A NEW FORMAL STRUCTURE FOR DERIVING A PHYSICAL INTERPRETATION OF RELATIVITY

PETER ROWLANDS

Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 3BX

ABSTRACT

It is suggested that the formal structure of relativity and its physical interpretation arise from quite different aspects of a more fundamental formal structure relating the four parameters space, time, mass and charge; and that the link between structure and interpretation derives from numerical relationships between the parameters which are intrinsic to the more fundamental formal structure.

The key formal structure in the whole of relativity, special or general, is undoubtedly the 4-vector relation between a 3-dimensional real space and a 1-dimensional imaginary time. 4-vectors are an extremely convenient way of representing the formal structure of space-time, but they do not arise in a natural way from the rules of algebra. They can be codified by algebraic rules, but there is nothing in algebra, for instance, which fixes the number of spatial dimensions at 3. So why does this peculiar, artificial structure fit so well to space and time? Hamilton thought he had guessed the answer in 1843. He discovered an algebraic structure virtually identical to the 4-vector except that it had three imaginary parts (i, j, k) and one real part (1), rather than three real (x, y, z) and one imaginary (ict), the relationship between these components being determined by the simple algebraic expression

$$i^2 = j^2 = k^2 = ijk = -1 .$$

The *quaternion* structure which Hamilton introduced arises purely from algebra (from the properties of imaginary numbers) and so is entirely natural, unlike the 4-vector. It is the only possible extension of the ordinary complex numbers (1, i) into higher dimensionality. That is, imaginary numbers are constrained by algebraic rules to have either dimension 1 (as in ordinary complex numbers) or dimension 3 (as in quaternions), *but no other*, and they cannot be defined without incorporating a corresponding real part.

Hamilton was so convinced that he had discovered the reason for 3-dimensional space that he immediately applied quaternions to the problem and even came close to the discovery of 4-dimensional space-time. There was, of course, an attempt after him to apply quaternions to vector problems, but the extra imaginary factors proved an inconvenience and, at the end of the nineteenth century, Heaviside and others devised a mathematical codification of vector algebra which, in its extension by Minkowki, was the "mirror image" of the quaternion system. Now, it is surely a most provocative fact that the formal structure most appropriate to space and time is not a natural algebraic system, but is the *mirror image* of one, and there has always been a feeling among some physicists that quaternions ought somehow to have a fundamental role to play in physics. In this paper, I am going to suggest that they *have* – in a system which is, for some reason, the mirror image of space and time: this is the system of *mass and charge*.

Simple dimensional analysis suggests that, apart from space and time, physics needs only the source terms for the fundamental interactions. There are four such interactions. Three are held to be alike, or would be under ideal conditions; these are the electromagnetic, strong and weak interactions, and we may describe their source terms as electromagnetic, strong and weak *charges* (which we may represent by e, s and w). The other is the gravitational interaction, whose source term is mass. It is well-known that like charges generate forces of opposite sign to like masses, a fact which has always defied rational explanation. There is no need for explanation, however, if one type of source is real and the other imaginary; the difference in sign of the force, depending on the source term squared, is then merely a result of following the rules of algebra.

It is clear that it is the charges which must be imaginary. Mass cannot be, as it has only one sign, and it is one of the most fundamental rules of complex algebra that positive and negative imaginary quantities must have equal status. In fact, it is this very rule which makes it necessary to have a system of antiparticles, as well as particles, for even particles with no *electric* charge, such as the neutron and the neutrino, still have strong and/or weak charges, and so antiparticles must exist with charges of opposite sign. Only those particles with no charges of any kind, such as the photon or neutral pion, have no antiparticles or are their own antiparticles. The three types of charge and mass thus form the components of a natural quaternion system (say, ie, js, kw, m), and, like the dimensions of space and time, they are combined by Pythagorean addition when the source terms are squared in physical forces. The quaternion system even predicts the Grand Unification now being sought by particle physicists and already partially achieved in the electroweak unification of Weinberg and Salam.

It is my belief that this remarkable symmetry between the mass-charge quaternion and the space-time 4-vector is no accident but is part of an even more fundamental set of relationships between the parameters space, time, mass and charge, and that these relationships bear directly upon the physical interpretation of relativity and its relation to the formal mathematical structure of the theory. It is, I would argue, a *result* of this symmetry that space-time is a 4-vector, the nature of this quantity being determined by symmetry with a mass-charge quaternion, whose nature follows directly from the rules of algebra:

mass-charge quaternion	space-time 4-vector
ie	x
js	y
kw	z
m	ict

4-vectors, of course, are responsible for the formal structure of special relativity and are also a necessary component of the General Theory. The *physical interpretation*, I would suggest, has a rather different origin. For this we need to examine the two "one-dimensional" parameters, mass and time. These two parameters have something particular in common – they are both *continuous* or indivisible. It is the continuity of mass which provides the physical interpretation of relativity, whether classical aether or quantum mechanical vacuum; and this property is quite independent of the 4-vector structure of space and time. Consequently, the inability of signals structured as 4-vectors to detect changes in the velocity of light is quite independent of whether an aether can be applied to fundamental physical questions.

Now, absolutely continuous quantities are quite naturally 1-dimensional, or more strictly *non*-dimensional, for the very existence of dimensionality necessarily breaks continuity. Continuity also explains without difficulty why masses are all of the same sign and why time is irreversible, any violation of these conditions leading immediately to discontinuity. Many people have considered it a profound mystery that time is physically irreversible while allowing both positive and negative solutions in equations, but there is no mystery at all if time is both continuous and imaginary: continuity makes it irreversible, whereas the imaginary nature makes it have two mathematical solutions of equal status. The *imaginary* nature of time also explains why quantities involving time squared, such as force and acceleration, are the important ones in physics, rather than those involving only time to the first power, and, indeed, why time is only "measured" under conditions in which the second power must be invoked.

If the nondimensional parameters are continuous or indivisible, then we might expect the *dimensional* parameters, space and time, to be, in some sense, noncontinuous or divisible. This is obviously true in the case of charge, but to understand its application to space, we need first of all to investigate another property which exhibits fundamental symmetries between the four parameters: this is the property of conservation / nonconservation. It is well-known, of course, that mass and charge are conserved quantities, space and time are not; but it is not so well-known that the "nonconservation" of space and time is not merely the absence of conservation, but is also its *exact opposite*. Elements of mass and charge have unique identities. This is what we mean when we say they are conserved; and the unique identities are shown by the fact that they obey *local*, rather than global, conservation laws. Elements of space and time have no unique identities. As a consequence, space and time are translation symmetric – each element is as good as any other; while space, as a dimensional parameter, is also rotation symmetric – each direction is as good as any other. In the same way, we can say that mass and charge are "translation" *a*symmetric, one element *not* being replaceable by another; and charge, as a dimensional parameter, is also rotation asymmetric, one *type* of charge not being replaceable by another. It is this last property which is responsible for the conservation rules of particle physics (and also the non-decay of the proton); electromagnetic, strong and weak charges are *not* interchangeable and are subject to separate conservation laws.[1]

Now, charge, as we know, is a discontinuous or divisible parameter – physically, it is found only in fixed units, each of the same size. Space is certainly a divisible parameter, for, if it were not, the entire act of measurement (which is made solely in terms of space) would be inconceivable; but it does not have fixed units. The reason is obvious as soon as we realise that space is *nonconserved*, for a nonconserved parameter could not possibly have fixed units; on the other hand, a conserved quantity *must* have them. It is unfortunate that mathematicians have chosen to apply the word "continuity" in this context and so make our use of the word "discontinuous" ambiguous; but there is no ambiguity if we describe space and charge as divisible, and mass and time as indivisible. It is interesting that it is *variability* (i.e. nonconservation) that is the root of *differentiability* (the mathematicians' continuity) and that this is independent of the distinction between divisibilty and indivisibility; this is why two independent sytems of differentiation were adopted by the followers of Leibniz and Newton (one discontinuous and the other continuous), and why the mathematical definition of differentiation has remained ambiguous to this day.

Dimensionality is clearly linked to divisibility, but we will not attempt at present to make a rigorous connection. However, we can say with rigour that a dimensional imaginary quantity is necessarily *three*-dimensional. The three-dimension-ality of space then follows automatically if it is exactly symmetrical in this respect to charge.

We have now established properties (reality, conservation and divisibility) in which each pair of parameters is alike, while exactly opposite in the other two. The information may be conveniently summarised as follows:

space	real	nonconserved elements nonunique	divisible dimensional
time	imaginary	nonconserved elements nonunique	indivisible nondimensional
mass	real	conserved elements unique	indivisible nondimensional
charge	imaginary	conserved elements unique	divisible dimensional

This a group of order 4. There is no reason to suppose that the symmetry is not both exact and exclusive. (The CPT theorem would be an obvious consequence of exclusivity.) In this group, all the elements have equal status: each can be structured as the identity element and each is its own inverse. The group nature is clearly seen if we assign symbols to the properties, for example, as follows:

+a	real
–a	imaginary
+b	nonconserved
–b	conserved
+c	divisible
–c	nondivisible .

The elements then follow rules of binary operation such as:

$$+a * -a = -a * +a = -a$$
$$+a * +a = -a * -a = +a$$
$$+b * -b = -b * +b = -b$$
$$+b * +b = -b * -b = +b$$
$$+c * -c = -c * +c = -c$$
$$+c * +c = -c * -c = +c\ .$$

This arrangement makes space the identity element, but this property can be reassigned to time, mass or charge by changing over the signs of a, b or c.

If the symmetry is exact, it has a very interesting consequence: the fixed numerical relation between mass and charge, determined by the quaternion structure, and the fixed numerical relation between space and time, determined by the 4-vector, must be extended by group symmetry to provide a fixed numerical relation between each single parameter and every other, and an additional fixed numerical relation between each single parameter and *the inverse of* every other.

Such relationships are already known. They require the existence of four fundamental constants – three for the direct relationships, and one to convert the direct relationships to inverse ones. The constants also exist, though, in practice, we do not use them direct, but in combinations which are the accidental result of our historically-derived systems of units. Incorporating one of these constants, for convenience, into the fundamental charge **q** (presumably the value obtained at Grand Unification, rather than that associated with the electromagnetic interaction), we obtain familiar dimensional relationships from which the complete set of direct and inverse relationships between the parameters space, time, mass and charge (**r**, **t**, **m**, **q**) may be derived:

$$\mathbf{r} = c\mathbf{t}$$
$$G\mathbf{m} = c^2\mathbf{r}$$
$$\mathbf{q} = G^{1/2}\mathbf{m}$$
$$\mathbf{m}c^2 = h/\mathbf{t}\ .$$

The basic laws of physics are derivations from dimensional relations of this kind, combined with information supplied by conservation laws, equations between physical quantities being nothing more elaborate, in principle, than numerical relations between different systems of units.[2] One equation in particular, $E = mc^2$, often associated with the special theory of relativity, though it is not, in fact, a *deductive* consequence of that theory (depending, as it does, on an arbitrary value for a constant of integration),[3] is seen to be based on something more fundamental, the direct relation which must exist, by symmetry, between the units of mass and space, or charge and space, or, indeed, between the quantities involved in the mass-charge quaternion and those involved in the space-time 4-vector. It is precisely this kind of relation which makes it possible to link the mathematical structure of the theory of relativity, which concerns space-time, with its physical interpretation, which concerns mass-charge.[4]

NOTES

1 Baryon number is conserved because baryons are the only particles with strong charges and these are always the same sign; lepton number is conserved because leptons are the only particles with weak, but no strong, charges. Protons cannot decay directly to positrons because this would require the spontaneous annihilation of single units of strong charge, or their conversion to charges of a different kind, and, hence, violate the rotation *a*symmetry of charge. Fermions, also, cannot be transformed into bosons because fermions have weak charges, whereas bosons do not.

2 The fundamental laws of physics are of two kinds: definitions of quantities in terms of other quantities, and statements (in the form of differential equations) that some quantities are conserved while others vary continuously. Both types of law derive naturally from the kind of symmetry proposed in this paper; rigorous derivation of the equations of classical mechanics and electromagnetic theory is a straightforward consequence. Also, quantum physics arises naturally from the existence of inverse, as well as direct, relationships between the parameters, while the constant h is shown to be fundamental to the structure of classical, as well as quantum, laws.

3 The equation $E = mc^2$ (as opposed to $\Delta E = \Delta mc^2$) is derived inductively in relativity by integration of the Lorentz-invariant equation $dT/dt = \mathbf{F}.\mathbf{v}$ and by arbitrary choice of the integration constant.

4 Many other aspects of physics are illuminated by the kind of fundamental formal structure discussed in this paper. For example, Noether's theorem links the conservation of energy (i.e. conservation of mass) with the translation symmetry of time (i.e. nonconservation of time); the group symmetry derives this directly. It also predicts links between the conservation (or translation asymmetry) of charge and the translation symmetry of space (or conservation of linear momentum), and the conservation of type of charge (rotation asymmetry of charge) and the rotation symmetry of space (or conservation of angular momentum). In addition, we may expect the simultaneous requirements of rotation asymmetry for charge and rotation symmetry for imaginary quaternion components to provide clues as to the nature of the quark system; while the origin of gauge invariance is clearly marked in the nonconservation properties of space and time.