector space-time with invariant c produces, by application of the Lorentz transformations, apparent effects of rotation of the coordinate system; classical mechanics nevertheless preserves its form by assuming the existence of absolute rotations which need not be generated by physical forces. We may assume that the apparent rotation produced by the 4-vector space-time is analagous to the magnetic force in electromagnetic systems and that it is equivalent to the effect of a fictitious inertial force, described by

$$F = -Kma$$

where K, the ratio of gravitational to inertial mass, is of the form $(4\pi/3)G\rho\,(r/c)^2$ for a sphere of radius r enclosing matter of mean density ρ. (There is some doubt about the exact expression for K, but it is certainly of this order). With K made unity by choice of units and taking ρ to be the mean density of the universe ρ_u, we may define a radius r_u (the Hubble radius) at which the measured force of inertia on mass m_1 would be identical to the fictious acceleration (a)-dependent inertial force produced by the enclosed matter (m_2)

$$F = \frac{Gm_1\, m_2\, a\,\phi}{c^2 r} = \frac{4\pi}{3}\; Gm_1\,\rho\,\frac{r^2}{c^2}\,a\,\phi$$

(ϕ, the angular dependence is assumed to be unity for the approximate calculation.) (9) With the gravitational force defined as the static part of the inertial interaction, we may derive the external gravitational field of the matter within the Hubble radius as Gm_u/r_u^2 where $m_u = (4/3)\,\pi\,\rho_u\, r_u^3$. At r_u, the inertial force of repulsion equals the gravitational attraction, and massive bodies have an apparent velocity c and acceleration c/t, where t is a time of interaction r_u/c.

Assuming that the inertial electron mass is fundamental and equal to that which would be "caused" by the inertial reaction of the other matter within the Hubble radius, we would expect the external gravitational field of the electron Gm_e/r_e^2 to be of the same order as the inertial reaction on unit mass (i.e. Gm_u/r_u^2); the electron of radius r_e thus remains stable because the apparent inertial reaction or repulsive force due to the other matter within r_u exactly equals the attractive force of gravitation. With the exact relationship $m_e r_e = e^2/4\pi\varepsilon_o c^2$ also known, we may derive individual values for the electron mass and classical radius for any assumed mean density of the universe.

In general, a body within the Hubble radius at distance r will have an apparent acceleration v/t and an apparent velocity v, where t is the time of interaction r_u/c. Hence

$$r = vr_u/c = v/H$$

where H is the Hubble constant. Any actual velocity of source would not alter the measured value of c but v is a fictitious velocity due to the choice of reference frame. The relationship between v and r produces a progressive apparent slowing down of light over distance, with a steady decrease of frequency due to the effects of the time of interaction r_u/c. This is equivalent to a de-energising of photons; the energy lost becomes background radiation because it is a measure of uncertainty or retardation due to the finite interaction time. The uncertainty is a measure of our lack of knowledge of the true (classical) space-time; such uncertainty is a fundamental component of matter and ensures that space always has a mean temperature whose precise value is related to the mean density of matter and may also depend on the ratio of matter to antimatter in the universe.

If any such theory is correct, the mean density of the universe is a significant quantity in determining the value of fundamental constants. In a universe of mean density equivalent to that of a fundamental charge (ρ_o), all forces would be of equal strength and c would be c_o. The differences between the strengths of the three nongravitational interactions appear to be related to the fine structure constant (α) and so it is probable that the latter is related to, and possibly even determines, the observed mean density of the universe. According to current estimates of the Hubble constant, $m_u r_u/m_e r_e \sim 10^{124}$ and $m_u/m_o \sim 10^{62}$, where m_o is the mass equivalent of a fundamental charge; the former number is of the same order as exp $(2\pi/3\alpha)$, which is of possible significance in the renormalization of the electron's mass, but whether this establishes a genuine connection between ρ_u and α has yet to be decided.

(25) The Unifying Symmetry

Physical theories are usually considered successful when they explain a considerable number of facts with a minimum of assumptions. Fundamental physics certainly presents us with a considerable number of facts to explain, but many modern theories, being speculative, have actually multiplied, rather than reduced, assumptions. It is especially significant that none has attempted to reduce the most fundamental assumptions of all, those concerning the fundamental parameters space, time, mass and charge. Here we propose that this task may be accomplished by applying to these parameters that concept of symmetry which has been so successful a unifying principle in other areas of physics.

There are reasons for believing that symmetry is a fundamental aspect of the natural world and our only assumption is that the general property of symmetry must apply in particular to the actual process of measurement. Thus, if we begin by defining the parameter space to possess the properties which we would expect of a system of direct measurement, then the existence of parameters with the characteristic properties of time, mass and charge is necessary to preserve the general symmetry of nature. The mathematical character of physical laws is a result of these inherent characteristic properties and the most convenient presentation of the general symmetrical relationship between the parameters also takes mathematical form as a group of order 4.

Individual laws of physics evolve from particular aspects of this general symmetry. A fundamental distinction between mass and charge enables us to define a system (i.e. method of measurement) which excludes the latter; the conditions which are necessary for such a system to exist are the laws of classical mechanics. A symmetry which exists between mass and charge then allows us to extend the concept of force to charges; at the same time we are obliged to recognize a complementary symmetry between space and time. The combination of these two conditions leads directly to the laws of electromagnetics and eventually to those of quantum mechanics and thermodynamics. A further symmetry, that between space and charge, when combined with the general CPT invariance, is responsible for the system of coloured quarks representing unit charges; from this we derive the properties of the fundamental particles and the identifying characteristics of the electromagnetic, strong and weak interactions. The unified gauge theory for these interactions, involving the process of spontaneous symmetry breaking, emerges naturally from the system. It enables us to find relationships between some of the fundamental constants and there is every probability that further such relationships will be discovered as the details of the theory are worked out.

Though there is possibly an area of doubt as to whether classical or relativistic theories should be accommodated to the general structure of fundamental physics (the former being favoured by the logical derivation), there is no major law or principle of physics which cannot be derived from the symmetrical properties of the parameters. By using simple conceptual arguments, we have demonstrated that the many assumptions concerning forces and particles can be replaced by a single unifying symmetry which is, in effect, the most fundamental principle of physics.

REFERENCES

(1) L.J. Schiff. Amer. J. Phys. 28, 340, 1960

(2) R.d'E. Atkinson. Proc. Roy. Soc. A 272, 60, 1963

(3) F.H.J. Cornish. Proc. Roy. Soc. A 273, 413, 1963

(4) R.H. Dicke. Reviews of Modern Physics 29, 363, 1957

(5) R.H. Dicke. Phys. Today 20,55-70. 1967

(6) R.H. Dicke and H.M. Goldenberg. Phys. Rev. Lett. 18, 313, 1967

(7) M.Y. Han and Y. Nambu. Phys. Rev. 139 B, 1006, 1965.

(8) K. Johnson, M. Baker and R.S. Willey. Phys.Rev.Lett. 11, 518, 1963
Th.A.J. Maris, V.E. Herscovitz and G. Jacob. Phys. Rev. Lett, 12, 313, 1964

(9) D.W. Sciama. "The Physical Foundations of General Relativity", London 1972, pp.36 ff.