THE FUNDAMENTAL PARAMETERS OF PHYSICS

by Peter Rowlands

1, Sheerwater Close, Bruche, Warrington,
Cheshire, England. WA1 3JE.

ABSTRACT

The four parameters space, time, mass and charge are shown to possess an exact symmetry as a group of order 4. The exact properties of the parameters as displayed in this group are then used to propose derivations of the fundamental principles of classical mechanics, electromagnetic theory, quantum mechanics, thermodynamics and particle physics, together with some applications in cosmology. The derivations suggest that the laws of physics and the fundamental particles have a single origin in the initial process of direct measurement.

(1) A Fundamental Symmetry

The fundamental axioms of physics are not the laws of physics but the physical parameters through which all physical measurements are made and in terms of which the laws of physics are defined. Space, time, mass and charge (where charge refers generally to the sources of electromagnetic, strong and weak interactions) are regarded as elementary components in nearly every physical theory but it is not generally recognized that they may be defined in an explicit way through the discovery of their exact properties. The most fundamental principles of such comprehensive systems as classical mechanics and electromagnetics are those which either define individual properties of the parameters (e.g. conservation of mass, conservation of charge) or describe the way in which physical measurements, interpreted through different parameters, are interrelated (e.g. the definition of force and all laws relating to force); they are not concerned with physical phenomena but with the process of measurement. In this case it ought to be possible to define the fundamental parameters in such a way that the laws of physics, in particular those of classical mechanics and electromagentism, become necessary consequences of the definitions. This may even lead us to discover why we use these particular parameters as elementary principles of physical measurement.

Now, it is a remarkable fact that, although we have four physical parameters, only one is known by direct measurement. Measurement of space (or, at least, of distance) involves an element of enumeration or division into a finite number of discrete components; it is a counting process based on the natural numbers or integers and it is impossible to imagine how any form of direct measurement could be otherwise. 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 measurement 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. Masses and charges are, of course, known only through the forces which result from their interactions and these are known only through measurements of space and time.

Since we believe the explicit properties of parameters to be the fundamental axioms of physics, we may assume that mass and charge are conserved because conservation is fundamental to their definition. We also know that all masses and all charges interact mutually to produce the forces which are responsible for all physical processes. At the same time, the other two parameters, space and time, are not conserved and do not interact. We may assume, therefore, that the property of conservation is linked to the property of interaction. Space and time produce the initial information about a system because they are variable or nonconserved and because each determination of space and time is separate and independent from any other. If nonconservation is thus linked to noninteraction, it is reasonable to suppose that conservation must be linked to interaction.

While mass and charge are thus so far similar, they are dissimilar in their types of interaction. Masses are always positive and always produce attractive forces; charges can be either positive or negative and can produce either attractive or repulsive forces; in particular, two charges of the same sign produce a force in the opposite direction to two masses. To explain this, we could introduce the mathematical device of making all masses real and all charges imaginary. This would complement the use of real space and imaginary time already established for 4-vector systems. Mass and the three charges (electromagnetic, strong and weak) could be represented as the real and imaginary parts of a quaternion in the same way as space and time are represented as the real and imaginary parts of a 4-vector. Imaginary charge would explain the existence of antistates to all known particles, since equations involving +i and -i are indistinguishable, while the use of a truly imaginary time would explain why this quantity is always measured through the effects of force and acceleration (as t^2) and never through the effects of motion (as t).

There is, however, yet another fundamental distinction between mass and charge. There is only one type of mass, but there are three types of charge, each of which interacts only with its own kind. At the same time, charge, unlike mass, occurs only in discrete units which can be counted; charge is localized in what appear to be point-sources, but mass (which includes energy) is a continuum in the sense that it is present in all systems, has an unlimited range of possible values, and is defined only in terms of a spatio-temporal distribution which can never be known exactly because it requires knowledge of the interactions between all masses. Now, space is three-dimensional like charge and is measured by dividing into units which can be counted, but time is one-dimensional like mass and is also a continuum measurable only by relationship to discontinuities in space. (The discontinuities arise because space is multi-directional or multi-dimensional; it is impossible to imagine a periodic discontinuity in space without a change in direction.) Space and charge may be described as denumerable quantities, mass and time as non-denumerable, and we may assume that the property of denumerability precedes that of dimensionality. In this context, it is significant that the dimensions of both space (in vectors) and charge (in forces) are added as squares according to the Pythagorean equation. Imaginary quantities, of course, have meaning only as the square roots of real quantities. For charges, squaring removes the imaginary quaternion operators and expresses the mutual interaction between one unit of charge and any other. This suggests that there must be a real meaning to adding the squares of charges, which now become real integers. The symmetric application of addition by squares to space then actually introduces its vector or dimensional properties, for it implies the existence of an alternative combination of distances (via the Pythagorean equation) to the direct addition which is involved in the measurement of an individual distance. Addition by squares also introduces space's property of reflection symmetry, a parallel to charge's introduction of particles and antiparticles. Once the concept of dimensionality for space and charge is established, it is easy to see why it has the particular order of 3, for this is the minimum number of independent imaginary numbers for which an algebra can be created.

Our discussion, so far, has revealed the existence of certain symmetries between each pair of parameters. It is instructive to combine these in a single table. Assigning the arbitrary symbols +a, +b, +c to the properties of space, and using the corresponding negative symbols, where appropriate, for the alternative properties, we may set out the primary properties of the four parameters as follows:

space	real	nonconserved	denumerable	+a +b +c
time	imaginary	nonconserved	nondenumerable	-a +b -c
mass	real	conserved	nondenumerable	+a -b -c
charge	imaginary	conserved	denumerable	-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:

$$+a * +a = -a * -a = +a$$
$$+a * -a = -a * +a = -a$$

and similarly for b and c. Here, space is the identity element , though the symbols could be rearranged to assign the property to any of the other parameters; each element is its own inverse; and the multiplication rule identifies the group as Abelian.

This remarkable and unexpected symmetry cannot be accidental and we may suspect that it is related to the fact that space is the only parameter used in direct physical measurement. Space appears to have all the properties required for a parameter of direct measurement - it is real, nonconserved (and,therefore, independently variable) and denumerable - but this does not explain why, having decided the utility of direct measurement, we are obliged also to use three additional parameters each of which has only one of the required properties while being in other respects the exact opposite. While avoiding extensive speculation on matters which are mostly relevant to philosophy, we may suggest that direct measurability may not be a property of the universe as a whole. The system of four symmetrical parameters, only one of which is related to direct measurement, may enable us to use our concept of measurement to describe a universe which is not conditioned by it.

Symmetry has, of course, been used with considerable success as a unifying principle in the other areas of physics, especially in the theory of particles, and there is no reason why it should not be a significant component of what may be the most fundamental set of all physical axioms. We may assume that the symmetry which we have discovered between the four fundamental parameters is exact and is, indeed, the <u>source</u> of the properties of time, mass and charge. In that case, we need only define space to have the properties required for use in direct measurement and the definitions of the other three parameters will automatically follow. Such mysterious physical phenomena as inertia and gravitation then become inevitable consequences of the actual process of direct measurement.

(2) The Properties of Space

The properties of space are of particular importance because it is the only parameter of which we have direct knowledge - all three properties are directly accessible. Each of the other parameters has one of these properties and so is partially accessible: masses are real, charges are countable and time is nonconserved, making it identifiable by its changes. The group of order 4 is, in fact, the simplest complete symmetry in which each parameter remains at least partially accessible; to replace it with a group of order 2, for example, would require the definition of a single parameter replacing time, mass and charge, which would be totally inaccessible, while a group of order 3 would not be a complete symmetry. It seems, in some respects, that the separation of the characteristics of space into three distinct properties is required by the necessity of incorporating it into a group of order 4, for the three defining properties seem to be, in some sense, aspects of one; thus the countability of numbers leads eventually to the real-imaginary distinction while at the same time suggesting, or being suggested by, the idea of change or nonconservation.

The fact that space shares only part of this fundamental identity with each of time, mass and charge leads to subtle distinctions in the characteristic ways in which these different parameters interpret the shared properties. Thus, the apparent differences between the countability of space and charge - charge having fixed units - arise from the fact that two properties are involved rather than one. Charge is unlike space because it is both countable and conserved, whereas space is countable but not conserved. Conservation fixes charge but nonconservation makes space arbitrary. With respect to this property, the two parameters must be exact opposites - if we could count fixed units of space in a system, then it would have to be a conserved quantity like charge. Because space and time are not conserved, we can put no restriction on their variation; if space existed only in discrete fixed units we would be imposing a restriction on its variation, but we impose no such restriction by saying that it is merely countable. Time, of course, offers an example of a parameter which is unrestricted in its variation but not countable; space, though arbitrary like time, is not, likewise, a continuum and so we can have points of zero space representing the arbitrary divisions which make it countable.

Of course, where only one property is involved, we have exact correspondences between space and the other parameters. The most significant example of this is the existence of three-dimensional Pythagorean systems for both space and charge. Three-dimensional space is necessary to give physical representation to quaternions (which explains why quaternions were used in early forms of vector analysis), and quaternions representing mass-charge are in every respect the symmetric complements of 4-vectors representing space-time. This is demonstrated by the interpretation of the quaternion product $ij = -k$. Though the result of the product of two imaginary quaternion operators is a third, this does not represent a third charge, derived from the product of two, any more than the area vector derived from the product of the two dimensions x and y represents a third dimension to be added to the original two. In the case of charge, the product of i and j represents the nature of the quaternion applied to the third charge, while, in the case of space, the product of x and y represents the direction of the third dimension.

(3) Classical Mechanics and Electromagnetic Theory

Since the parameters are the sole means by which the raw data of physical measurement is channelled into usable physical information, we may assume that the information derived through the members of the parameter group is not only subject to its absolute symmetry but is also absolutely exclusive; it is possible to see, even without elaborate mathematics, how the elementary principles of classical physics become a matter of organizing this exclusive information.

In general, the variable, nonconserved parameters, space and time, are used to derive information about the invariant, conserved parameters, mass and charge; the former represent information available, the latter information required. A "system" (by which we mean nothing but a method of organizing measurements) may be defined as a set of interrelated values of space, time, mass and charge, its permanent feature being the invariant value of the two latter quantities. There is no limit to the combination or division of systems, with the corresponding combinations or divisions of mass and charge, except that charge exists only in discrete units. These units may be negative or positive, and so it is possible to have a system with zero charge. Mass, however, is a continuum and so cannot be excluded from a system. This means that mass cannot take both positive and negative values, for systems with zero mass would otherwise be possible. (To say that a parameter is a continuum means that it cannot take zero or negative values.) A remarkable corollary of this property of mass is the unidirectionality of time, for time is the other parameter which is a continuum; this is a striking confirmation of the exact symmetry between the parameters. Time, however, is imaginary and has positive and negative mathematical representations which are physically indistinguishable. Thus, even though it is not possible to reverse time, it is possible to discuss the symmetry of physical laws under an imagined reversal of time. Since we can never specify the actual sign of time, we are unable to ensure that the laws of physics maintain the same form following a sign reversal in any individual parameter; we can only guarantee symmetry under sign reversal for the group as a whole (the CPT theorem).

The laws of classical mechanics and electromagnetic theory may now be derived from the individual properties of the parameters combined with the general properties of the D2 group. The quaternion representation of charge requires completion by the addition of a real term, which in this case must be mass (since space is three-dimensional). This establishes a fixed numerical relationship between mass and charge, expressed (if we neglect the quaternion operators) by the equation $Gm^2 = Q^2$. By symmetry, there must also be a fixed numerical relation between space and time, which we may express in the form $r = ct$; 4 - vector space - time is thus introduced by direct symmetry with quaternion mass-charge. However, an examination of the D2 group reveals that every element exists, in terms of group multiplication, in a fixed relation with every other. Also, in this particular case, there is no special characteristic of any of the four parameters that specifies its identification with any particular element, and any of the parameters can be placed as the identity element. Consequently, if the fixed group relation is identified as a fixed numerical relation in any particular case, it must also be a fixed numerical relation in all other cases. This leads to the direct relationship between mass and space, $Gm = c^2r$, more familiarly expressed as $E = mc^2$.

The direct relationships between all the parameters thus become obvious consequences of group symmetry and explain the existence of the fundamental constants e (or Q), c and G; three independent relationships require three independent constants. The elimination of these constants leads to fundamental units of each of the four parameters: $G^{1/2}\,Q/c^2$; $G^{1/2}\,Q/c^3$; $Q/G^{1/2}$; Q. However, the D2 group also suggests the existence of another set of relations between the parameters for each element is its own inverse and relations involving elements must also be valid for relations involving their inverses. Thus, for instance, where we have a relation between mass and time, we must also have a relation between mass and the inverse of time, and, as the first was numerical, so must be the second. For the first we have $m = c^3 t/G$ and for the second we introduce a new constant h so that $m = h/c^2t$ or $E = h\nu$. From this equation we can derive relationships between any individual parameter and the inverse of any other. Identifying Q^2 or Gm^2 with hc (as in a Grand Unified Theory) would be equivalent to making each parameter numerically identical to its own inverse. It is significant that quantization of energy does not depend on the validity of this identity - it only requires that each parameter has an inverse with identical group properties - but the analysis suggests that such identification is strongly probable when all the forces become equal in value.

The relations between the parameters enable us to set down four numerical equations:

$$\frac{Gm^2}{r} = \frac{Q^2}{r} = mc^2 = \frac{h}{t} = \frac{hc}{r}$$

which are in themselves sufficient for the derivation of the laws of classical mechanics and electromagnetic theory. (The equations involving h may require some adjustment of the constants, but this does not have a significant effect on the argument.) The derivations all make an explicit introduction of the conservation of mass within a system. This is accomplished by isolating a term equivalent to mx a constant on one side of the equation and differentiating with respect to r or t to establish the invariance of m with respect to the arbitrary parameters of the system. The first result is that energy (Gm^2/r or Q^2/r, considered as a summation within a system) must also be conserved ($= \sum mc^2$). (Adding kinetic energy, of course, changes the system by adding mass and changes the relationship between mass and the potential energy term, Gm^2/r or Q^2/r, accordingly.)

For Newton's second law of motion, we begin with $Gm^2t = hr$, take differentials, and substitute t/r^2 for m/h, to find that

$$\frac{Gmt}{r^2} = \frac{mc^3t}{h} = \frac{dr}{dt} .$$

Then, multiplying throughout by m and differentiating with respect to t, we find that

$$\frac{Gm^2}{r^2} = m\frac{d^2r}{dt^2} .$$

This gives us the definition of force and ensures that gravitational and inertial mass are identical. The r term on the right hand side makes force a vector (t being the same throughout the system), and a full analysis, with imaginary signs included, would make the gravitational force vector take the opposite direction to the space vector because of the negative sign of t^2. This means that the gravitational force must be attractive. The numerical identity of Gm^2/r^2 and Q^2/r^2 then introduces the concept of electrostatic force. Because the sign of Q^2 / r^2 is already negative for identical charges, the electrostatic force between these must be in the same direction as the space vector, and hence repulsive.

Newton's third law of motion is derived by differentiating the equations

$$\frac{Gm^2}{r} = \frac{Q^2}{r} = mc^2$$

with respect to r. If the total mass is conserved within a system, then the sum of the force terms represented by Gm^2/r^2 and Q^2/r^2 must be zero. In the simplest case, this means that every action has an equal and opposite reaction. Conservation of momentum, which follows from Newton's third law, also follows from conservation of energy; numerically, momentum is $m \times Gm\ t/r^2 = Gm^2/rc$ and this is conserved if Gm^2/r is conserved.

The differential forms of Newton's law of gravitation and Coulomb's law of electrostatics can be derived by defining gravitational and electrostatic field terms, $Gm/r^2 = mc^3/h$ and $Q/r^2 = mc^3/G^{1/2}h$, which, when differentiated with respect to r in a system in which mass is conserved, become zero. The field terms are equivalent to d^2r/dt^2 and are additive directly as vectors, without regard to the masses with which they are associated, and so these laws extend the relationships between the parameters into <u>universal interactions</u> involving all masses and charges, and make explicit the link between conservation and interaction.

Newton's laws of motion and gravitation and Coulomb's law of electrostatics are, together with the 4-vector representation of space-time, all the principles we need for classical mechanics and electromagnetic theory. Classical mechanics assumes that the effects of the 4-vector system are negligible, while the principles of electromagnetic theory, such as Maxwell's equations and the Lorentz transformations, express the application of the 4-vector system to Coulomb's law of electrostatics; in such cases the original r is simply replaced by the 4-vector term. Fundamental particles, in particular, as electromagnetic systems, may be assigned a 4-vector dynamics without the introduction of phenomenological assumptions. Special relativity, which makes such assumptions as the constancy of the velocity of light in all inertial frames of reference and the use of a particular definition of time simultaneity, may be regarded as a heuristic device introduced to explain the intrinsically established 4-vector system.

(4) Symmetries and conservation laws

Space and time are subject to certain symmetries precisely because they are nonconserved parameters. Systems are defined by their unchangeable masses and charges but are arbitrary in their choice of space and time. Consequently, space and time are both symmetric under arbitrary tranlations, and space, as a three-dimensional parameter, is also symmetric under arbitrary rotations. These properties are not characteristic of the conserved parameters mass and charge; thus, if charge possessed the same rotation symmetry as space, the electromagnetic, strong and weak charges would not be distinguishable from each other and would not be independently conserved.

Using Noether's theorem about invariance under continuous groups of transformations, it becomes possible to associate the classical principles of conservation of linear and angular momentum with the respective translation and rotation symmetries of space, and the conservation of energy with the translation symmetry of time. In 4-vector systems, the conservation of 4-momentum becomes associated with a translation-rotation symmetry of space-time. Also, since energy is directly related to mass, the fundamental principle of conservation of mass becomes a particular consequence of time's translation symmetry.

This is a remarkable illustration of the fundamental symmetry between the parameters. Mass and time are opposite in regard to the property of conservation; mass is conserved because time is not, in the same way as charge is conserved because space is not. The conservation properties of mass and charge are the precise consequences of the nonconservation properties of space and time. (The nonconservation of space and time also leads to the fact that there are no fixed coordinates for systems of masses and charges; these quantities are known only through their interactions and knowledge is restricted, in the first case, by the fact that mass is a continuum, and in the second case, by the process of retardation; here, again, the property of conservation is linked to that of interaction.)

(5) The Quarks

We are, of course, aware that there are three types of charge and that electromagnetic theory alone will not suffice to explain the properties of strong and weak interactions. For this we need also a comprehensive theory of particles. However, 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. This is because there is a problem in relating the imaginary quaternions i, j, k to the physical manifestations of the electromagnetic, strong and weak charges (e, s, w). 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. Mathematical imaginary quaternions are not physically distinguishable, unlike spatial dimensions, but the three charges ought to be identifiable because, though space has rotation symmetry, charge does not, and e, s and w charges must be distinguishable from each other. By analogy with space and mass, unit charges ought to have unit interactions with other charges of the same kind and we must know how one system of charge containing e, s, w interacts with any other. Another problem arises from the fact that, for a quaternion representation, we would expect all the unit charges to be of equal value, whereas 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.

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 (1), they possess either unit or zero charge. (In either case, the average charges would be fractional and, for permanently confined quarks, the models would be phenomenologically indistinguishable.) Quarks may be considered as particles which would, ideally, contain one unit, either positive or negative, of each of e, s and w. There are eight such combinations, to which we may assign the labels of the u, d, c, s quarks and the $\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$ antiquarks:

$$(\pm e + s + w) \quad u, d \qquad (\pm e + s - w) \quad c, s$$

$$(\mp e - s - w) \quad \bar{u}, \bar{d} \qquad (\mp e - s + w) \quad \bar{c}, \bar{s}$$

Now, we cannot simply assign quaternion values to e, s, w in these combinations because i, j, k would then be physically identifiable, but if each were, at any one time, one of three representations, arbitrarily defined by "colours" - say blue (ie, js, kw), green (ie, ks, jw), red (je, is, kw) - then we would not be specifying a unique quaternion value for any particular charge. There are two obvious ways in which this could be done:

(i) ie, js, kw
ie, ks, jw
je, is, kw

(ii) ie, js, kw
ke, is, jw
je, ks, iw

With the actual quaternion labels arbitrary, all other assignments would be essentially identical to one of these.

Nevertheless, the system would be inconsistent if charges of the same type could be associated with different quaternion values. Thus, if we assigned a unit value to ie in the blue quark, then we could not also assign a unit value to je in the red quark. It would, therefore, be necessary to find a system combining quarks and antiquarks of each type or flavour in three-colour or colour-anticolour combinations in which a blue quark with unit ie would be indistinguishable from a red quark of the same flavour with <u>zero</u> je. (The system would have to be of type (i) because type (ii) would then associate charges with specific quaternion values.)

We can show that such a system does exist, but this depends on two conditions: the first is that the division between quarks and antiquarks must be made by assigning a positive unit of one type of charge (s) to all quarks and a negative unit to all antiquarks; the second is that at least one other type of charge (w) must be allowed to change sign, when appropriate, by assuming a simultaneous sign reversal in other parameters such as space and time. These conditions are ultimately responsible for all the differences between the three nongravitational interactions. Typical arrangements of the quarks in this system are given in the following tables:

(a)

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

(b)

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

There are also arrangements which allow the possibility of further quarks (t, b):

(c)

		Blue	Green	Red
u	+e +s +w	1i 0j 1k	1i 1j 0k	0k 0i 0j
d	−e +s +w	0i 0j 1k	0k 1j 0i	1i 0k 0j
c	+e +s −w	1i 0j 1k	1i 1j 0k	0k 0i 0j
s	−e +s −w	0j 0i 1k	0k 1j 0i	1i 0j 0k
t	+e +s ±w	1i 0k 0j	1i 1j 1k	0j 0i 1k
b	−e +s ±w	0j 0k 0i	0i 1j 1k	1i 0j 1k

(d)

		Blue	Green	Red
u	+e +s +w	1i 0k 0j	1i 1j 0k	0j 0i 1k
d	−e +s +w	0j 0i 0k	0k 1j 0i	1i 0j 1k
c	+e +s −w	1i 0k 0j	1i 1j 0k	0j 0i 1k
s	−e +s −w	0j 0i 0k	0k 1j 0i	1i 0j 1k
t	+e +s ±w	1i 0j 1k	1i 1j 1k	0k 0i 0j
b	−e +s ±w	0j 0i 1k	0i 1j 1k	1i 0k 0j

With the values of 1 or 0 as assigned to the unit charges in these tables, 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 their respective antiquarks presents us with a unique way of combining the charge components of different systems without specifying their quaternion values; no other representations are actually possible. 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 related 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 invariance of quarks to colour transformations within "colourless" combinations. The laws of physics are so defined that the sign reversal in charge usually requires a simultaneous reversal in the sign of the space coordinate or parity operator, but for t and b quarks it may be necessary for the sign reversal in w component to be independent of that for the u, d, c and s quarks. This means that a six-quark model requires the alternative possibility of a simultaneous reversal in the sign of time. It is significant that, in the unified theories of weak and electromagnetic interactions, six quarks are the minimum necessary to give violation of time-reversal symmetry via the weak coupling matrix.

(6) 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±w	$p\pi^-$; $n\pi^0$
Σ^-	dds	−e+s±w	$n\pi^-$
Σ^0	uds	+s+w	$\Lambda\gamma$
Σ^+	uus	+e+s±w	$p\pi^0$; $n\pi^+$
Ξ^-	dss	−e+s±w	$\Lambda\pi^-$
Ξ^0	uss	+s±w	$\Lambda\pi^0$
Baryon singlet			
Λ	uds	+s±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±w }	
Σ^{*0}	uds	+s±w }	$\Lambda\pi$; $\Sigma\pi$
Σ^{*+}	uus	+e+s±w }	
Ξ^{*-}	dss	−e+s±w }	$\Xi\pi$
Ξ^{*0}	uss	+s±w }	
Ω^-	sss	−e+s−w	$\Xi^0\pi^-$; $\Xi^-\pi^0$; ΛK^-

These structures explain many significant facts. Thus it is obvious that π^0, which has zero charge structure, must be its own antiparticle, whereas K^0 or $\bar{K}^0$ which may have a charge structure ± 2w, is not. 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 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 any of the individual C, P or T symmetries. But when K^0 and $\bar{K}^0$ and out of phase (the K_2^0 decay), the charge structure is nonzero and the time-reversal symmetry must be violated in order to allow the particle to decay to $\pi^+ + \pi^-$ with zero charge structure.

Also, it is clear why Ω^- is the only member of the spin 3/2 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. The average number of negative weak units of charge (n) in the members of a multiplet is expressed in terms of the strangeness (-3n), a quantity which is conserved in strong interactions because the latter involve only the s component of the charge structure; any decay of Ω^- will reduce this number. 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: the strangeness-conserving current (which retains the sign of w) and the strangeness-changing current (which reverses the sign of w). The two are produced by the Cabibbo mixing, by which the eigenquarks of weak isospin become

$$d_\theta \equiv d \cos \theta + s \sin \theta$$
$$s_\theta \equiv s \cos \theta - d \sin \theta$$

in contract to the d, s eigenquarks of strong interactions. The mixing is the direct expression of the fact that the weak interaction cannot distinguish between +w and -w, and ensures that there is an effective change of sign of the weak charge in every weak interaction involving a quark transition. Symmetry-breaking, therefore, takes place in transitions <u>within</u> quark generations ($u \leftrightarrow d$, $c \leftrightarrow s$) as well as in transitions <u>between</u> quark generations ($u/d \leftrightarrow c/s$); the weak interaction is so structured that it requires the existence of a unique <u>state</u> of weak charge whereas the mixing ensures that each quark generation requires the existence of weak charge of two <u>signs</u>. (The Cabibbo angle which determines the extent of the d, s mixing may perhaps be derived from the relative masses of the two quarks. The masses of d $\sim$ 313 MeV and s $\sim$ 489 MeV derived from the splitting of the lowest-lying baryon states N, Λ predict $\sin \theta \sim 0.22$ for equality between the masses of d_θ and s_θ .)

Now, if there were only one generation of quarks (u,d), there would be only one sign of weak charge and, hence, no need to introduce symmetry-breaking in the weak interaction; the sign of w could be fixed with that of s. However, with two quark generations, it is necessary to introduce two signs of weak charge and, at the same time, a symmetry-breaking mechanism such as nonconservation of parity. When a third generation of quarks (t,b) is introduced, it is necessary to create another degree of freedom for sign changes in the weak charge and this leads to further mixing involving a phase angle δ , which accommodates a violation of time symmetry, and a generalisation of the Cabibbo angle to allow for all possible mixings of the various quark doublets.

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 strange and charmed quarks are a natural pairing because they have the same value of w and because the weak decay of a charmed quark is predominantly via a strange quark mode.

Strangeness is also related to the concept of isospin, which groups together all particles with the same values of s and w, regardless of the value of e. Strong interactions conserve isospin but the weak decay of Λ breaks isospin symmetry along with strangeness. A similar concept of <u>weak</u> isospin applies in those weak interactions which involve such transitions as $u \leftrightarrow d$ or $c \leftrightarrow s$ where only the value of the electromagnetic charge component is exchanged.

From the decays of mesons and baryons we may also derive the charge structures of leptons: ν_e (w), e^- (−e+w), ν_μ (−w), μ^- (−e−w); the transition τ^- u → d (p → n) and the continued failure of experimental attempts at detecting ν_τ may also suggest τ^- (−e) and ν_τ (o). These structures immediately explain why $\bar{\nu}_e$ and ν_μ are not antiparticles of each other and why muons do not decay into electrons by the emission of photons. Also, baryon number is conserved because baryons are the only particles with s component, while lepton number is conserved because leptons are the only particles with ± w component but no s component (with the exception of K mesons, which may have ± 2w component, but which tend to behave, in most respects, like particles with zero weak charge component). The conservation laws for fundamental particles are, thus, essentially consequences of the separate conservation of e, s and w charges due to rotation asymmetry.

(7) Baryon Masses

The charge structures of particles suggest the origin of their particular masses, for there is evidence that these masses provide the missing energy due to the zero charges in the component quarks. In general, the rest masses of fundamental particles can be expressed 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, and it turns out that the masses of baryons and mesons can be expressed in terms of m_e/α by formulae which directly depend on the number of their zero charges. In effect, m_e/α may be taken as the mass-equivalent of each missing charge.

If we take the rest mass of the electron as fundamental and determined by the value of the electric charge, we may suppose that masses which originate in the strong coupling between quarks should take values related to the electron's mass by the term $1/\alpha$, which is the ratio of the strong and electromagnetic couplings. The value of the electron's classical radius ($e^2/m_e c^2$) is also fundamental for it is the Compton wavelength or maximum range for a virtual particle of mass m_e/α, and, by implication, of the mass-equivalent of a fundamental unit of charge.

The significance of this may be seen if we examine the mass of the pion. This is, at $2\ m_e/\alpha$, the lightest possible quark-antiquark combination. The Compton wavelength or maximum range which this produces is also the minimum distance of separation for an electron-position pair ($r_e/2$); it is also the minimum distance of separation for a nucleon-antinucleon or nucleon-nucleon pair bound by the strong interaction. All this evidence goes to suggest that the characteristic distance associated with a quark and presumably, therefore, with a fundamental unit of charge, is identical to that associated with the only stable charged lepton. Leptons and quarks are believed to be closely related; both types of particle are apparently point charges and lepton-antilepton pairs are produced through the decay of quark-antiquark or baryon-antibaryon combinations. There is thus every reason to suppose that they are related by a common value of size and that this size directly determines the mass associated with their unit charges.

The rest masses of baryons may be worked out from the formula

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

where n_o is the total number of zero charges in the components of a multiplet of multiplicity M , and M_o is the highest multiplicity in the relevant 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.

This formula arises from the charge independence of the strong interaction. This occurs because the electromagnetic charge is not confined within a particle, and so baryons may be organized into isospin multiplets, the components of which differ only in the values of their electromagnetic charges. The isospin multiplets represent a single particle in different states of electromagnetic charge; all states of one multiplicity exist simultaneously, and so the zero charges of all states must be accommodated in determining the mass of the particle. The zero charges of the Δ particles are simply those of all four states added together, but the four Δ states when excited have to be averaged between three Σ states, so each state represents an average 4/3 states, and the average charge has to be multiplied by 4/3, etc. The four Δ states are eventually excited to one Ω state, so each Ω state represents an average four states.

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)	uud, uud	<u>9</u>, 11, 13	3	2	13.5	13.4
	Λ	uds	<u>5</u>, 7	3	1	15	15.9
	Σ	dds, uds, uus	15, <u>17</u>, 19	3	3	17	17
	Ξ	dss, uss	11, <u>13</u>	3	2	19.5	19
Decuplet	Δ	ddd, udd, uud, uuu	<u>20</u>, 22, 24	4	4	20	18
	Σ^*	dds, uds, uus	<u>15</u>, 17, 19	4	3	20	20
	Ξ^*	dss, uss	<u>11</u>, 13	4	2	22	22
	Ω	sss	<u>6</u>	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^o 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. The theoretical masses of 20, 20, 22 and 24 m_e/α for Δ, Σ^*, Ξ^* and Ω may be compared to the experimental values 18, 20, 22 and 24. In fact, differences of $2m_e/\alpha$ can be expected at each level in the decuplet because the excitation from Δ^o to Σ^* to Ξ^* to Ω occurs due to the successive transitions of one d quark to one s quark with the net loss of two w charges, and this might suggest a theoretical Δ mass of 18 m_e/α. The difference between the two possible masses of the Δ particle is then accounted for by a high kinetic energy, of order $2m_e/\alpha$, which explains the particle's rapid decay. The full width of the Δ state at half maximum of energy is $\sim$ 120 MeV, which is of order $2m_e/\alpha$ and considerably greater than that for any other member of the decuplet. (The predicted values for Δ, Σ^*, Ξ^* and Ω become significantly closer to the more exact experimental values of 17.7, 19.8, 21.9 and 23.9 m_e/α with increasing mass, no doubt because of the fewer available states for the heavier particles.)

In the spin 1/2 baryon octet (M_o = 3), the minimum values for n_o predict masses of 13.5 m_e/α for N and 15 m_e/α for Λ which are comparable with the observed values of 13.4 and 15.9, especially as the observed mass of Λ may be affected by mixing with singlet Λ states. (The mixing occurs because SU(3) is not an exact symmetry, being broken by the negative w charge and greater mass of the strange quark.) In the next highest multiplet, Σ^o has the same quark structure (uds) as Λ, so assuming that the latter occupies the lowest value of n_o, we take the second lowest (n_o = 17) for Σ and predict a mass of 17 m_e/α, which is almost identical to the observed mass. Continuing with the second lowest value, n_o = 13, for Ξ, we predict a mass of 19.5 m_e/α where the observed value is 19. Here, again, we may have to take into account an expected mass difference of 2 m_e/α between Σ and Ξ.

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.

Masses for the three lightest quarks (u, d, s) may be derived, in the usual way, from the masses of baryons; the relative heaviness of the strange quark is responsible for the apparent symmetry-breaking effects of the so-called "semistrong" force. These masses are, of course, effective masses produced by the strong coupling which provides the minimum energy requirements due to the missing charges; the true quark masses may be as low as 0 for u and m_e for d (u being lower because of its average extra charge). The masses of the heavier quarks would then be minimum energy requirements due to some additional condition linked with grand unification.

(8) 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' singlet 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 $(n_o/M)(m_e/\alpha)$. 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

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

From our full analysis of weak interaction processes (in section 12), we can discover that decays of the form $K^+ \rightarrow \mu^+\nu_\mu$, $K^- \rightarrow \mu^-\bar{\nu}_\mu$ involve only $u \rightarrow c$ transitions and, hence, no Cabibbo mixing. All such decays, and those such as $\eta \rightarrow 3\pi$, which involve no quark transition at all, give decay products with the exact masses expected. From the decays of K mesons to $\mu\nu$ and $\pi\mu\nu$, we may fix the mass of μ (which originates only as a decay product of K or π) at $1.5\ m_e/\alpha$; this large mass is presumably responsible for the particle being unstable against decay to the lighter electron.

Other decay modes of K and that of π to $\mu\nu$ all involve quark transitions of the form $d \longleftrightarrow s$, which are complicated by Cabibbo mixing, and all give decay products with masses differing from the predicted values by $\pm \frac{1}{2}\ m_e/\alpha$. The mixing must certainly affect the mass of the final states.

(9) 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 intrinsically defined object.

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 could be said to show a highly 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, η mesons of respective mass 6, 7, 8 m_e/α, but the correspondences are not especially striking and arguments based on missing units of charge are no doubt insufficient to account for high energy states. The same could be said about comparisons between the baryon octets of spin 1/2, 3/2, 5/2 and 7/2, and combinations of the ground-state baryons with π mesons of respective mass 6, 8, 10 and 12 m_e/α. In all these cases, the missing charges may be presumed to make a <u>contribution</u> to the particles' masses, but much more information will be needed to complete the high energy picture.

(10) The electromagnetic interaction

The charge structures of mesons and baryons also emphasize the fundamental differences between the three nongravitational forces, which were introduced with the restrictions on s and w necessary to complete the quark system. Thus, e, which is unrestricted and which differs from s and w in being neither required to be present in a baryon state nor absent in a meson state, appears as an independent term, not bound to the particle in the same way as the other two charges. The two consequences of this are that the electromagnetic interaction is infinite in range and that the strong interaction has the property of isospin conservation or charge independence. (The electromagnetic interaction breaks isospin conservation because of the asymmetric grouping of electromagnetic charges in the quark system, but it is this same asymmetry which is responsible for the electromagnetic force remaining independent of the particle systems with which the charges are associated.)

(11) The strong interaction

Strong charges, unlike electromagnetic charges, are confined to the interiors of hadrons; baryons may be combined with mesons to produce other baryons and energy may, therefore, be transferred between the strong charges within baryons by the emission and absorption of mesons; the nonzero mass of the meson reduces the range of the interaction to that described by the Yukawa potential. The strong interaction, in addition, is generally closely associated with colour; in colour-anticolour combinations, such as mesons, the s component disappears; the absence of colour thus 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. Now, the confinement of quarks within "colourless" hadrons is actually a necessary consequence of a system whose purpose is to prevent the physical identification of the quaternions i, j, k. If the three quarks in a baryon were exactly coincident in space, then no force would be required to bind them together. However, if quantum mechanical or other considerations prevented exact coincidence, we would expect the quarks to have a binding force which would become greater with their respective separations from each other. The single strong charge would be exchanged between the three quarks to prevent the identification of any one quark by its colour, and even the notion of an instantaneous location for s would become impossible on account of the gauge invariance; the exchange particles, which would be massless for a renormalizable theory, would couple to each other because they would also be "coloured" and the interaction would lose strength at short distances. In other words, the "colour" force would have the precise characteristics of the strong interaction.

An examination of the quark tables suggests that for u, d, c, s the strong charge could move from the blue to the red quark by the transition from system (a) to system (b) and then from the red to the green quark by the further transition to system (c). This may, indeed, be the mechanism by which the strong interaction operates, and it may be of some significance that any transition to system (d) would require the additional exchange of a weak charge.

The colour force must, of course, be the one involving the transfer of mesons between baryons because it is the only mechanism for causing combinations of quarks and antiquarks, the process involved in the absorption and emission of mesons. The mesons must be assumed to carry gluons because of their colour-anticolour combinations. These must then be responsible for the strong interactions between the heavy particles. (Quantum field theory, of course, requires an attractive force between identical particles mediated by spin 0 mesons, as opposed to a repulsive force between identical particles mediated by spin 1 gluons.)

Since the strong interaction involves no symmetry breaking, the energy produced by this force is presumably the maximum available. This may explain why the energy of a pair of strong charges involved in the pion exchange interaction (which determine the maximum range of the strong force) is of the order required to confine two wavefunctions at a wavelength equivalent to the distance of separation ($\sim 2hc/r$), and why the interaction at energies corresponding to the mass of the proton (at which energies the quark-gluon interaction becomes effective) is characterized by approximately unit strength ($s^2 \sim \hbar c$), producing two or more strongly interacting particles with the same probability as one. Now, the strong interaction, as the force responsible for colour, must decrease at short distance (and high energies) and this must be so arranged as to be due to the coupling of gluons. At low energies s must be greater than e. This means that the size of the proton, pion, etc. must be so organised as to make the strong interaction greater in strength than the electromagnetic interaction by a value which is related to the fine structure constant ($\alpha = e^2/\hbar c$). In other words, it is because s must be greater than e to fulfil conditions imposed by a value of α which is less than unity that the quarks in a particle cannot be coincident in space as we would otherwise expect them to be. (Quarks need not be coincident in space, because space is an arbitrary factor in a system, but the minimum energy of a particle must be such that it would make the quarks coincident).

(12) The weak interaction

The weak interaction is more complicated than either the electromagnetic or strong interactions; it always involves fermions or particles with a weak charge component $\pm$w, produces changes in quark flavour, and is closely connected with the structure of leptons. The leptons are quarklike in that they are assumed to be point charges and are grouped with quarks into families (u, d, ν_e , e^-) (c, s, ν_μ, μ^-) (t, b, ν_τ, τ^-) which are distinguished by the respective values (+w, -w, 0) of their weak charge components. It is significant that the six red quarks with a weak charge in table (d) have identical charge structures to their respective leptons. As the strong interaction is characterised by its confinement of the strong charge within colourless hadrons, so the weak force is distinguished by its violations of parity conservation or time-reversal symmetry, brought about by exchanges in charges between particles; these again require the existence of intermediate particles. (The weak interaction produces the necessary <u>mechanism</u> for the sign reversal in individual weak charge components; the symmetry-breaking which this requires thus becomes the <u>characteristic</u> property of the weak interaction. Thus, since the weak interaction cannot distinguish between +w and -w, the <u>mechanism</u> by which this operates must be incorporated into the mathematical description of <u>every</u> weak interaction.)

Weak interactions between baryons involve a double exchange of quark flavour which leaves the total e and w components of the interacting particles exchanged (except for the sign of w); such typical interactions as

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

always involve <u>four</u> particles or virtual particles. Leptons are also paired off in weak interactions, muons and electrons being always accompanied by their appropriate antineutrinos when they are produced by the interaction of virtual baryon-antibaryon pairs; we may suppose that weak interactions involving leptons also take place through exchanges in quark flavour. A possible explanation of such processes may be that quarks and antiquarks are exchanged in such a way that the strong charge components disappear and each baryon or antibaryon undergoes a symmetry-breaking total charge conjugation in one of its component quarks. An interaction such as

$$\underset{udd}{n} + \underset{\overline{uud}}{\bar{p}} \longrightarrow \underset{dd\bar{d}}{e^-} + \underset{\overline{uuu}}{\bar{\nu}_e}$$

would leave e^- and ν_e with structures equivalent to the red d and u quarks in table (d). In addition to explaining why leptons are quarklike point-charges, such a process would provide a direct mechanism for the expected quark-lepton transition.

In fact, the weak interaction may be found to occur because the (d) system of quarks cannot be obtained in general from any of the others without an exchange of weak or electromagnetic charges. However, in certain circumstances, a particular quark <u>combination</u> in the (c) system may be identical to one in the (d) system and a transfer to that system may occur. Thus Λ decay may be represented as follows:

(c) system

<u>u</u>(R)	<u>u</u>(G)	<u>d</u>(B)
0	1	0
0	1	0
0	0	1

p + Λ

<u>u</u>(R)	<u>d</u>(G)	<u>s</u>(B)
0	0	0
0	1	0
0	0	-1

$\longrightarrow$ (d) system

<u>d</u>(B)	<u>d</u>(G)	<u>u</u>(R)
0	0	0
0	1	0
0	0	1

n + p

<u>d</u>(B)	<u>u</u>(G)	<u>u</u>(R)
0	1	0
0	1	0
0	0	-1

The (c) and (d) representations appear to be virtually identical, but, because these charge structures are indistinguishable from others in the (d) system, a significant exchange of e and w charges has taken place which involves a change in quark flavour. There is no mechanism for the quarks in the (d) system to exchange strong charges and, therefore, nothing to hold them together; it thus becomes possible for two particles to exchange quarks or charges. Since quarks must be bound, exchanges only occur where three-quark or quark-antiquark combinations are the product. In addition to the s $\longrightarrow$ d transition, the weak decay of Λ may be considered as a u(G) (c) $\longleftrightarrow$ d(G)d exchange occurring simultaneously with a u(G)(d) $\longleftrightarrow$ d(G)(c) exchange. (Other representations are possible, including an exchange of red and blue quarks or an exchange of all three quarks.)

All weak interactions may be represented as real or virtual interactions between a particle and an antiparticle (e.g. $\Lambda + \bar{n} \rightarrow p + \bar{p}$). Normally, the strong interaction will predominate within a quark combination, but if a particle is brought into close contact with an antiparticle, the strong charges may effectively cancel, allowing the weak interaction to take place; the necessity for close contact between interacting particles means that the intermediate exchange particles ($W^{\pm}$, Z^{0}) must be highly energetic or very massive. When the strong charges actually annihilate, the weak interaction becomes one of those in which leptons are produced. In the production of τ leptons it is necessary to annihilate weak charges as well as strong charges. In this case it seems that we need the energy to produce a real rather than a virtual baryon-antibaryon combination because the mass of the τ lepton approaches that of the lightest such combination ($p\bar{n}$).

The production of leptons may be studied in the process of neutron decay $n\bar{p} \longrightarrow e^{-}\bar{\nu}_{e}$. Assuming transfer to the (d) system, we can write the structures

u(B)	d(G)	d(R)	
1	0	-1	
0	1	0	for the neutron
0	0	1	

and

ū(B̄)	d̄(Ḡ)	ū(R̄)	
-1	0	0	
0	-1	0	for the antiproton .
0	0	-1	

Exchange of the neutron's u(B) for the antiproton's d̄(Ḡ) would result in d(R) (electron) for the neutron and ū(R̄)(anti-electron neutrino) for the antiproton. Alternatively, with w sign change, the products may be s(R) (muon) and c̄(R̄) (anti-muon neutrino). The leptons may be represented by single quark/antiquark structures because blue and green quarks of one flavour in the (d) system are identical in the absence of strong charges and so $dd\bar{d} \longrightarrow d$ and $uu\bar{u} \longrightarrow \bar{u}$. This suggests that leptons are indeed point-like structures; a united system of three quarks or antiquarks would require no strong force.

For Λ decay, the quark exchange occurs in such a way that the colour system may be restored by (d)$\longrightarrow$(c) reversal (which must occur if the strong charges remain), but, for the decay of the neutron, the effective exchange of colour as well as flavour means that the strong charges disappear and the colour system cannot be recovered.

If the exchange of quarks between particles is characteristic of the weak interaction, it becomes possible to explain the weak interaction via a neutral current. Thus, in addition to the direct exchange of $u\bar{u}$ or $W^{+} + W^{-}$ between an electron and a neutrino, a proton (uud) may exchange a u quark with an electron neutrino ($\bar{u}uu$) or a muon neutrino ($u\bar{u}c$) without any simultaneous change in their electromagnetic charges:

ν_μ	+	p	$\longrightarrow$	ν_μ	+	p
$cu\bar{u}$		uud		$cu\bar{u}$		uud
ν_e	+	p	$\longrightarrow$	ν_e	+	p
$uu\bar{u}$		uud		$uu\bar{u}$		uud

The comparable interaction representing neutron decay ($n\nu \longrightarrow pe$ or $p\mu$) may be represented as an exchange of the red quark of the neutrino with the red d quark of the neutron:

ν_μ	+	n	$\longrightarrow$	μ	+	p
$u\bar{u}c$		udd		$u\bar{u}s$		udu
ν_e	+	n	$\longrightarrow$	e	+	p
$u\bar{u}u$		udd		$u\bar{u}d$		uud

The weak interaction is distinct in a very important way from the strong and electromagnetic interactions in that it can <u>only</u> take place through the transfer of massive particles. Weak interactions take place through the exchange of quarks between particles such as baryons and leptons. The quarks have masses and the exchange particles are of the type $u\bar{d}$, $d\bar{u}$, $d\bar{d}$, etc. In quantum field terms, this becomes equivalent to the lowest energy or vacuum state acquiring a net weak charge, and it is significant in this context that only the weak charge may change sign and "create" net charge in the vacuum state by a violation of parity conservation or time-reversal symmetry. The electromagnetic charge may not and so is transmitted by a <u>massless</u> particle, associated with a noncharged vacuum state. (The necessity for massive intermediate particles also ensures that weak interactions <u>must</u> violate a symmetry.)

Nevertheless, the weak interaction is obviously of the same <u>form</u> as the electromagnetic interaction, which could be considered in principle, as transmitted by a particle of the form $u\bar{u}$ if $u = \nu_e$ has no intrinsic mass. In fact, of course, the situation is slightly more complicated, with γ and Z^0 being mixed states of B^0 and W^0, and massive $u\bar{u}$ bosons equivalent to $u\bar{d} + d\bar{u}$ being responsible for some of the interactions involving Z^0. However, the analysis does suggest that there must be an interaction involving a noncharged vacuum state, which, because it does not violate any symmetries, cannot be the weak interaction.

(13) Grand Unification

The fundamental equivalence of the three nongravitational forces is now a generally accepted principle, and it is believed that their apparent differences are due to a spontaneous symmetry breaking which results from the particles involved in the interactions assuming various masses; the forces remain symmetric in principle while an asymmetry is introduced by the quantum or energy state of the system. (Spontaneous symmetry breaking is a natural result of a quark theory based on unit charge; such a process may be expected to occur if the masses of particles result from the absence of charges whose presence would otherwise maintain symmetry.) Apparently, the strength of each interaction is not constant but depends on the energy with which the particles interact. With increasing energy the strong force grows weaker and the weak and electromagnetic forces grow stronger, the three converging at an energy of order 10^{14} - 10^{15} GeV.

According to the quantum theory of fields, any charge is screened and effectively reduced in value by the quanta of field radiation which surround it; since, the effect of this screening is reduced as we approach the charge, the effective value of the charge then increases. For the electromagnetic charge, determined solely by Coulomb forces, the effective value of the charge should simply increase with increasing energy of interaction. The behaviour of strong and weak charges is, however, more complicated. At low energies, the effective values of the strong and weak charges (or the coupling constants) are higher than the electromagnetic charge because their quanta of field radiation are also carriers of the charges and therefore sources of the radiation field, but the strong charge is greater than the weak charge because the strong force is carried by eight coloured gluons compared to only three intermediate bosons for the weak force. (This establishes the hierarchy $s > w > e$ for the charges at low energies; the relative strengths of the interactions, as determined by times, seem to be related to the number of symmetries broken by each force.) This effect is predominant for the strong force, but not for the weak, and so, at higher energies, s decreases while e and w increase. At grand unification energies, massive X bosons (which are carriers of all three charges) contribute to the effective charges in such a way that their coupling constants tend to equality.

The strong interaction has been accommodated via quantum field theory into a colour SU (3) group and the weak and electromagnetic interactions have been combined into an SU (2) x U (1) representation. In the latter case, scalar fields are introduced into the unified quantum field theory to produce the symmetry-breaking Higgs mechanism which assigns a net weak charge to the vacuum state; the self-couplings of the scalar fields determine the actual symmetry-breaking and, hence, the masses of the intermediate bosons.

Any grand unified theory must successfully incorporate the colour SU (3) and the electroweak SU (2) x U (1) groups into a single gauge group which imposes relations between the coupling constants for the individual interactions. The simplest combination, SU (3) x SU (2) x U (1), cannot be the unified gauge group (G) because it does not truly unify the weak and electromagnetic interactions, but the combination must be a subgroup of G. The smallest possible grand unification group containing SU (3) x SU (2) x U (1) turns out to be SU (5). (2) This seems to work reasonably well for left-handed fermion states if we assume neutrinos of zero mass, and fixes the ratio of electromagnetic to weak couplings ($\sin^2\theta_W$) at 0.375, but this is only valid at extremely high energies and Buras et al. have estimated that renormalization would reduce this value to ~ 0.2, in good agreement with experimental values. (3) Other grand unification theories, based on the left-right symmetric electroweak model of Pati and Salam (4), require the neutrino to have a finite mass.

The figure of 10^{14} - 10^{15} GeV for the energy of grand unification is obviously of some special significance, and, since it is only a few orders of magnitude below the Planck mass, it may be related to the size of the gravitational interaction. Here, we may return to our original concept of group symmetry between the parameters. If each parameter is numerically identical (in fundamental units) to its own inverse, then

$$Gm^2 = Q^2 = hc$$

where Q is the value of a charge at grand unification. If the grand unification mass (GUM) were equal to the Planck mass (or 2π x the Planck mass) we would have no problem in identifying Gm^2, Q^2 and hc, and deriving the true value for Planck's constant; but, even if the GUM were less, as seems likely, we could still identify the three quantities provided that c decreased with increasing electromagnetic charge. This is certainly possible according to group symmetry since c (the space / time ratio) is derived only from the equally fixed ratio between mass and charge, and e is the only charge which fulfils comparability with m, being unrestricted and infinite in range. It is probable that, under normal circumstances, the predominant value of c is determined by the existence of a predominant value of charge less than the fundamental charge Q. As the unit of charge is increased relative to mass, so the unit of time is increased relative to space.

If we assume that, at the energy of grand unification ($U = mc^2$), there is a value of the velocity of light (c_o) such that

$$Gm^2 = Q^2 = hc_o$$

we can calculate the expected value of Q^2 for any given GUM, for we can write

$$U = mc_o^{\,2} = mQ^4/h^2$$

and so $$Q^2 = GU^2h^4/Q^8$$

giving $$Q^{10} = GU^2h^4 \quad .$$

For $U = 10^{14} - 10^{15}$ GeV, this expression gives values of Q^2 in the expected range between e^2 and s^2 (and greater than w^2 as defined by $e^2/\sin^2\Theta_W$). For $U = 10^{15}$ GeV, we obtain $Q^2 = 3.2 \times 10^{-27}$ Nm^2, and for $U = 1.4 \times 10^{14}$ GeV, $Q^2 \sim 2\pi e^2$. In this case e^2 would be hc_o (unit strength for $c = c_o$, or $\hbar c$ at grand unification) and c_o would equal αc. Thus the maximum and minimum values of charge would be exactly defined by the maximum and minimum values for unit strength of interaction. The fine structure constant would give the ratio of the smallest unit coupling to the largest unit coupling. This possibility is certainly enhanced by the most recent estimates of Q^2 and U, based on extrapolation from low energy values, for these seem to suggest values in exactly this region. (It is significant that Q^2 is not particularly sensitive to the precise value of U, and is only affected to a relatively minor degree by changes in the form of its expression, e.g. $U = 2mc_o^{\,2}$ or $U = mQ^4/\hbar^2$. For $Q^2 = \hbar c_o$, which is possible on a priori grounds, we have to take $U \sim 5.5 \times 10^{15}$ GeV for $Q^2 \sim 2\pi e^2$; then $e^2 = \hbar c_o/2\pi$ and $c \sim 2\pi\alpha c_o$. This seems to be a less likely possibility.)

The probable existence of relatively simple relationships between the fundamental charges and α suggests that we are, at last, close to an independent derivation of the value of this constant. According to current theories, at grand unification energies, the masses of leptons become equal to those of the related quarks, i.e.

$$m_e \approx m_d, \quad m_\mu = m_s, \quad m_\tau = m_b$$

Buras et al. (3) have used input values of s^2, m_μ and m_τ to derive from the SU(5) scheme predicted values of m_s and m_b which are in good agreement with experimental results. Now, if our previous arguments are valid, we have independent reasons for fixing the values of m_μ and m_s with respect to m_e/α , and so, by assuming the value of α (and, hence of s^2), we can, in principle, derive the value of the GUM. Furthermore, our independent knowledge of m_d in terms of m_e/α (in this case compared to m_e) also suggests that it should be possible to use the GUM to derive an independent estimate of α. The value of m_b should follow from an exact knowledge of m_τ. (No reason is yet known for the mass of the charmed quark, but the masses of $c\bar{u}$ and $c\bar{d}$ are close to those of $p + \bar{p}$ and $p + \bar{n}$.)

In fact, the theory gives us a programme for working out the entire range of fundamental constants. First, we take three constants - say G, h and Q - to establish the numerical relationships between the four fundamental parameters. The values of these constants cannot be "explained" for they have no intrinsic meaning; they merely relate the units of entirely separate systems of measurement. From these, we can immediately work out the GUM, c_o and e; w follows from e if the gauge group is known. Then, using m_d/m_e and/or m_s/m_μ, we can derive α, s and c. (If e is not simply related to Q or c_o, it can still be worked out with s from a combination of m_d/m_e and m_s/m_μ.) This reduces the problem of deriving the fundamental constants relevant to particle physics to that of explaining m_e, the one constant which provides a scale for matter and gives physical meaning to all the others.

The three nongravitational interactions are believed to be mediated by spin 1 bosons. For a grand unification involving the SU(5) group, the gauge bosons for the first generation of quarks and leptons may be represented by

$$5 \times \bar{5} = 1 + 24$$

where $$24 = (8,\ 1) + (1,3) + (1,1) + (3,2) + (\bar{3},2).$$

The first term (8,1) represents the gluons, colour-anticolour combinations representing +s,0, -s charges; though there are nine such combinations, one is a colour singlet.

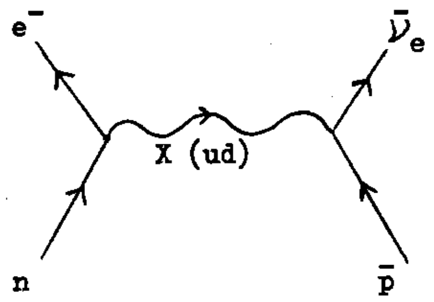
The terms (1,3) and (1,1) represent the exchange particles for the combined weak and electromagnetic interactions: W^+ (which is equivalent to $u\bar{d}$), W^- (equivalent to $d\bar{u}$) and Z^o, γ ($d\bar{d}$, $u\bar{u}$), all of which are colour singlets. (m_Z is greater than m_W because $m_{d\bar{d}} > m_{u\bar{d}} = m_{\bar{u}d}$.)

The last two terms have been thought to represent a particle carrying both colour and flavour which would be responsible for a decay of the proton. An interaction of the form

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

would be mediated by a particle $X = \bar{d}e^+$, equivalent to $\bar{d}\bar{d}$. However, an interaction of this form would violate the rotation asymmetry of charge, the conservation of charge in general and the conservation of strong charge in particular. It would only be possible if the identification of e, s, w charges (and even charge as distinct from mass) could be described as probabilistic, depending on the energy of interaction, and with the interactions inseparable at the GUM.

The facts, however, also support an entirely different explanation, for the particles ud (RG, BG, BR, GR, GB, RB) and their respective antiparticles could also be represented by (3,2) and ($\bar{3}$,2) and appear to be those exchanged when baryons decay to leptons. The Feynman diagram for this interaction could be represented thus:



The strong and weak (or electroweak) forces are in fact combined in the decay of virtual baryon-antibaryon pairs to leptons and antileptons, and the interaction also represents a quark - lepton transition. We may assume, therefore, that this interaction is the one mediated by the X particle and that ud is in fact that particle. The interaction does of course, include the ordinary weak $n\bar{p} \longrightarrow W^- \longrightarrow e^- + \bar{\nu}_e$ but it requires a strong charge combination first; m_X must be greater than m_W because, in this case, the strong charges must annihilate. All weak interactions involving bound quarks require the mediation of strong interactions; this is why Georgi and Glashow (2) found it impossible to find a gauge group which would unify electromagnetic and weak interactions independently of strong interactions.

The theory of grand unification suggests that, if measured under the same conditions, all three nongravitational forces would have the same size and identical characteristics. This does not mean that they are all aspects of the same force, for the three charges retain their individual identities as determined by rotation asymmetry. The forces are, of course, "unified" in the transitions mediated by the X particle, and in this sense the charges act as a true unit when their values are equalized, but there is no reason to suppose that charges of any one type can be at any time converted to charges of any other. Also, though the grand unification energy seems to be that at which gravitational forces become equalized with those due to charges (as required by the quaternion representation), there is no reason to suppose that these two very different types of force at any time lose any of their individual and distinctive properties.

(14) Quantum Mechanics : the Schrödinger Equation

Quantization of energy is a fundamental aspect of the group symmetry between the parameters. It is introduced with the numerical relation $mc^2 = h/t$ between mass and the inverse of time and it becomes a physical relationship in 4-vector systems due to the physical linking of mass and charge brought about by the quaternion representation. Whatever the actual value of h or $\hbar$ and whatever its relationships with other constants such as Q, c and α, it must be universal; there must be a single fundamental constant in all 4-vector systems which relates energy to the inverse of time. Simple physical considerations also suggest that quantization of energy in 4-vector systems would follow solely from the effect of field retardation. Ideally, 4-vector forces would all be of equal magnitude, all of inverse-square law form, and all transmitted by massless particles at the speed of light. Then, the energy of two typical charges of strength Q, separated by a distance r, would be given, in appropriate units, by Q^2/r, and the same expression would represent the energy due to the repulsion of the self-charge for a sphere of radius r uniformly charged over its surface. With the interaction travelling as a retarded wave at the velocity c, the energy would be related to the frequency of interaction (c/r) by a constant Q^2/c. For fundamental (equalized) charges at the energy of grand unification, there are reasons to believe that Q^2/c would be equal to h (or $\hbar$).

The ultimate derivation of both from the 4-vector system suggests that the mathematical and physical arguments for quantization are really equivalent, and the latter adds the requirement that the interactions must be "gauge invariant" for there can be no way of knowing the absolute phase or "position" of the retarded wave which carried the interaction. (This contributes to the "uncertainty" element in quantum mechanics.) The 4-vector system also unites the energy of a system with its momentum into a 4-momentum (p, E/c) and, if the energy is quantized, so must be the momentum. As the energy is quantized in units of inverse time, so the momentum (representing the 3-dimensional part of the 4-vector) is quantized in units of inverse distance:

$$E = h/t = \hbar\omega$$

$$p = h/r = \hbar k \quad .$$

Now, the systems of particles show that charges do not exist as isolated units but are found only in association with definite masses. Quantum mechanics originates in this situation. Particles are 4-vector systems and have an energy relating to their units of charge which is quantized; certain arrangements of the units involving zero charges are necessary to fulfil the requirements of the quaternion representation; however, to maintain the symmetry of unit charges 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 $p = \hbar k$. (This actually determines the behaviour of the subnuclear forces. For a system of bound quarks, for instance, the energy due to missing charges is mainly supplied by the energy due to a strong coupling via massless gluons at $s^2 \sim \hbar c$.)

As we have seen, quantized energy may be expressed in the form of retarded field or wave energy. Particles may therefore be expressed as wave motions with wave number k and angular frequency ω . The interpretation of the expressions for E and p in terms of gauge invariance then leads directly to Heisenberg's uncertainty principle and the probabilistic explanation of a particle's wave motion. And, since the mass energy of a particle is the quantity associated with the definition of its position (in fact with the definition of the particle as a <u>system</u>), we can derive the quantum mechanical equations of Schrödinger and Dirac by making certain assumptions about the nature of electrons and protons as "real" particles.

We assume, in fact, that the particle is a <u>classical</u> wave, which means that it can be described by a function of the form

$$\psi\ (x,\ t)\ \propto\ \cos\ (kx - \omega t + \alpha)$$
$$\text{or}\quad \psi\ (x,\ t)\ \propto\ \sin\ (kx - \omega t + \beta)$$

or some combination of such sines and cosines. By applying the correspondence principle that quantum mechanics must effect a smooth transition to classical conditions as quantum effects become insignificant, we find that this wavefunction must be constructed in such a way that the initial wavefunction at t = 0 determines the instantaneous wavefunction at all times.

The only way to incorporate this condition into an expression which describes the simple harmonic plane waves required for 4-vector field retardation is to combine sine and cosine terms into a single expression, which becomes

$$\psi_1\ (x,\ t) \propto \cos\ (kx - \omega t) + K_1 \sin\ (kx - \omega t)$$

for one direction, and

$$\psi_2\ (x,\ t) \propto \cos\ (kx + \omega t) + K_2 \sin\ (kx + \omega t)$$

for the other. To fulfil the requirement that ψ_1 and ψ_2 should be linearly independent at all times, it is necessary to equate the two constants K_1 and K_2 to $\pm$ i. The wavefunction for a particle then takes the form

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

where the constant A may be made unity by normalization. The form of this wavefunction, in fact, reflects the fact that the terms kx and ωt are imaginary quantities. Though x is real, k is imaginary because p is imaginary, and, though ω is real because E is real, t is, of course, imaginary. The imaginary nature of the wavefunction is, therefore, not a mathematical accident but a direct consequence of the physical nature of the quantities involved in it.

With the exact form of the wavefunction established, it is of course easy to derive the energy and momentum operators

$$E = i\hbar \frac{\partial}{\partial t}\quad ,\quad p = -i\hbar\frac{\partial}{\partial x}$$

and to substitute these into the classical expression

$$E = \frac{p^2}{2m} + V$$

to derive the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{\hbar^2}{2m}\quad \nabla^2\psi\ + V\,\psi\,.$$

(15) Bosons and Fermions

Now, as is well-known, the Schrödinger equation is not a 4-vector equation and is not covariant with respect to Lorentz transformations. A second derivative with respect to space is equated to a first derivative with respect to time. To introduce Lorentz invariance and the 4-vector space-time symmetry into the Schrödinger formulation it is necessary either to square the energy operator $i\hbar\,\partial/\partial t$ to make time a component of a 4-vector space or to take the square root of the momentum operator $(\hbar^2/2m)\,\nabla^2$ to make space a component of a 4-vector time. Because of the limitations imposed by the classical nature of the original wavefunction, these two processes are not identical and neither can be applied to all types of wavefunction. The first only applies to wavefunctions which are symmetric to a reversal in space and time coordinates and the second only to wavefunctions which are in the same sense antisymmetric. The fact that both types of wavefunction are possible is due to the classification of particles as fermions and bosons: fermions are described by the Dirac equation and bosons by the Klein-Gordon equation.

Fermions are particles with a total weak charge component of $\pm$ w. This weak charge is allowed to change sign to preserve the quark colour invariance and this is achieved by violation of parity conservation, the time-reversal symmetry being preserved; in terms of the CPT theorem the symmetry of CP is preserved at the expense of that of PT. Now, in order to introduce the real-imaginary space-time symmetry into the quantum mechanics of the fermion, it is more convenient to introduce space, which now has only one parity state but two mathematically indistinguishable solutions, rather than time, with its uniquely specified direction, as the imaginary component of the relationship. By restricting space to one physical state, we now allow <u>two</u> physical states to time. By application of the <u>group</u> symmetry, this mathematical device should make the mass of fermions appear to be imaginary, with both positive and negative solutions, and charge appear to be the real quantity with a single solution. Now, one of the remarkable facts about the Dirac theory is its prediction of negative mass-energy states; at the same time, the equations are worked out for particles with a single sign of charge. The theory has to restore the true status of mass and charge and remove the implied negative masses by the <u>ad hoc</u> introduction of antimatter or particles with the opposite sign of charge. Bosons, on the other hand, have 0 or $\pm$2w component of weak charge; this does not require any reversal in the sign of spatial coordinates or in the status of space and time when the imaginary space-time symmetry is introduced into the otherwise asymmetric wave equation. For bosons, space remains the real and time the imaginary parameter, mass-energy remains positive and the antiparticles are simply a result of CPT symmetry; with two equally possible charge states the PT symmetry is preserved.

The properties of bosons and fermions may thus be summarised as follows:

Bosons	Fermions
2 charge states	1 charge state
1 mass state	2 mass states
2 space states (or parity)	1 space state with 2 signs
1 time direction with 2 signs	2 time directions
C = PT preserved (wavefunctions symmetric)	C = PT violated (wavefunctions antisymmetric)
P = CT preserved	P = CT violated
T = CP preserved	T = CP preserved

(The CP and PT symmetries are violated in the <u>decay</u> of K^0 bosons, but parity is preserved.)

(16) The Dirac Equation

Since the Schrödinger equation applies to symmetric wavefunctions and positive energy states, it can be adapted to a 4-vector description of bosons without any modifications of the original wavefunction. The Klein-Gordon equation

$$-\hbar^2 \frac{\partial^2\psi}{\partial t^2} = -\hbar^2 c^2 \nabla^2 \psi + m^2 c^4 \psi$$

is easily derived by the substitution of the operator expressions for E and p into the relativistic 4-momentum equation

$$E^2 = p^2c^2 + m^2c^4.$$

However, for fermions, the original symmetric wavefunction is no longer valid and modifications have to be made both to the wavefunction and to the operators derived from it. This is because space, unlike time, has properties of rotation as well as of translation. In introducing space as a component of time, we also introduce a rotation term as part of the 4-momentum; this manifests itself as an intrinsic spin or angular momentum. Fermions have two directions of angular momentum, because this quantity depends only on time, and therefore have two wavefunctions. Combining these into one, using spinors, makes the total wavefunction antisymmetric with the spin responsible for the asymmetry. The intrinsic spin of fermion theory is the physical manifestation of the two physical states of time introduced with the original violation of space parity. Since this construction is not necessary to the derivation of the Klein-Gordon equation there is no need to postulate an intrinsic spin for bosons; however, particles of intrinsic spin equivalent to those of two fermions would have symmetric wavefunctions and obey the Klein-Gordon equation in the same way as spin 0 bosons.

The Dirac equation reduces the relativistic energy-momentum equation to linearity by defining operators α and β such that

$$i\hbar \frac{\partial\psi}{\partial t} = \alpha i\hbar c \nabla \psi + \beta mc^2 \psi .$$

This is only possible, of course, because negative energy states are permitted. αc represents a velocity term and β corresponds to $(1 - v^2/c^2)^{1/2}$. The new degrees of freedom represented by the introduction of these terms are those which describe the particle's spin, which in the case of fermions has half-integral values because there are two possible spin states.

The Dirac equation originates in the assumption that fermions have only one charge state; the antiparticles are not conjugate charge states but negative mass states. (Conjugate charge states are introduced ad hoc to explain the negative mass states.) The ultimate reason for this is the fact that one of the three charges which specifies the nature of the particle (w) must have its positive and negative values made indistinguishable when it is involved in an interaction. This means that all particles with unit values of weak charge must have individual wavefunctions which are antisymmetric and all weak interactions involving such particles must break a symmetry involving C (either CT or CP) even if such interactions do not involve a change of sign in the weak charge.

Now, weak interactions between fermions involve exchanges in weak charges; an exchange of space and time coordinates between two fermions would be equivalent to an exchange of charges and would, therefore also be equivalent to a weak interaction with violation of PT symmetry. Since the space and time coordinates of particles are determined by their wavefunctions, the total wavefunction of a group of fermions would be antisymmetric to an exchange of space and time coordinates between any two fermions. (For a group of bosons, of course, any exchange of coordinates would preserve PT symmetry and the total wavefunction would be unchanged in sign and, therefore, symmetric.) Particles with antisymmetric wavefunctions 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.

Now, the total wavefunction of a fermion is made up of the product of the orbital and spin wavefunctions; the spin wavefunction is the modification which makes the total wavefunction asymmetric. According to the relativistic theory of the fermion, the orbital angular momentum and spins of a moving particle are not separately conserved, but the component of spin in the direction of momentum is. Since the Dirac Theory assumes a single state of charge for fermions, CPT symmetry ensures a symmetry violation associated with the PT operator. At the same time, the preservation of CP symmetry means that fermions are reduced to a single state of parity. This means that fermions, or particles with unit weak charge components, may have only one helicity or spin orientation with respect to the direction of motion when $E/c \gg mc$, for the helicity, as determined by spin angular momentum/linear momentum, depends only on the state of parity. This orientation must be antiparallel or left-handed for particles of positive mass-energy because the spin must act in such a way as to reduce the effect of a linear momentum of the form $i\hbar\nabla$ when $(i\hbar c\nabla + mc^2)^2$ is effectively reduced to the $-\hbar^2c^2\nabla^2 + m^2c^4$ of the Klein-Gordon equation. Of course, the presence of a large mass m compared to a low energy E could also reduce this term, and so lead to the production of right-handed fermions, but these would, presumably, not take part in weak interactions, where parity conservation must be violated, and for massless neutrinos right-handed helicity states would be strictly forbidden. (The evidence seems to suggest that right-handed fermions have 0 or $\pm 2w$ weak charge, and so the additional mass-energy required for their production presumably comes from the suppression of a unit of weak charge.)

(17) The Quantum Theory of Fields

The quantum mechanical formulations of Schrödinger and Dirac are semiclassical because they regard particles as their fundamental objects and treat particle waves as nonquantum, but 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 avoids the problems produced by the existence of two incompatible covariant one-particle equations by making all particle states positive energy states and by introducing antiparticles as charge-conjugates to all particles. Thus, Dirac's relativistic theory of the fermion may be 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 postulate an unobservable sea of negative energy states. Quantum field theory thus restores the true status of the parameters with the "second" quantization and emphasizes the origin of quantization in the effects of field retardation.

(18) Thermodynamics

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 4-vector properties of systems containing 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 large-scale systems 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.

Another way of looking at the second law is to suppose that, in quantum mechanical processes, some energy becomes unavailable because of the extra uncertainty produced by the <u>retardation</u> which determines the quantum mechanical conditions; in any quantum mechanical interaction, there will be an increase in the effect of retardation. Because time flow is unidirectional, all field processes due to charges are retarded by an amount depending on the velocity of light; at the quantum mechanical level, this effect will produce a greater degree of uncertainty in the space-time distribution of energy and momentum, and hence a greater degree of disorder or entropy in the system. The second law of thermodynamics and the irreversible increase in entropy due to natural processes thus become associated with the unidirectionality of time, and the direction of time becomes associated with the progression of a sequence of irreversible dynamic events.

The increase of entropy defines the "thermodynamic" arrow of time; retardation is also responsible for the "cosmological" and "historical" arrows. In the first case, redshift always increases with the distance of the source because more information is lost by retardation when light from distant sources reaches the observer. In the second case, retardation creates a partial record of the past but none of the future; we gain information about the past at the expense of information about the present, or we gain information about the remote past at the expense of information about the recent past. Thus, the irreversible increase of entropy does not represent a total loss of information or a tendency to disorder in the universe as a whole; it merely represents a loss of information or a tendency to disorder within any defined system. In the universe as a whole, there is an exchange rather than a loss of information; the historical arrow replaces information about the microsystem with information about the macrosystem; the actual systems and their boundaries are arbitrary and defined by the observer. Energy which becomes unavailable for work within any defined system can be made available for work outside it. The second law of thermodynamics, in effect, states only that the parameters of physical systems cannot be fixed and that all systems have a natural tendency to change from one state to another.

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 true statement of the third law of thermodynamics.

(19) The Prevalence of Matter over Antimatter

All weak interactions violate either conservation of parity or time-reversal symmetry. In fact, the laws of physics are so structured as to determine that space parity is violated in preference to time symmetry whenever possible, or that time symmetry is violated only because of extra terms in the weak coupling matrix produced by the fifth and sixth quarks; the very high energies needed to produce such objects, and their consequent rarity, is a direct expression of this preference. The preference occurs because time is an imaginary parameter and it is inconvenient to have to express a preference between two imaginary directions of time symmetry of which we have no direct knowledge. Since the violation of time symmetry is known only through its effect on space, and since the space of a system is arbitrary, it is automatically chosen so as to minimize the effects of any violation in time symmetry. Thus the violation of the symmetry of time reversal occurs only when it is impossible to invoke the alternative violation of the symmetry of space reflection to accommodate an effective change in the sign of weak charge. Nevertheless, <u>time itself</u> is unidirectional and to express this in the laws of physics while preserving, except in a few cases, the equivalence of the two directions of time symmetry, it would be necessary to assume the existence of a dominant state of space parity or of charge. In fact, since space is the only parameter of direct measurement, we retain the two signs of parity and, to incorporate the unidirectionality of time into the laws of physics, we create a preference in nature towards a single state of charge or a preference between matter and antimatter. This preference is broken only to the extent that the violation of time-reversal symmetry in certain interactions breaks the equivalence between the state of charge and the direction of time. Again, the Dirac theory offers an interesting illustration in the case of fermions. In this theory, which makes time temporarily take over the role of space, there is a prevalence of positive mass-energy states over negative mass-energy states; reversing the mathematical convention to remove the unwanted "negative" energy, we arrive at the true position of a prevalence of matter over antimatter.

(20) A Weak Version of General Relativity

The derivation of the laws of classical mechanics in section (3) left open the question of how far the 4-vector system and the curl terms which are associated with electromagnetic theory may be applied in the theory of gravitation. It would be possible to assume that 4-vector space-time is intrinsic to gravity, giving rise to curl terms which differ from those of electromagnetic theory because the gravitational field acts as a generator of itself. With the proper mathematical development this information could presumably be formulated into a theory which would be in every respect identical to Einstein's General Theory of Relativity in which gravitational effects are ascribed to a curvature in 4-dimensional space-time.

There are, however, certain problems in formulating such a "strong" theory, mainly on account of the fact that the gravitational field as a self-generator would violate the conservation of energy, while the 4-vector space-time, if intrinsic to gravitation, would cause problems with quantum gravity. The problems are fundamental and seem to originate with the introduction of the 4-vector system. Nevertheless, it is possible to derive a perfectly coherent theory with a full description of 4-vector and curl-induced effects provided that these are not considered as intrinsic to gravitation. Because this theory is in every way parallel to the strong theory, with the exception of the self-generating effects produced by the gravitational field, it may be described as the "weak" version of general relativity.

The origin of this theory lies in the initial assumptions which are necessary for the foundations of classical mechanics and electromagnetic theory. These determine that classical mechanics is not intrinsically a 4-vector system. Now, the difference between classical mechanics (i.e. the classical theory of gravity) and electromagnetic theory with respect to the 4-vector system derives from the fact that mass is a continuum present in all systems but it is possible to define a system without charge. The systems of classical mechanics are those defined to be without charge. Charges are by definition localized units, their actions taking place over a finite time, but masses ought to present a true continuum of interaction, suggesting that the gravitational action must be instantaneous.

Now, the 4-vector system makes time part of the space vector in addition to creating a numerical relation between space and time units. With the absence of charges within a system, the necessary connection between space and time is broken, though not the numerical relationship between their units. The actual, as opposed to numerical, space/time ratio might be expected to be infinite, and, since only 4-vector dynamics demands the interconversion between mass and energy, the actual, as opposed to numerical, mass produced by gravitational energy would then be zero. (If $e = o$; $m/e = r/t = c = \infty$. $E = mc^2$; $m = o$.) The potential energy of gravitation would still be numerically equal to mc^2, but it would not take part in energy exchanges. (This would not, of course, cause problems with conservation of energy, since energy is never actually exchanged between gravitational and electromagnetic systems; though these are frequently coupled in physical processes, the numerical equations $Gm^2/r = Q^2/r = mc^2$ actually require the independent conservation of gravitational and electromagnetic energies. The condition would, however, ensure energy and mass conservation in all physical systems, even in black holes.)

The rejection of the 4-vector description for purely mechanical systems leads us on to the question of the status of the strong theory of general relativity, which is founded on an extension of this description. The strong theory is concerned with establishing a 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 transformations 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 associated with 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.

Without this theoretical justification, the case for the strong theory as the exclusive theory of gravitation must rest on the experimental evidence. The gravitational redshift and bending of light, of course, present no problem, even to classical theory, for light is a 4-vector system and the effects are easily derived from a combination of 4-vector dynamics (or special relativity) and the equivalence principle. (5) (The redshift can be derived even more simply from Newtonian mechanics combined with $E = mc^2$ and $E = h\nu$.) Thus, only the perihelion precession of the planets stands as evidence for the need to introduce a curved space-time and a self-generating field into gravitational theory.

Nevertheless, it is probable that this phenomenon is not due to gravity at all for it seems possible to derive an entirely different explanation for the <u>inertial forces</u> which are responsible for perihelion precession. This is that the space-time actually measured in astronomical observations is operational (4-vector) and not identical to the absolute (3-vector) 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 a 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, there is not a single reason, theoretical or experimental, for the introduction a priori of a curved space-time, representing a self-generating gravitational field, into the theory of gravitation. The weak theory suggests that the 4-vector and curl-induced effects are similar in most respects to those predicted by the strong theory, but that, in this case, they are the extrinsic effects produced by the operational use of a 4-vector coordinate system. (The natural explanation of inertial forces and Mach's principle that this entails will be left to Section 22).

This does not mean that the geometrical methods introduced by the strong theory are not of value - they may provide the only kind of general method by which interacting gravitational and electromagnetic systems may be studied; but it does mean that they do not provide a true description of space and time. The strong theory is, in fact, severely limited by its violation of the conservation laws of mass and energy; if we are to use its geometrical methods, we must be careful to take this into account. We are, of course, entitled to use any <u>geometry</u> which may be convenient in individual cases, but the four-dimensional Euclidean metric seems to be the only true description of the <u>space and time</u> of the universe.

(21) Gravitational Waves and Gravitons

A weak version of general relativity will require a reinterpretation of the gravitational waves and gravitons predicted by the strong theory. In the weak theory, the gravitational field is linear, like the electromagnetic field, and the free-space equations involving ω, the rotational term representing inertial forces, will simply be the analogues of Maxwell's equations for electromagnetic fields:

$$\nabla \cdot g = 0 \qquad \nabla \times g = -\dot{\omega}$$
$$\nabla \cdot \omega = 0 \qquad \nabla \times \omega = \dot{g}/c^2 \quad .$$

Since the inertial forces are fictitious, the inertial frame of reference, with $\omega = 0$, will describe a Newtonian system, with infinite speed of interaction; all calculations involving energy must be referred to this frame. In noninertial frames, g and ω will appear to satisfy the equations $\Box^2 g = 0$ and $\Box^2 \omega = 0$, which predict the existence of gravitational (or, rather, inertial) waves, the quantum of which will be the graviton.

However, the graviton will not be a quantum of gravitational energy, which interacts with the gravitational field itself, because the inertial forces are not true carriers of gravitational energy. The equations for $\nabla \times g$ and $\nabla \times \omega$ do not involve the gravitational field as such; they merely express the numerical value of the inertial force correction. In fact, the gravitational field remains linear (and is, therefore, not a source of itself) precisely because these curl terms and the related 4-vector effects are imposed by the use of electromagnetic space-time and not by the intrinsic nature of gravitational energy. The true energy balance can only be worked out in an inertial frame.

The main problems of unified field theory and quantum gravity appear to derive from the supposed nonlinearity of the gravitational field. Without this effect, it should be possible to develop a rational theory in both cases; however, in the latter case, the theory will be concerned with quantizing the fictitious inertial field rather than the real field due to gravitation. It is important however, to note that the weak or "inertial" theory predicts the same "gravitational wave" effects (and periastron precession) as the strong theory in such systems as the binary pulsar PSR 1913 + 16 is assumed to be, since these effects are not due to field nonlinearity. (There is, in fact, no experimental evidence from any observational test for a nonlinearity of the gravitational field, and there is, consequently, no reason to believe that objects such as Cygnus X-1, which have been claimed as "black holes" or field singularities, are anything other than highly dense neutron stars.)

It is also significant that the main observational test of general relativity, the planetary perihelion precession, cannot be predicted on the basis of gravitational waves or gravitational field retardation because the analyses cannot be applied to direct transit between the sun and the orbiting planet, i.e. in the direction of the real gravitational field; it can only be applied in directions associated with the fictitious inertial forces. Hence, it is only the latter which can be associated with retardation at the velocity of light, and only they can be responsible in this way for the perihelion precession.

(22) Mach's Principle and the Hubble Universe

The introduction of a non-self-generating gravitational field leads, perhaps inevitably, to a nonexpansionary cosmology. Though this goes against the grain of current cosmological theories, it is perhaps worth exploring just how much the evidence requires an actual expansion of the universe. If expansion turns out to be correct after all, then almost all the arguments used here could be easily adapted to an expansionary model.

An important question to be decided in the formulation of a fundamental cosmology concerns the exact status of Mach's principle, which attributes the inertia of matter and the existence of inertial forces to interactions involving all the matter in the universe. Now, we have found that inertia is 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 theories with intrinsic 4-vector space-time, if we could remove from the latter the supposed nonlinear effects of gravity.

Using this theory of inertial forces, it becomes possible to develop a version of Mach's Principle in which such forces really are fictitious. We suppose that 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. (6) Taking ρ_u as the mean density of the universe, we may define a characteristic radius r_u (the Hubble radius) at which K could be made equal to unity.

Now, within a radius r, the measured force of inertia on a mass m_1 would be identical to the fictitious acceleration (a) - dependent inertial force produced by the enclosed matter (m_2)

$$F = \frac{Gm_1m_2}{c^2r}\, a\, \phi = \frac{4}{3}\pi Gm_1\rho \frac{r^2}{c^2}\, a\, \phi .$$

(ϕ, the angular dependence, may be assumed to be unity for the approximate calculation.) (6) 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 $m_u = (4\pi/3)\, \rho_u r_u^3$. For K to be made equal to unity, the inertial force of repulsion due to all the matter in r_u would be equal to the gravitational attraction, and bodies at the Hubble radius would have an apparent outward velocity c and acceleration c/t_u, where t_u is a time of interaction r_u/c.

A body inside the Hubble radius would be subjected to a fictitious inertial force equivalent to the gravitational effect of all the bodies within r_u. Thus a body of mass m at distance r from an observer at the centre of a sphere defined by the

Hubble radius would be seen to experience an inertial force

$$\frac{Gmm_u a}{c^2 r} = \frac{Gmm_u}{r_u^2}$$

and hence an apparent acceleration $a = c^2 r/r_u^2$. Writing $a = dv/dt$, where v is the apparent velocity, we can say that

$$\frac{dv}{dt} = \frac{dv}{dr}\,\frac{dr}{dt} = \frac{c^2 r}{r_u^2}$$

and so $v^2/2 = c^2 r^2/2\, r_u^2$ and $v = cr/r_u$. This means that, to an observer at the centre of a sphere defined by the Hubble radius (a Hubble "universe"), each body at a distance r would appear to be moving away at a velocity proportional to this distance. Any actual velocity of source would not alter the measured value of c, but, since v is a fictitious velocity due to the choice of reference frame, the relationship between v and r would also produce a slowing down of the light from the body, with a decrease in frequency which would become progressively greater with distance.

This "redshift" (which, incidentally, explains Olbers' paradox) may be considered as 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 time of interaction (r_u/c). The uncertainty is a fundamental component of matter and ensures that space always has a mean temperature whose precise value is related to the densities of matter and radiation in the universe.

That the redshift is caused by a fictitious velocity and not by a true expansion may be suggested by the fact that a body at r_u would still have an acceleration given by c^2/r_u even though its velocity of c could not be exceeded. Now, if we take

$$v^2 = \left(\frac{dr}{dt}\right)^2 = \frac{c^2 r^2}{r_u^2},$$

and substitute Gm_u for $c^2 r_u$, we obtain

$$\left(\frac{dr}{dt}\right)^2 = \frac{Gm_u r^2}{r_u^3} = \frac{4\pi}{3} G \rho_u r^2,$$

from which

$$\frac{1}{r^2 c^2}\left(\frac{dr}{dt}\right)^2 = \frac{4\pi}{3}\,\frac{G\rho_u}{c^2}.$$

This is formally similar to the equation for an expanding universe, with the terms for curvature k/rc^2 and cosmic repulsion Λ/c^2 automatically equated to zero. The extremely low observed values for these terms are unexplained in the expansionary model but, in a nonexpansionary cosmology, there is simply no reason for their introduction.

The nonexpansionary model also requires a new interpretation of the Hubble "universe" In this case, the Hubble universe is simply a sphere surrounding each body in space, of such magnitude that the inertia of the body can be ascribed to the inertial reaction of all the matter within it. Any matter outside the Hubble radius would not contribute to the inertial reaction because its apparent velocity would be greater than c, but there would still be matter outside the Hubble radius and this would contribute to the Hubble "universe" surrounding some other body. The size of this "universe" would vary with variations in the density of matter surrounding the body.

(23) The Mass of the Electron

The properties of the Hubble universe are determined not only by its overall density, but also by the relative populations or densities of all the major groups of particles within it. The most important of these are protons, neutrons, electrons, neutrinos and photons.

Since protons are the only stable heavy particles, most of the matter in the Hubble universe should be composed of protons; but, since protons exist in isospin symmetry with the only slightly heavier neutrons, and since these two particles are able to transmute into each other via weak interactions involving stable leptons, there may also be a significant number of neutrons. Now, the universe as a whole must have a positive baryon number because of the asymmetry between matter and anti-matter. This means that the universe has a preference for a single state of strong charge, but this has no large scale significance because the strong charge is confined. However, there can be no reason for a preference for a single state of the unconfined electromagnetic charge, and, because electrons are the only stable negatively charged particles, the number of electrons must be approximately equal to the number of protons. The number of neutrinos is a more problematical question since the state of the w charge is difficult to determine, but it must be at least as great as the number of neutrons to maintain a fixed proton/neutron ratio, and is possibly far greater. Of the exchange particles, only the photons should be significant on the large scale since the electromagnetic force is the only large-scale nongravitational interaction. The relative populations of all these groups of particles are determined to some extent by the fixed value of the mass of the electron.

The electron is the only stable point-charge particle which has a nonzero rest mass. It has a mass, unlike the neutrino, because it is a carrier of the infinite-range electromagnetic charge; the correspondences between the electron and the d quark and the electron neutrino and the u quark also suggest that $m_e > m_\nu$ because $m_d > m_u$. Now, because a point-charge has no identifiable structure and is an intrinsically elementary object, this suggests that the electron mass is a fundamental quantity and provides a minimum mass for a point-charge with an electromagnetic component, whether it exists as an isolated lepton or as a bound quark. The electron mass is also related via the constants e and c to a characteristic distance described as the classical radius ($r_e = e^2/m_e c^2$). This distance seems to be characteristic of other point-charges and it is probable that it in some sense <u>determines</u> the value of m_e.

Since the electron appears to be both elementary and stable, we may suppose that the m_e-r_e relationship may be determined in some way by the application of Mach's Principle. If we assume that the inertial electron mass is a fundamental quantity and cannot be explained, like m_μ, by some secondary derivation, then it must be equal to that which would appear to be "caused" by the inertial reaction of the other matter within the Hubble radius. In this case, the external gravitational field of the electron Gm_e/r_e^2 would become equal to the inertial reaction on unit mass (i.e. Gm_u/r_u^2). The apparent inertial reaction or repulsive force due to the other matter within r_u would exactly equal the attractive force of gravitation. With the other relationships $m_e c^2 = e^2/r_e$ and $Gm_u = r_u c^2$ also known, we could, therefore, derive individual values for the electron mass and classical radius for any given mass or radius of the Hubble universe.

This argument could not, of course apply to massless neutrinos, nor to composite particles with "secondary" masses such as protons, but it could apply, in principle, to any isolated unit charge, because an <u>isolated</u> charge would necessarily take on the characteristics associated with an unrestricted interaction, i.e. a coupling strength equal to e and an interaction speed equal to c.

The homogeneity and isotropy of the universe can be seen as the result of the application of Mach's Principle to the electron. Each electron is surrounded by a Hubble universe in which the relation $Gm_e/r_e^2 = Gm_u/r_u^2$ is valid, and, since electrons are universal components of matter, the universe as a whole must be organized so as to create an average density even more uniform than the distribution of electrons. Of course, because of the great scale of r_u and m_u compared with m_e and r_e, and because of the finite time of interaction, considerable variation in local distribution is allowed, but the various processes of the universe take place in such a way as to maintain the general equilibrium.

A homogeneous and isotropic universe will also be expected if the universal background radiation is assigned to the de-energising of photons, since this is another effect due to the inertial reaction of the matter within the Hubble radius. In this case the contributions due to the inertial reactions averaged out for many electrons combine to produce a background radiation equivalent to a redshift at velocities asymptotically approaching the velocity of light. A background radiation of this form would be remarkably close to defining the absolute standard of rest demanded by classical Newtonian dynamics. The background radiation may even be responsible for maintaining the overall electron/photon ratio. If the electron mass is equal to that which would appear to be caused by the inertial reaction of the matter within the Hubble radius, and if this reaction actually produces the de-energising of photons which causes the background radiation, then it may be that we can equate the energy density due to the electrons ($n_e m_e c^2$) with that of the photons ($n_\gamma kT$) due to the background temperature T. With the relation $n_\gamma kT = aT^4$ already known from Stefan's Law (where a is the radiation constant), we can derive both n_γ/n_e and T. With the number of electrons assumed equal to the number of protons ($n_e \sim 10^{80}$), the relation $aT^4 = n_e m_e c^2$ gives values of T^4 of the right order to accord with a measured background temperature of 2.76K, and using this value of T, we find that $n_\gamma/n_e \sim 2 \times 10^9$, which is in good agreement with observed values.

The temperature of the background radiation may also be said to determine the proton/neutron ratio. Thus, with $n_p = n_e$, the amount of thermal energy available to each proton is equal to $n_\gamma kT/n_e$, and, therefore, to $m_e c^2$. With the neutron-proton mass difference at $2.3m_e$, and the relative abundances determined by the Boltzmann factor $\sim \exp(-2.3)$, the neutron/proton populations are in the approximate ratio 0.1/1. (This neutron/proton ratio fixes the hydrogen/helium mass ratio of the universe at ~ 4.)

Now, if $Gm_e/r_e^2 = Gm_u/r_u^2$, then $m_u/m_e = (r_u/r_e)^2$ and $m_u r_u / m_e r_e = (r_u/r_e)^3$. According to current estimates, $r_u \sim 1.5 \times 10^{26}$m and $m_u \sim 10^{53}$kg; this requires values of $\sim 10^{124}$ for $m_u r_u/m_e r_e$, 10^{41} for r_u/r_e and 10^{83} for m_u/m_e. It has been frequently noticed that many of the dimensionless quantities which are important in cosmology occur in powers of order 10^{20}. Assuming one ratio such as m_u/m_e or r_u/r_e, some of these apparent coincidences may now find their explanation. Thus, the ratio of the Hubble time (r_u/c) to a characteristic nuclear time ($h/m_p c^2 \sim r_e/c$) is of order 10^{40} because it is close to r_u/r_e, and the number of charged particles in the universe ($\sim 2m_u/m_p$) is of order 10^{80} because it is approximately $10^{-3} \times m_u/m_e$. Using the further relation $Gm_u = r_u c^2$, expressing the mass-space relationship or conservation of energy for the Hubble universe, it is possible to show that the inverse of the gravitational fine structure constant, $\alpha_G^{-1} = \hbar c/Gm_p^2$, is of order 10^{40}, and the ratio of the nuclear time to the Planck time, $(G\hbar/c^5)^{\frac{1}{2}}$, is of order 10^{20}. The derivation of α_G is particularly important, for astrophysicists have shown (7) that this constant determines the number of particles in a star, the number of stars in a galaxy and hence the number of galaxies in the Hubble universe; it is also responsible for the lifetime of a typical star being of the same order as the Hubble time. Thus, without regard to the details of any particular cosmology, the large-scale structural features of the universe are determined by the fixing of the value of a single fundamental quantity—m_e, m_u or r_u; but, if this quantity can be shown to be invariable, then the universe cannot possibly be in a state of expansion.

Without the concept of physical expansion many of the current problems of cosmology disappear. We can thus explain the formation of stars and galaxies as a process of local gravitational condensation which occurs continually and which alternates with disruption and dispersal of the component particles brought about by the weak interaction, the whole process having a time period comparable to the Hubble time. There is no reason to suppose that this process had any definite beginning or will have any definite end. With the lifetimes of stars and galaxies only approximately related to the Hubble time, their ages will show statistical variations and it is possible that some of these objects may actually be older than r_u/c.

One question then remains to be answered: is there any physical argument which will fix absolutely the size of m_e or m_u? An independent relationship between e and m_e or m_e and m_u would complete the derivation of the fundamental constants and would finally explain why gravity is so much weaker than all the other forces. The best possibility for such a relationship would seem to lie in the derivation of the electron's mass in quantum electrodynamics by renormalization. Here, we may adapt to our purposes an argument of Browne (8), according to which the maximum self-energy of the electron due to the emission and absorption of photons would be equal to the particle's rest energy ($m_e c^2$) if the maximum intermediate state energy were of order exp $(2\pi/3\alpha)$ in units of the rest energy. This number is of order 10^{124} and remarkably close to the accepted value for the ratio Gm_u^2/e^2. Since e^2 represents the rest energy of the electron for a unit classical radius, Gm_u^2 could represent the maximum available energy over the same unit of distance; the ratio Gm_u^2/e^2 is also equivalent to $m_u c^2 r_u/e^2$, which is the ratio of the maximum amount of energy available within the Hubble radius to the minimum energy due to the self-charge of the electron taken over the whole Hubble universe. Since Johnson et al. (9) and Maris et al. (10) have shown that a nonperturbative calculation of the self-energy may be made assuming a zero baremass for the electron, we may suppose that, if m_e is directly determined by this ratio, $m_u r_u$ would then be derived from the fine structure constant. (A relation between electromagnetism and gravity involving an exponential of the fine structure constant has long been suspected.) If this argument, derived from a quantum electrodynamics, is valid, we now have four equations relating m_e, r_e, m_u and r_u which between them determine that the value established for each of these quantities is unique.

The four equations suggest the following explicit expressions:

$$m_u \sim \left(\frac{\hbar c\alpha}{G}\right)^{\frac{1}{2}} \exp \frac{\pi}{3\alpha}$$

$$r_u \sim \frac{G}{c^2} \left(\frac{\hbar c\alpha}{G}\right)^{\frac{1}{2}} \exp \frac{\pi}{3\alpha}$$

$$m_e \sim \left(\frac{\hbar c\alpha}{G}\right)^{\frac{1}{2}} \exp \frac{-\pi}{9\alpha}$$

$$r_e \sim \frac{\hbar\alpha}{c} \left(\frac{G}{\hbar c\alpha}\right)^{\frac{1}{2}} \exp \frac{\pi}{9\alpha}$$

With $1/\alpha = 137.04$ and exp $(\pi/3\alpha) = 2 \times 10^{62}$, they give $\sim 10^{53}$ kg, 10^{26}m, 10^{-30} kg and 10^{-15}m for the respective values of m_u, r_u, m_e and r_e, which are all well within the expected orders of magnitude.

The derivation of the mass of the electron via renormalization over the Hubble universe suggests that the quantum state of the vacuum is such that its total effect is the production of this mass. The electron mass is also fixed at that which would appear to be due to the inertial repulsion of all the matter within the Hubble radius. For the simultaneous application of these two conditions, the gravitational attraction due to the quantum state of the vacuum would have to be numerically equal to the total inertial repulsion. In other words the exact value of the sum of the two forces (the so-called "cosmic repulsion") would always be zero.

(24) 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. The quaternion representation of mass-charge thus becomes symmetrical with a 4-vector representation of space-time. The relationship between mass and charge within the quaternion fixes the relationships between all the parameters in the group, and these relationships lead to the establishment of the laws of classical mechanics and electromagnetic theory. Classical mechanics defines conditions which exclude charge from a system, while classical electromagnetic theory, special relativity, quantum mechanics, relativistic quantum mechanics and the quantum theory of fields represent the successive and increasingly complex attempts at applying 4-vector space-time to systems involving charge. Thermodynamics may be seen as a particular application of quantum mechanics. The nature of classical mechanics leads to certain applications in cosmology.

A further symmetry, derived essentially from a comparison 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, and it enables us to find relationships between some of the fundamental constants; others may be derived from cosmology. With all these derivations established without the use of phenomenological assumptions, there is sufficient evidence to suggest that the original source for both the laws of physics and the fundamental particles is to be found in the exact properties of the fundamental parameters and, hence, in the actual process of measurement.

References

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

(2) H.Georgi and S.L. Glashow. Phys.Rev.Lett. 32, 438, 1974.

(3) A.J. Buras et al. Nucl.Phys. 135B, 66, 1978.

(4) J.C. Pati and A. Salam. Phys.Rev. D10, 275, 1974.

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

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

(7) P.C.W. Davies, "The Accidental Universe", Cambridge 1982; and references therein.

(8) P.F. Browne. Intnl. J. Theoret. Phys. 15, 73, 1976; Found. Phys. 7, 165, 1977.

(9) K.Johnson, M.Baker and R.S. Willey. Phys Rev.Lett. 11,518, 1963.

(10) Th.A.J.Maris, V.E.Herscovitz and G.Jacob. Phys. Rev. Lett. 12, 313, 1964.

(11) H. Georgi, H.R. Quinn and S. Weinberg. Phys. Rev, Lett 33, 451, 1974.

(12) A.J. Buras et al. Nucl. Phys. 131B, 308, 1977.

(13) W.J. Marciano. Phys. Rev. 20D, 274, 1979.

ADDITIONAL NOTES

Classical Mechanics and Electromagnetic Theory

(P.5) The logical order for the equations is $r = ct$, which fixes c; $Gm = c^2r$, which fixes G; $Gm^2 = Q^2$, which fixes Q.

Grand Unification

(P.26) For quarks with unit charges, an SU(5) grand unification actually predicts $\sin^2\theta = 0.25$. Thus, for a V-A theory with left-handed (ν_e,e) and (u,d) doublets and right-handed singlets of SU(2) XU(1), a single irreducible representation of the unified gauge group requires that $\sum(I_3^2)$ u+d+e+ν = 2 and $\sum Q_i^2 = 8$, and so $\sin^2\theta = 2/8$. This is much closer to the experimental value of 0.23 $\pm$0.015 than the 0.375 required by the fractional-charge quark theory.

(P.28) The fifteen left-handed fermion states can be accommodated into SU(5) as follows:

$$\bar{5} = (\bar{3}, 1) + (1,2)$$
$$= (\bar{d}) + (\nu_e, e^-)$$
$$10 = (\bar{3}, 1) + (1,1) + (3,2)$$
$$= (\bar{u}) + (e^+) + (u,d) \quad .$$

If the X particle is identified as ud or $\overline{ud}$, then such transitions as ν_e from (ν_e, e^-)$\longleftrightarrow\bar{d}$, u from (u,d)$\longleftrightarrow e^+$, and d from (u,d)$\longleftrightarrow\bar{u}$ will all be possible. Some transitions, such as $n\bar{p}\rightarrow\bar{\nu}_e e^-$ could be expressed in terms of a particle equivalent to 2X (uudd), which is similar to that involved in the supposed decay of protons to $e^+\pi^0$. (The particles listed by Buras et al. (3), which, according to the SU(5) unification scheme, actually mediate baryon-lepton transitions, are all equivalent to ud, uu, dd or combinations of these.)

Gravitational Waves

(P.41) Since inertial forces are fictitious, we can only employ them to "explain" such nonphysical effects as inertia. Inertia is a fundamental property of mass, derived from symmetries within the parameter group, and has no need of a "physical" explanation. It is significant that inertial forces are repulsive, unlike gravitational forces; this makes the term ω positive, unlike g, and so, for comparability with Maxwell's equations, in which E and B are both positive, we take $-g$ instead of g in the equations from which "gravitational waves" are derived. This means that our equation analagous to $\nabla.E=0$ is really $\nabla.(-g)=0$, where $-g$ represents the static inertial repulsion which apparently balances the gravitational attraction. The form of this equation is, in addition, insensitive to the velocity at which the interaction is transmitted; thus, the real gravitational attraction, transmitted at infinite velocity, and the fictitious inertial repulsion, transmitted at the velocity of light, are both described by the equation $\nabla.g=0$. The effect of this is that, when the inertial forces are introduced with the 4-vector space-time of electromagnetic theory, the equations which describe them are of identical form to the equations of that theory, the inertial fields are of exactly the same form as the electric and magnetic fields, and the quanta of the inertial fields or gravitons are massless spin 1 particles of the same kind as the photon. They are <u>not</u>, however, quanta of the real gravitational field and are, therefore, <u>not</u> sources of that field, which means that they are not spin 2 particles as supposed in the conventional (strong) theory. (They could not, of course, be spin 2 particles if they transmit a force of <u>repulsion</u>.)

Mach's Principle

(P.43) Some authorities maintain that r_u must be the Schwarzschild radius for the mass m_u. This would make $2Gm_u = r_u c^2$, and the equation for apparent expansion would then become

$$\frac{1}{r^2\ c^2}\left(\frac{dr}{dt}\right)^2 = \frac{8\pi}{3}\ \frac{G\rho_u}{c^2}$$

However, use of the Schwarzschild radius for r_u would imply that the Hubble universe must be a closed system to which the usual mass/distance ratio (c^2/G) would not apply. If we reject nonlinearity of the gravitational field, then there is no system (not even a fundamental charge at the GUM) which is confined within a space less than twice the Schwarzschild radius.

Background Radiation

(P.45) The background radiation is isotropic because it is taken over an infinite assemblage of electrons, each of which creates uniform density in the Hubble radius surrounding it; over an infinite time period, thermal equilibrium will average out the contributions from all sources. Since this radiation represents information lost from the energy state of photons, it must become totally disordered thermal energy and cannot be identifiable with photon energy of particular wavelengths. This means that, to become equivalent to defining a uniform and isotropic background temperature, the energy lost from the photons must be redshifted out of existence as energy defined in terms of frequency and wavelength. This is equivalent to photon energy redshifted at the velocity of light, but, since the radiation is the result of the distribution of matter in the entire universe, any motion with respect to this matter will produce a local anisotropy in the temperature of the background radiation because the equivalent velocity of photons in the direction of this motion will no longer be exactly equal to the velocity of light.

If the thermal energy available to each proton from the background radiation ($= m_e c^2$) is described as kT_F, then, using Stefan's law, we may suppose that the rate at which this energy becomes available to protons is $\sim (Gk^4T_F^4/3\hbar^3c^5)^{\frac{1}{2}}$, the rate at which $m_e c^2$ of photon energy is converted into thermal energy via redshift. If this is equated to the reaction rate for neutron-proton transmutation, $\sim G_F^2\ k^5T_F^5/\hbar^7c^6$, then we obtain a value of $\sim 1.4 \times 10^{-62}\,\mathrm{Jm}^3$ for the Fermi constant G_F, and from the expression

$$G_F = 2^{\frac{1}{2}}e^2\hbar^2 \ /\ m_W^2c^2\sin^2\Theta_W\ ,$$

we obtain $\sim 80\ \mathrm{GeV}/c^2$ for the mass of the W boson.

Galaxies

(P.45) The time period associated with the formation of stars within a galaxy is comparable to the Hubble time because of the uniformity and isotropy of the universe. Thus, the time required for the contraction of a cloud of gas of mass m and radius r to form stars is of order $(r_u^3/Gm_u)^{\frac{1}{2}}$ because, for a universe of uniform density, $r^3/m = r_u^3/m_u$.

The Classical Radius

(P.45) The term r_e in such equations as $Gm_e/r_e^{\,2} = Gm_u/r_u^{\,2}$ need not necessarily be taken as the classical radius, e^2/m_ec^2 — $e^2/2m_ec^2$ and $2e^2/3m_ec^2$ are alternative possibilities depending on the model of the electron used; but no such expression is likely to be exact and the classical radius is probably sufficiently accurate for the approximate calculation.

The Hubble Constant and the Mean Density of the Universe

(P.46) Currently accepted values for the Hubble constant (c/r_u) create serious difficulties for the big bang cosmology, for they suggest that the age of the universe cannot be greater than about 9 billion years, whereas there is already firm evidence that some of the stars in our own galaxy are at least 14½ billion years old. Another problem - not related to the size of the Hubble constant - is raised by the so-called "missing mass" of the galactic clusters and quasars; if the redshift of the clusters is a Doppler effect produced by a real recessional motion, then the masses of the clusters predicted by the virial theorem ($=v^2R/G$) are an order of magnitude greater than those found by summing the masses of the component galaxies; also, if the redshift of the quasars is really cosmological and due to a universal expansion, then their distance - luminosity relation suggests that the curvature parameter q_D (~ 0.7) is sufficient to "close" the universe, whereas the accepted value for m_u is only $\frac{1}{4}$ of the mass required to do this. It seems that, on both counts, the universe is insufficiently dense to support the expansionary explanation of redshift. Although various ad hoc hypotheses - black holes, neutrino masses, etc. - have been suggested to overcome the problem of the "missing" mass, it does not seem very likely that any acceptable explanation will be found for the relatively high value of the Hubble constant.

Neither problem, of course, arises in a nonexpansionary cosmology based on a non-Doppler explanation for the cosmological redshift, but, if Mach's Principle is assumed to apply within the Hubble radius, then the mean density of the universe remains an important parameter. However, estimates of this quantity based directly on the Hubble constant appear to be inaccurate, and it is probable that a better value may be obtained in terms of the completely isotropic microwave background radiation. The equations $Gm_u = r_u c^2$ and $Gm_u/r_u^2 = Gm_e/r_e^2$ are approximations based on the assumption that the Hubble universe is a sphere of uniform density; the effect of this appears to be an overestimation of the value of ρ_u. Now, the mean density (ρ_T) required to produce an exact balance between $3n_e m_e c^2/4\pi r_u^3$ and aT^4 at $T=2.76K$ is $\sim 10^{-27}$ kgm^{-3}. This is, as we would expect, slightly lower than the observed densities of the galactic clusters and can be regarded as a reliable average value. (It would be even lower if Mach's Principle makes any contribution to the mass of the proton.) Also, for $m_u/r_u^2 = 4\pi\rho_T r_u/3$ and $r_u \sim 10^{26}m$, the apparent discrepancy between Gm_u/r_u^2 and Gm_e/r_e^2 based on theoretical estimations of m_u is reduced.

Cosmological Constants

(P.46) With the addition of Stefan's Law and the formulae $n_\gamma kT = n_e m_e c^2$ and $n_e = n_p \sim 0.9\ m_u/m_p$ or $0.9\alpha/13.4\ m_e$, it becomes possible to derive three further relationships:

$$n_e \sim \frac{0.9\alpha}{13.4} \exp \frac{4\pi}{9\alpha},$$

$$\frac{n_\gamma}{n_e} \sim \left(\frac{13.4 \times 4\pi^3\alpha^2}{45 \times 0.9}\right)^{\frac{1}{4}} \exp \frac{\pi}{18\alpha}$$

and $$kT \sim c^2 \left(\frac{\hbar c\alpha}{G}\right)^{\frac{1}{2}} \left(\frac{45 \times 0.9}{4\pi^3 \times 13.4\alpha^2}\right)^{\frac{1}{4}} \exp \frac{-\pi}{6\alpha}.$$

(We may also note that $\rho_u = (3c^5/4\pi G^2\hbar\alpha) \exp(-2\pi/3\alpha)$.) These, again, predict values correct to within an order of magnitude for the fundamental constants they contain, and depend only on the three irreducible constants G, h and c (or c_o), and α.

The Fine Structure Constant

To calculate α itself, we assume s^2 for the proton is $\sim\hbar c$ and the GUM is $(hc\alpha/G)^{\frac{1}{2}}$. For $c_o = c\alpha$ at grand unification, the energy ratio becomes $(hc\alpha/G)^{\frac{1}{2}}\alpha^2/m_p$. Then, adapting the formulae given by Georgi et al. (11), to apply to an integrally charged quark model, we find that

$$\log_e \left(\left(\frac{hc\alpha}{G}\right)^{\frac{1}{2}} \frac{\alpha^3}{13.4m_e}\right) = \frac{4\pi}{11\alpha} (\sin^2\theta - \alpha).$$

Substituting for m_e gives

$$\log_e \left(\frac{(2\pi)^{\frac{1}{2}}\alpha^3}{13.4}\right) + \frac{\pi}{9\alpha} \sim \frac{4\pi}{11\alpha} (\sin^2\theta - \alpha).$$

The more exact expression requires that $\sin^2\theta \sim 0.22$ for $\alpha = 1/137$. The corresponding value of α_s ($= s^2/\hbar c$) at 10 GeV is determined from $\log_e 10/0.938 = 4\pi(1-\alpha_s)/11\alpha_s$ as ~ 0.32, in agreement with the result from electroproduction analyses (12), while the value for α_{GUM} or $11\alpha/4\pi\ (\sin^2\theta - \alpha)$ is 1/32.6. The required value for $\sin^2\theta$ is of exactly the right order to accord with that renormalized for radiative corrections, according to the formalism of Marciano (13), from a theoretical 0.25 maximum.

Now, if we adapt the equations of Buras et al. (3) to a six-flavour SU(5) system with integral charges, we find that

$$\log_e \frac{m_s}{m_\mu} \sim \frac{8}{21} \log_e \frac{\alpha_s(m_s)}{\alpha_{GUM}},$$

where $1/\alpha_s(m_s) = 1 - (11/4\pi) \log_e m_p/m_s \sim 1/2.1$. This predicts $m_s/m_\mu \sim 4.9$ for $1/\alpha_{GUM} = 32.6$. Since we already have independent knowledge of the m_s/m_μ ratio, we can, in principle, use it to determine α_{GUM} and, hence, α, though at this stage only a highly approximate result can be expected. The corresponding equation for $\log_e m_b/m_\tau$ requires $m_b/m_\tau \sim 2.7$, since $\alpha_s(5\text{GeV}) \sim 0.406$. For $m_\tau \sim 1.8$ GeV, this means that $m_b \sim 4.8 - 4.9$ GeV.

It is possible also that α may be derived even more directly by an argument connected with the cosmic production of neutrons from electron-proton pairs. Thus, if we assume that protons are nearly always found associated with electrons, we may take the probability of finding an electron-proton pair within the Hubble universe as the probability of finding an electron; this will be the ratio of the total Compton volume of the electrons within the Hubble radius to the Hubble volume itself, $n_e \hbar^3/m_e^3 c^3 r_u^3$. Now, if we take the probability of an electron-proton pair actually creating a neutron at any instant as n_e/n_γ exp 2.3, where n_e/n_γ is the probability of the proton receiving $m_e c^2$ of thermal energy and 1/exp 2.3 is the Boltzmann factor, then the probability of a neutron being created anywhere within the Hubble universe must be $\sim n_e^2 \hbar^3/m_e^3 c^3 r_u^3 n_\gamma$ exp 2.3. If this is equated to the probability of finding a neutron, $n_e \hbar^3/m_n^3 c^3 r_u^3$ exp 2.3, then $n_\gamma/n_e \sim m_n^3/m_e^3$. If we now substitute $(m_n/m_e)^3$ or $(13.4/\alpha)^3$ into the equation for n_γ/n_e, we find that

$$\frac{1}{\alpha^{14}} \sim \left(\frac{4\pi^3}{45 \times 0.9 \times 13.4^{11}}\right) \exp \frac{2\pi}{9\alpha} \quad .$$

Hence $\alpha \exp \pi/63\alpha \sim 7.09$ and $1/\alpha \sim 138$.

APPENDIX

Replies to points raised by reviewer of paper as published in "Speculations in Science and Technology", volume 6, 1983, taken from correspondence with the editor.

First Report

(1) The reviewer appears to have misunderstood the purpose of my paper which is, as clearly stated in the abstract and on p.1, to discover the true properties of space, time, mass and charge. I am, of course, aware that the imaginary co-ordinate ict as used in special relativity is merely a mathematical device and I am not misled by this at all. (I actually describe a similar mathematical "trick", with respect to mass and charge, as a comparable "device".) It is one of the discoveries of my paper that a time which is imaginary would at once explain why this mathematical trick works and also explain why time is measured only through acceleration (i.e. using real t^2) and why frames of reference in uniform motion (i.e. using imaginary t) are inertial. It would further explain, as I show later, the attractive nature of gravitational forces and the existence of a time-reversal symmetry, even though time itself is unidirectional; it even has important consequences in the Dirac theory of fermions. It would, in any case, be a remarkable coincidence that a similar "trick" may be used to set up a quaternion mass-charge which is to all appearances symmetrical to the 4-vector space-time and which neatly summarizes the elementary properties of the four interactions.

(2) The reviewer's statement is imprecise and it is difficult to make any specific comment, but it is possible to say that both ideas (time continuum and countable space) have a respectable ancestry, even going back to the Greek philosophers. I have also stressed in the paper that we have to set up special conditions to measure time which involve a periodically discontinuous space. I believe that my new understanding of space and time provides the first explanation of this principle of time measurement.

(3) In the paper I show that the interaction of masses which we call gravitation and which I describe as a fundamental property of mass must be equivalent to a new quantity acceleration, whose vector properties derive from those of space. I do not, therefore, assume that gravitation is a vector field or that this field is an acceleration field. I do imply that the vector-field is curl-free because the 4-vector representation of space-time (which is responsible for the curl terms) is introduced only for a system which contains both mass and charge and this does not apply in the purely gravitational case. If by my "assumption" of the equivalence principle the reviewer means that I equate gravitational and inertial mass I would certainly accept this as it is fundamental to my definition of the parameter, but I do not believe that my assumption is phenomenological. I have not considered equations involving source-terms because these are not really fundamental and may be derived by standard arguments from the source-free equations.

Second Report

(1) The direct velocity-dependence occurs in fluids where the expression for frictional force involves the product of two imaginary quantities, coefficient of viscosity and velocity. By averaging out the effects of an enormous number of electromagnetic interactions we can find systems in which the coefficient of viscosity is appreciably constant and can be replaced in the equation by a fixed numerical equivalent, but this is really only equivalent to setting up the special conditions which make it possible to measure time. Thus, a device which operates under simple harmonic motion can be used to measure time, because we have achieved some special conditions which make the imaginary angular frequency a close approximation to a constant. In both cases force remains real and time imaginary; a

more exact physical analysis would prevent the elimination of coefficient of viscosity and angular frequency because they are not true (or universal) constants.

(2) This is a very interesting question because it leads to an entirely new illustration of the fundamental symmetry between the parameters in the particular case of mass and time. Noether's theorem about invariance under continuous groups of transformations apparently leads to the fact that the conservation of the Hamiltonian, or mass-energy of a system, is a consequence of the invariance or translation-symmetry of time, while the classical conservation laws of momentum and angular momentum become respective consequences of the translation - and rotation - symmetries of space. It would seem, therefore, among other things, that the conservation of mass, which I claim to be fundamental to the definition of the parameter, is dependent on a more fundamental property of time. But if we examine what exactly is meant by the symmetry properties of space and time we find that they are expressions of these parameters' nonconservation; only a nonconserved parameter could possess translation - or rotation - symmetry. Thus, for example, the electromagnetic, strong and weak charges would not be distinguishable and independently conserved if charge possessed the same rotation-symmetry as space. Systems are defined by their unchangeable masses and charges but are arbitrary in their choice of space and time. In other words, the consequence of Neother's theorem is that mass is conserved because time is not. This is, in fact, one of the fundamental principles of my theory. The fundamental properties of the four parameters are determined by an exact symmetry between them; mass and charge are conserved because space and time are not (though this does not, of course affect the status of conservation as a fundamental property of mass and charge).

Here we see a remarkable application of the general symmetry of the parameter group in a particular area of physics. A very similar thing occurs in the Dirac theory of fermions, which, in reversing, for mathematical convenience, the roles of space and time, also reverses the roles of mass and charge, producing negative energy states; other examples - involving the unidirectionality of time, 4-vector space-time, independent conservation of charges, etc. - are integrated into the main argument of the text, but the Noether's theorem example seems to be an exceptionally clear-cut instance, and it is of extra significance as the first response of the theory to an entirely new question.

(3) The apparent differences between the countability of charge and space arise from the fact that two properties are involved rather than one. Charge is unlike space because it is both countable and conserved, whereas space is countable but not conserved. Conservation fixes charge but nonconservation makes space arbitrary. With respect to this property the two parameters must be exact opposites. If we could count fixed units of space in a system then it would have to be a conserved quantity like charge. The point about space and time is that, because they are not conserved, we can put no restrictions on their variation - this is what is really meant by their translation - and rotation - symmetries (c.f. (2)); if space existed only in discrete fixed units we would be imposing a restriction on its variation, but we impose no such restriction by saying that it is merely countable. (That a parameter can be equally unrestricted in its variation and not countable is shown by the argument about time on p.1).

(4) Since the discussion of mechanical systems is concerned with purely classical conditions (i.e. excluding curl and 4-vector terms), it is as valid, in this context to define force as mass x acceleration as it is to define it as rate of change of momentum, and I choose the former for convenience of presentation. However, I introduce 4-vector space-time for systems involving charges (which, in particular, means particles), and when I refer in general to such relativistic ideas as the Lorentz transformations and the equivalence of mass and energy, I mean to imply here, also, that this automatically includes the relevant 4-vector modifications of classical dynamics, such as the definition of force as rate of change of momentum. I thus define force as mass x acceleration ($\equiv$ rate of change

of momentum) where the classical case is sufficient and as rate of change of momentum where I need to use the relativistic or 4-vector case.

(5) I think that most of the reviewer's concern over my treatment of Newton's law of gravitation has arisen from the fact that I did not make it sufficiently clear that I regarded it as a limiting case entirely parallel to Coulomb's law of electrostatics. I believe that my arguments are actually quite rigorous for the restricted conditions for which they apply. I stated that $\nabla . g$ must vanish because it represents both information that should be available and information that cannot be recovered, and only zero information could apply to both conditions. In less restricted (i.e. nonclassical) conditions, of course, this argument may no longer apply and $\nabla . g$ could possibly be of the form $-\mu^2 \phi$. In this paper I am particularly interested in defining the conditions in which both $\nabla . g$ and $\nabla . E$ vanish because this parallel is an important aspect of the relationship between the parameters mass and charge; I am not specifically concerned with devising a theory of gravitation. I have been concerned throughout with justifying my claim that the symmetrical group containing the four parameters includes all the information which can be discovered about them; this means that those laws of physics (such as Newton's law of gravitation), which are concerned only with fundamental properties or interrelationships of parameters, must be interpreted in terms of the organization of information from the parameter group.

Third Report

(1) I interpret the arguments presented in the paper as suggesting that, if, in principle $\nabla . g$ could be of the form $-\mu^2 \phi$, even under classical conditions, in fact it is not. There is no reason to believe that any other equation than $\nabla . g = 0$ is necessary to describe gravity under classical conditions. The Newtonian approximation is assumed arbitrarily in General Relativity but I believe that it has a fundamental justification which is not purely phenomenological in origin, and we must still explain its success as a limiting case.