

Absolute relativity in classical electromagnetism: the quantisation of light

J.G.Williamson^a, S.J. Leary^{a,b}

^a University of Glasgow, College of Science & Engineering, Glasgow G12 8LT, Scotland;

^b Now at: SAS Institute Inc., Glasgow, Scotland.

ABSTRACT

A rigorous introduction of the underlying nature of space and time forces qualitatively new kinds of solutions in the classical theory of electromagnetism. A class of relativistic wave-functions are derived which are solutions to the first-order, free-space Maxwell equations. These describe all photons from radio to gamma waves and are governed by a single parameter: the exchange frequency. Though the theory remains that of classical, continuous electromagnetism, allowed travelling-wave solutions are quantised in that they come in “lumps” and their characteristic energy is proportional to frequency.

Keywords: light quantisation

1. INTRODUCTION

Since the early twentieth century theoretical effort has focused largely on the understanding of quantum mechanics and the development of gauge theories following on from the hugely successful theory of quantum electrodynamics. This paper picks up on an older path, that of classical electromagnetism, and develops it within a relativistic mathematics designed to parallel experiment as closely as possible. The point of departure is represented by Maxwell’s classic text-book,¹ rather than more recent formulations of electromagnetism with a more complex superstructure.² The Maxwell theory has been re-cast in a minimal mathematics forced to parallel closely the experimental, relativistic, properties of space and time. The algebra generated is closely related to certain Dirac³ and Clifford algebras,⁴ but is more restrictive in some respects and somewhat generalised in others. In particular, a principle is adopted which forces all quantities, in all equations, to all orders to take their proper space-time form. This may be viewed as bringing in the experimentally observed nature of space and time everywhere at a fundamental level, and as an extension and sharpening the principle of (special) relativity. Because it is both, this will be referred to here as “the principle of absolute relativity”. The severe constraints of this approach allow the usual four Maxwell equations to be written as a single equation (rather than in two pairs as is more conventional²) in a form similar to that of the Dirac equation. The resultant equation may be expanded into a set of eight coupled differential equations, four of which take the form of the Maxwell equations. The new theory is more general than that of either Maxwell or Dirac and has new kinds of solutions corresponding to both light and material particles. The latter are circulating, necessarily charged and with half-integral spin. Such solutions are beyond the scope of the present work. The aim here is to explore only light, leaving an investigation of the origin of charge and half-integral spin to a companion paper.⁵

Here, the solutions of that subset of the new theory corresponding to the free-space Maxwell equations alone will be explored. It is found that the only allowed propagating solutions are quantised. That is: it is claimed that the rigorous implementation of the principle of absolute relativity leads to the quantisation of light into photons. The structure of the paper is as follows. Firstly, the mathematical framework will be defined. Secondly, a new set of equations, encompassing the Maxwell equations, will be derived. Thirdly, a new kind of wave function, incorporating the principle of absolute relativity, will be discussed. Fourthly, on the basis of this, it will be shown that field only solutions correspond to the properties of light as observed experimentally. Electromagnetism remains continuous. Locally, fields are unquantised, but propagation over longer distances forces solutions to be quantised as a consequence of the deeper principles of energy and field linearity and absolute relativity.

Further author information: (Send correspondence to J.G. Williamson)

J.G. Williamson: E-mail: john.williamson@glasgow.ac.uk, Telephone: +44 141 330 4923

2. SPACE, TIME AND ROOT-ENERGY

Often, it is argued that a more general mathematics is more powerful than a simpler one. If one wishes to make an attempt to properly parallel reality, as in a solution of Hilbert's sixth problem for example, then one needs to find the simplest mathematics that parallels reality, just and no more. Mathematics is no more than a formal structured language used to aid thinking. Any language may aid thinking, but if improperly defined may cloud it. The mathematics will be designed here to enforce the known experimental properties of relativistic space and time as starting axioms. This mathematics will incorporate a principle: that no term may appear without its proper relativistic form. This is the principle of absolute relativity. The further extension of this structure to describe energy in the electromagnetic field and related quantities will be kept as simple and restrictive as possible. Most of the structure and results of this paper will arise from a consideration of the proper underlying nature of space and time and its products and quotients. This is not merely a philosophical choice: it is precisely the severe constraints imposed on the mathematics which lead to the necessity of travelling wave solutions of continuous classical electromagnetism being quantised in the following.

Because it is the cause of much confusion, even amongst professionals, a discussion on units is required. In a single equation, all quantities must appear in the same units. In S.I. units in free space where ϵ_0 is the permittivity, the energy density of an electric field \mathcal{E} is $\frac{1}{2}\epsilon_0\mathcal{E}^2 = \frac{1}{2}\mathcal{E}.\mathcal{D}$. Here, the electric displacement \mathcal{D} represents the same physical field as \mathcal{E} , but with units such that the product with the electric field has the units of energy density. Maxwell¹ identified three such pairs whose product had the dimensions of an energy density, the "electrostatic pair" (the electric field \mathcal{E} and displacement \mathcal{D}), the magnetic pair (the magnetic flux \mathcal{B} and the magnetic field \mathcal{H}) and the "electrokinetic pair" (the vector potential \mathcal{A} and current density \mathcal{J}). It is worth noting that, for Maxwell, the relationship between vector potential and current was similar to that between field and flux or displacement - representing the same underlying physical quantity in free space.¹ This has become less current in theoretical circles, but is well-understood in the areas of quantum transport and practical engineering.⁶ Since there is, here, one and only one physical field, current, mass and so on the mathematics is used to enforce this. Define the "electric field" $E = \sqrt{\mathcal{E}.\mathcal{D}}$ and the "magnetic field" $B = \sqrt{\mathcal{B}.\mathcal{H}}$. The symbols E and B have chosen as they relate directly to the quantity of charge.¹ The quantity B will be referred to here as the "magnetic field" even though the letter is more usually reserved for the "magnetic flux". This allows all field, mass and current quantities to be expressed in the same units as $\sqrt{(Jm^{-3})}$.

Relativistic space-time is paralleled here with a sub-algebra, isomorphic to the Dirac-Clifford algebra,^{3, 4, 7, 8} encompassing the relativistic properties of points, lines, planes, volumes and a 4-volume. This algebra has sixteen independent degrees of freedom. In addition, a seventeenth degree of freedom is introduced, corresponding to a square-root energy density. This may manifest in any of the geometric forms outlined above and transforms between them in a way well-described by the new linear equation to be derived. A word is needed to label this concept which, curiously, given its ubiquity, seems absent in contemporary physics parlance. Like the electric field E or the quantum probability-density Ψ "square-root-energy-density" must be squared and integrated over an appropriate volume to yield an energy value. Here, to make the connection with earlier work^{9, 10} we will use the invented word "vot" to describe it, because the invariant rest-mass density which was earlier referred to as pivot becomes p-vot. In an obvious notation any such quantity may be denoted with a suffix: Ψ -vot for quantum mechanical mass-energy density, E -vot for the electric field, and so on.

To keep things as simple as possible three things and three things only will be introduced a-priori: space, time and vot. Since, from observation, space has three linearly independent "directions", this is five degrees of freedom. Four frame-independent unit elements are introduced. These elements represent unit lines in one dimension of time and three (orthonormal) dimensions of space. The elements themselves are frame-independent. Magnitudes or extents (which may transform between frames) are represented by quantities with the appropriate units. These are used to express a magnitude of "vot", or the apparent magnitude of (4- or multi-) vector elements in a particular frame of reference (3 (light) nanoseconds in x , for example). So far this is just (4+1) degrees of freedom (more if one ascribes dimensionality also to the physical units), of which only one, the positive definite real (vot) quantities, represents a magnitude, the others being strictly unit 4-vector elements. Additional inner complexity arises rapidly under a proper, physical, relativistic, definition of "multiplication" or "division" of the unit vector elements amongst themselves. Multiplication generates a unit point (for example of a unit line multiplied by itself), six unit planes (line times perpendicular line), four unit volumes and a unit 4-volume,

making, together with the four unit lines, sixteen linearly-independent unit elements representing the form of space and time. In addition to the sixteen linearly independent “directions”, a set of signs are required for each taking, potentially, the values $+$ and $-$ only. In particular cases, however, they may not be required (or allowed) at all. For example, clearly, one needs to distinguish “forwards” and “backwards” in Cartesian space. It is at least debatable, however, whether a minimal description of reality will require both “forwards” and “backwards” in time. Root-energy density may take either sign but it is arguable that its integrated square, the total energy, is always positive definite. In the definition of the positive direction of the unit plane formed from the product of two perpendicular lines, one should distinguish the left-handed and right-handed choices with different signs. Note, in particular, that it has no meaning to add or subtract the unit elements themselves, but only the magnitudes which condition them: $1\alpha_0 + 1\alpha_0 = (1 + 1)\alpha_0 = 2\alpha_0$ (seconds, for example). That this property is necessary should be obvious: there is no such thing as a double-sized unit element. The sign appearing in the addition or subtraction of real numbers is, again, different conceptually from the signs of the unit elements themselves. Also an extension of the simple basis here into the standard model may require more signs for various quantum numbers to distinguish such aspects as positive and negative charge, spin, lepton number and so on. It should be clear that the potential number of different “algebras” which may be defined in this way is rather large. Which, if any, is necessary as an element of the eventual solution of Hilbert’s sixth problem is left to future work. The approach followed here has been to choose a minimal system which works at the level of the Maxwell equations and which, further, corresponds as closely as possible with the conventions adopted in the standard textbooks.²

To further symmetrise the units of space and time, in the following any spatial quantity should be thought of as being in seconds by dividing by the (constant) speed of light c . That is for z in metres z in seconds is such that, $z = \frac{z}{c}$. If one wishes the units to be maintained over differentials then one must further introduce a proper (relative) unit time to create a dimensionless derivative, as discussed below. Unless otherwise stated, the unit units in the equations to follow may be taken to be those of root energy density ($\sqrt{(Jm^{-3})}$). Obviously, any such equations derived have equal validity in any other set of units with similar properties, such as Tesla, $\sqrt{(kg\ m^{-3})}$ or Vm^{-1} as an equivalent equation may be obtained by multiplying through by the appropriate constant.

A four vector is written ($\mathbf{v} = a_0\alpha_0 + a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3$). Note carefully that absolute relativity is imposed by using the a_μ to express a real magnitude (e.g. 3 Amps m^{-2}) or an extent (e.g. 42 light-nanoseconds) and the α_μ to represent the proper unit-element vector form.

A product of these base unit elements with themselves is defined such that the unit time vector, α_0 squares to the positive invariant scalar unity (the unit point) $\alpha_0^2 = \alpha_P$ and the three spatial vectors α_1, α_2 and α_3 square to the negative scalar unity $\alpha_i^2 = -\alpha_P$. This is the point at which absolute relativity is introduced and is all that is required such that all derived quantities transform correctly, relativistically, under all products and quotients. The quantity α_P represents a physical point, not in size but rather as opposed to a line or a plane or a volume. Note that, for neither product, is the value assigned to the real number 1. The quantity α_P is distinguished, here, from the real or natural number unity (1) in that it is the scalar unit element invariant under a Lorentz transformation and may take only the two values $\pm\alpha_P$. The positive value is idempotent such that $+\alpha_P^2 = +\alpha_P$. Here, in order to keep integral energies positive definite, the negative value is also taken to square to the positive unit scalar, $(-\alpha_P)^2 = +\alpha_P$. It is worth noting that, properly, the multiplication (or division) of unit vectors, of magnitudes and of numbers are, in principle, three different kinds of operations. The first results in an object of a different form, the second in quantity with a different dimension and the third in merely a different magnitude. A consequence of the definition of multiplication above is that the square of a four vector is $(a_0^2 - a_1^2 - a_2^2 - a_3^2)\alpha_P$, a manifestly Lorentz invariant quantity, as it is experimentally. For the α_μ taking the dimensions of length, for example, the square is the positive-definite invariant interval squared.

The ordered product or quotient of one spatial unit element with another, for example $\alpha_1\alpha_2$ leads to a unit right-handed ordered spatial plane (bivector) element. This unit spatial plane is denoted $\alpha_1\alpha_2 = \alpha_{12}$. The reverse ordering gives a plane in the opposite (left-handed) direction, that is $\alpha_{12} = -\alpha_{21}$. There are three such right-handed objects: $\alpha_{12}, \alpha_{23}, \alpha_{31}$. Because this is a four-dimensional basis there are three further space-time planes, represented by products such as $\alpha_1\alpha_0 = \alpha_{10}$. Because of the properties of the base elements introduced above and the nature of the product, these elements transform relativistically as the magnetic (α_{ij}) and electric

(α_{i0}) field elements which take this form. The 2-component of the electric field ($E_{20}\alpha_{20}$) may then be denoted 20-vot (pronounced two-zero-vot).

This is a general and defining feature of the algebra being developed: anything with a particular unit element form inherits the relativistic transformation properties of that form. There are 4 tri-vectors representing unit volume elements ($\alpha_{123}, \alpha_{012}, \alpha_{023}, \alpha_{031}$). The latter three take the form of a momentum density multiplied by a perpendicular unit vector, and therefore transform as the components of an angular momentum density. Finally, there is a quadri-vector (α_{0123}) which, just as the scalar, is invariant under a Lorentz transformation but may change sign under other operations such as Hermitian conjugation.¹²

Several considerations should be noted. The system is non-commutative, hence the implicit ordering of quantities is important. In the sequel, a system has been chosen which works, at least up to the derivation of the Maxwell equations. In principle, the elements derived from ordered multiplication or ordered division may be different. In particular, quantities of this form scale differently under a Lorentz transformation, as discussed below. The ordering of division (whether one divides by or divides into a quantity) introduces a sign change. Further, there are several choices to be made about the handedness and ordering of the operations between the various unit elements. In particular, the time element may be taken to come first or last (implying a change of sign and of handedness of the base elements in which it appears). Importantly, both choices give a same-handed set of products amongst each other ($\alpha_1\alpha_0 \times \alpha_2\alpha_0 = \alpha_0\alpha_1 \times \alpha_0\alpha_2 = \alpha_1\alpha_2$). This would imply that there exists an intrinsic sign of and an intrinsic handedness between certain elements. The conventions adopted here work with the standard left-to-right ordering of products, the standard (right-) handedness of co-ordinate systems and the standard signs chosen for the directions of the electric and magnetic fields. This can equally be made to work with a left-handed basis. A comment is in order here: nature is intrinsically handed. The feeling of the author is that the left-handed choice is very likely to be more correct, though the right-handed choice has the advantage that it currently the usual convention. The conventional signs in the Maxwell equations then arise if one adopts the convention that the multiplication of a unit vector in the 1 direction into an inverse unit vector in the 2 direction has the reverse sign to the simple product. That is $\alpha_1/\alpha_2 = -\alpha_{12}$. It should be immediately apparent that, with these degrees of freedom, there is more than one way of choosing a consistent system at the level of the Maxwell equations. Further, conventionally, the scaling and sign properties are taken up by a real number factor (introducing positive and negative reals then) and, rather than introducing many more base elements and obfuscating the simple development to follow, that approach will be followed here. Provided one is not working with addition or subtraction, but just with multiplication and division this is not an issue. Where it becomes an issue (in the addition of energies and fields), it proves to become more of a solution than a problem as forcing both to be linear derives the transformations of special relativity, as will be shown. In the following, the proper form of quantities will be represented by a unit token with ordered lettering, thus $\alpha_{\mu\nu}$ represents a general bivector and, $\alpha_{0ij}, \alpha_{ij}$ and α_{i0} are right-handed tri-vectors, space-space bi-vectors and space-time bi-vectors respectively. Given this, the “direction” of a hypercomplex element is assigned to the way it transforms under a planar rotation.¹² For example, the unit volume α_{012} rotates in the same way as α_3 hence, in Cartesian co-ordinates it represents the “z” component of the angular momentum density. Here and in the sequel, Greek indices run from 0 to 3 and Roman from 1 to 3.

The multiplication and division of unit vector elements has been defined. The division of 4-vectors within the algebra is now discussed. Note firstly that the algebra developed is not a division algebra. There are many regions, apart from zero, where division is not defined.⁷ Primary amongst these is the vector itself. Consider the 4-vector case:

$$\mathbf{v} = \alpha_0 v_0 + \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = \alpha_0 v_0 + \alpha_i \vec{v} \quad (1)$$

$$\mathbf{v}^{-1} = \mathbf{v}/\mathbf{v}^2 = v/(v_0^2 - v_1^2 - v_2^2 - v_3^2) = \frac{\mathbf{v}}{(v_0^2 - v_1^2 - v_2^2 - v_3^2)} \quad (2)$$

The inverse is in the same direction as the original vector, but with a different (real) scale factor, corresponding to the usual relativistic scaling. The over-arrow is used to denote the components corresponding to a conventional 3-vector. Here these are just three real numbers, with the proper (4-dimensional) unit elements being given by the α_i factors. Note, for the case of the space-time coordinates, the divisor corresponds to the invariant interval squared and that all inverses are scaled relativistically, by construction, according to this quantity. The underlying

unit elements, when squared, give quantities of opposite sign, as they must relativistically. At the same time, if the real number factors for the magnitude of the spatial and temporal parts are equal, as they are everywhere on the lightcone for example, the interval goes to zero. Hence there is no inverse, not only at zero, but also in the crucial case of anywhere on the lightcone such that $(v_0^2 - v_1^2 - v_2^2 - v_3^2) = 0$. That is, the plane where division is undefined corresponds precisely to the physical limitations imposed by the speed of light. There are other combinations as well (such as that corresponding to the photon energy and momentum, for example) where division is undefined as well. Further discussion of where division is and is not defined is of great interest in itself, but not relevant to the simple cases discussed here. It is reserved for future work.

For Cartesian co-ordinates a 4-vector 4-differential is defined within this framework as:

$$\begin{aligned}\mathcal{D}_4 &= \frac{\partial}{\alpha_\mu \partial x_\mu} = \partial_\mu / \alpha_\mu \\ &= \alpha_0 \partial_0 - \alpha_1 \partial_1 - \alpha_2 \partial_2 - \alpha_3 \partial_3 = \alpha_0 \partial_0 - \alpha_i \vec{\nabla}\end{aligned}\quad (3)$$

A dimensionless derivative for any system of proper unit time \mathcal{T} is given by $\mathcal{T}\mathcal{D}_4$. An appropriate local unit time for any given system is the inverse Compton frequency. The subscript on the left denotes the number of distinct unit elements in the expression, here that is a 4-vector derivative. Note the imposition of the principle of absolute relativity by including the (quotient of) unit vector elements explicitly. It is this that leads to the change of sign in the spatial part above since $1/\alpha_0 = \alpha_0$ and $1/\alpha_i = -\alpha_i$. Note also that the scale properties of division discussed above are simply unity, since the differential is taken with respect to each base unit vector element locally. Note that, for the case of the definition of the 4-vector derivative, division always has unit scaling since it is with respect to each component in turn. It is precisely this form of the vector derivative that appears in both the Maxwell and Dirac equations. The emitter, absorber and the exchange field may each have different definition of “locality” and different scales for their local rulers and clocks, but the proper vector 4-differential has a scale factor of unity in each and every frame. Conversely, if one were to require the process which differentiation represents in reality to be a unitary process, only derivatives with respect to the individual vector unit elements, as in equation (3) above, would be allowed.

Writing a (proper absolute relativistic) v-vot (related to the vector potential) as:

$$A_4 = \alpha_\mu A_\mu = \alpha_0 A_0 + \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 = \alpha_0 A_0 + \alpha_i \vec{A} \quad (4)$$

Conventionally, it is the 4-differential of a 4-vector potential alone that yields field components. The 16 ($= 1 + 3 + 3 \cdot 2 + 3 \cdot 2$) terms of the 4-vector-derivative of the v-vot $\mathcal{D}_4 A_4$ may be gathered together and written as:

$$\mathcal{D}_4 A_4 = \alpha_P (\partial_0 A_0 + \vec{\nabla} \cdot \vec{A}) - \alpha_{i0} (\partial_0 \vec{A} + \vec{\nabla} A_0) - \alpha_{ij} \vec{\nabla} \times \vec{A} = P\alpha_P + F\alpha_{\mu\nu} \quad (5)$$

which is the sum of a scalar (p-vot) part $P\alpha_P$ and a bivector (field) part $F\alpha_{\mu\nu}$.

This looks very similar to the conventional derivation of field from potential where in Eq. (5) the term in α_{i0} is usually identified with the electric field $\vec{E} = -\partial_0 \vec{A} - \vec{\nabla} A_0$ and that in α_{ij} with the magnetic field $\vec{B} = \vec{\nabla} \times \vec{A}$ where, taking the convention in Jackson,² $\mathcal{F}_6 = \mathcal{F}\alpha_{\mu\nu} = E_i \alpha_{i0} - B_i \alpha_{jk}$. The standard electric field then maps to the set of three ordered right-handed space-time unit elements $\alpha_{10}, \alpha_{20}, \alpha_{30}$ and the magnetic field to the terms $\alpha_{23}, \alpha_{31}, \alpha_{12}$ respectively. It should be noted that there is a possibility, in the full theory to be developed below, that the 4-trivector may also yield field components. Here, the (electromagnetic) field is associated with the bi-vector form, whatever its origin.

Over each of the sixteen multivector-quantities defined above, a general dynamical multi-vector field G is defined over a scalar term $P\alpha_P$, a vector term $A\alpha_\mu$ a field term $F\alpha_{\mu\nu} = E_i \alpha_{i0} - B_i \alpha_{jk}$, a tri-vector term $T\alpha_{\mu\nu\rho}$ and an eventual quadri-vector potential $Q\alpha_{0123}$ such that: $G_{16} = P\alpha_P + A_0\alpha_0 + A_i\alpha_i + E_i\alpha_{i0} - B_i\alpha_{jk} + T_k\alpha_{0ij} + T_0\alpha_{123} + Q\alpha_{0123}$. In an obvious notation, the constant terms are defined as $C = C_P\alpha_P + C_0\alpha_0 + C_i\alpha_i + C_{i0}\alpha_{i0} - C_{jk}\alpha_{jk} + C_{0ij}\alpha_{0ij} + C_{123}\alpha_{123} + C_Q\alpha_{0123}$.

Writing, by analogy with the form of the Maxwell equation $\mathcal{D}_4 F_6 = J_4$, $\mathcal{D}_4 G_{16} = C_{16}$ and using the conventional 3-space patterns for reference, one obtains from the odd terms a set of generalisations of the Maxwell

equations as:

$$\alpha_0(\vec{\nabla} \cdot \vec{E} + \partial_0 P) = C_0 \alpha_0 \quad (6)$$

$$\alpha_{123}(\vec{\nabla} \cdot \vec{B} + \partial_0 Q) = C_{123} \alpha_{123} \quad (7)$$

$$\alpha_i(\vec{\nabla} \times \vec{B} - \partial_0 \vec{E} - \vec{\nabla} P) = C_i \alpha_i \quad (8)$$

$$\alpha_{0ij}(\vec{\nabla} \times \vec{E} + \partial_0 \vec{B} + \vec{\nabla} Q) = C_{0ij} \alpha_{0ij} \quad (9)$$

and four further equations in the even terms:

$$\alpha_P(\vec{\nabla} \cdot \vec{A} + \partial_0 A_0) = C_P \alpha_P \quad (10)$$

$$\alpha_{0123}(\vec{\nabla} \cdot \vec{T} + \partial_0 T_0) = C_Q \alpha_{0123} \quad (11)$$

$$\alpha_{i0}(\partial_0 \vec{A} + \vec{\nabla} A_0 + \vec{\nabla} \times \vec{T}) = C_{i0} \alpha_{i0} \quad (12)$$

$$\alpha_{jk}(\partial_0 \vec{T} + \vec{\nabla} T_0 - \vec{\nabla} \times \vec{A}) = C_{jk} \alpha_{jk} \quad (13)$$

These represent a set of linear coupled differential equations describing the flow of square-root energy density (vot) in various space-time forms. Which equations are relevant in any context depends on which, if any, of the new dynamical terms and the new constants are non-zero. The first set of four equations correspond clearly to the Maxwell equations. The main difference is the appearance of two new dynamical terms P and Q transforming (under a Lorentz transformation) as invariant masses. The constant terms on the right correspond to the electric and magnetic charge and current. Setting P and Q zero, but setting C_0 to the charge and C_i to the current density these four equations are just the standard, inhomogenous Maxwell equations. Note that, in this limit, all four Maxwell equations are present at once in the present formalism, with all the correct signs, in contrast to the usual derivation not using the principle of absolute relativity.² It is the clear correspondence in the form of the equations that justifies the identification of the electromagnetic field as b-vot.

In the second set of four equations, provided the constants on the right are zero, the main new feature is the presence of the tri-vector terms T . If this term is not identically zero, this may be expected to be of value in understanding the underlying nature of angular momentum in particles. The new equations express (potential) degrees of freedom. With $C_P = 0$, equation (10) is just the Lorenz gauge condition and one obtains other conventional gauges by setting this constant to other quantities. In other words, C_P non-zero expresses a gauge degree of freedom. Here, there are other constants which, if expressed, would introduce new physics. In particular, the dual term C_Q expresses a further degree of freedom. In principle, the odd set and the even set should both constrain the physics, but the even set will not be used in deriving the main results of this paper. Indeed, the main results here will be obtained by demanding that allowed light-speed travelling wave solutions satisfy the standard set of free-space, field only Maxwell equations alone ($\mathcal{D}_4 F_6 = 0$). The full set is given for reference and will be used here only in developing the linear, first order wave-functions in the construction of the solutions. Note that, with $T = 0$ the final two equations are just the standard expression for the electric and magnetic field in terms of the vector potential. The new set of equations is the first result of this paper.

Within the formalism, the physical effect of the new term P is to allow a curvature of the momentum transport direction. If non-zero, this leads to the possibilities of a pivoting of the field flow around the mass leading to new kinds of self-confined circulating solutions with rest-mass.^{9,10,13} These solutions may underpin the underlying quantised nature of charge.¹⁶ The possibility that this new framework constitutes a new general, linear theory of light and matter, treating leptons and photons on the same footing, will be explored further in a companion paper.⁵ Here, all the constant terms on the left will be set to zero. This corresponds to the free-space (Lorenz gauge) condition for the conventional Maxwell equations alone. This is the appropriate framework for the description of the photon.

3. A NEW PHOTON WAVE-FUNCTION

Using the algebra new kinds of wave function may be generated with properties more strongly constrained than is possible conventionally. There are a plethora of such solutions. Complexity has arisen rapidly from the simplicity

of the four-dimensional basis due to the properties ascribed to multiplication and division. As a prequel, it is worth exploring these new solutions in general.

Conventionally, one often writes wave functions introducing a complex scalar i and exploiting the property that:

$$e^{i\theta} = e^{\theta i} = (e^i)^\theta = (e^\theta)^i = \cos(\theta) + i\sin(\theta) \quad (14)$$

The ordering and nesting of the exponents in the equation above is unimportant for complex numbers, as all factors commute, but will prove crucial in the more complex discussion to follow. As is well-known, a non-relativistic wave function propagating in the z direction may be written:

$$F_{NR} = Ae^{i(kz-\omega t)} = Ae^{i(kz-\omega t)} = Ae^{i\theta} \quad (15)$$

Where k is the wavenumber in z in m^{-1} and k is the spatial frequency in s^{-1} , ω is the temporal angular frequency. A is a factor setting the normalisation i is the unit imaginary and θ is a scalar, real number factor. Such forms have wide practical application. They may be further modified by well-known generalisations of the harmonic functions, Bessel functions, spherical harmonics, half-integral Legendre polynomials and so on to describe exponential-like solutions in cylindrical, spherical and toroidal systems. Despite their power and elegance, they have one major flaw if one wishes to use them in a relativistic theory: space and time appear in the combination $(kz - \omega t)$ as a (Lorentz) scalar factor. Since space and time, however, transform differently under a general Lorentz transformation, such a wave function does not reflect the differences between space and time properly. Such wave-functions do not, therefore, conform to the principle of absolute relativity. Rectifying this by imposing the proper relativistic transformations of space and time in absolute relativity at all levels, most especially in the exponent, leads to qualitatively different solutions exhibiting a kind of quantisation. This section will develop the general case, the next- the photon.

In the present formalism, there is no single simple complex scalar i , but several quantities which may play the same role in describing travelling wave solutions. Three of the base unit elements and seven of the unit elements derived from these square to negative unity^{4,8} (explicitly these are: $\alpha_1, \alpha_2, \alpha_3, \alpha_{12}, \alpha_{23}, \alpha_{31}, \alpha_{012}, \alpha_{023}, \alpha_{031}, \alpha_{0123}$). Any of these may be used to describe travelling waves. For example, by analogy with complex numbers one may expand an exponential with α_{12} , corresponding to a rotation of angle θ in the 12 plane as:

$$e^{\alpha_{12}\theta} = \alpha_P \cos(\theta) + \alpha_{12} \sin(\theta) \quad (16)$$

In the physical association made above, this would describe an oscillation back and forth between a rest mass component (α_P) and a magnetic field component (α_{12}). Such a formalism is descriptive in a similar way to complex numbers. Using the scalar α_P and α_{0123} alone provides an even more precise parallel, since the sub-algebra containing this pair only is isomorphic to complex numbers. Such exponents will be denoted in general as hypercomplex exponents. Though this may sound like some progress, eq. (16) retains the problem of general covariance alluded to above and such solutions are not necessarily proposed as representing a physical process (governed by a 4-vector derivative) as the proper elements corresponding to neither space nor time are present. To describe physics, they would require a bi-vector or quadri-vector derivative in an angular measure to operate. This would retain the essential feature of the deficiency sketched above, that space and time are treated identically. They serve merely to point the way to proper progress. Note that α_P has been used in the expansion above. This is because energy conservation considerations require that any physical wave should transform between elements of substance constituting, at the very least, an equal integrated energy. Here, local microscopic conservation of energy (and momentum), is ensured in that any transformation of vot from one space-time form to another is unitary.

To make progress, a second extension is required such that elements may be nested with each other, leading to a combined motion observed as a wave. This leads to a far richer structure than is available in a merely complex algebra. The new axiom requires the inclusion of the proper (in the Lorentz sense) relative transformation properties of space and time directly into the exponential. This is implemented by associating the proper unit element directly with the appropriate propagation direction in the hypercomplex exponent, as is now shown. For

example, for propagation in the 3 (z) direction, solutions are sought for some appropriate unit element, denoted $\alpha_?$, of the form:

$$F_{SR} = Ae^{(\alpha_3 kz - \alpha_0 \omega t)\alpha_?} \quad (17)$$

Absolute relativity is then imposed by the mathematical form in the above in that space and time have their proper form (here: α_3 for the z component in Cartesian space and α_0 for the time). By analogy with the expansion of real exponentials to complex exponentials, an element $\alpha_?$ is required to convert this to a travelling wave-function. Note that though equation (17) is a conceptual extension, it also embodies a physical restriction in that the factors corresponding to space and time are forced to have their proper relative form.

For a wave-like overall solution and by analogy with the properties of complex numbers, $\alpha_?$ is required to be some unit element which ensures that both the spatial and the temporal element of the development is governed by a unit element squaring to negative unity. Within the principle of absolute relativity it is axiomatic that space and time, and any other quantities such as angular momenta, should appear everywhere with their proper form.

By inspection, it is apparent that substituting neither α_P , nor the real number 1 for $\alpha_?$ leads to a travelling wave solution. In both cases the temporal development will square to positive unity, leading to falling exponential-like solutions rather than waves. For a given propagation direction there are six unit elements $\alpha_?$ which convert both space and time to a travelling wave-form, each leading to new kinds of wave-particle solutions. For the particular case of the 3 direction in eq. (17), these are explicitly: $(\alpha_{012}, \alpha_{23}, \alpha_{31}, \alpha_{123}, \alpha_{10}, \alpha_{20})$. The first three themselves square to negative unity; the second three to positive unity. This means the first three afford the possibility of inserting a scalar phase factor, in harmony with the multi-vector component, into the hypercomplex exponent. Of these, α_{012} (corresponding to the introduction of a unit vector in the direction of the angular momentum) leads to the possibility of substituting for A in equation (17) above, a pure field solution. This is presented below. The corollary to the principle of absolute relativity here is then that travelling wave solutions in space and time require a unit angular momentum in order to propagate. Such solutions are then a first order solution of the Maxwell equations - describing all six components of the electromagnetic field in any proper frame - but they are necessarily associated with a unit angular momentum.

Equation (17), with A scalar and $\alpha_? = \alpha_{012}$, though it represents a wave, is not itself of solution of the free-space Maxwell equations $\mathcal{D}_4 F_6 = 0$ as it contains terms transforming as a rest-mass (P-vot and Q-vot) as well as field terms. It is, however, a solution of the more general set of dynamical equations with the constant terms zero. That is: $\mathcal{D}_4 F_{SR} = 0$, as is readily verified by substitution. It is tempting to associate such a wave-function directly with a massive source particle such as an electron. This is partially true, but this form is still too simple to fully encompass the complexity of such particles as the scalar pre-factor is far too simple.^{5, 10, 16}

It is worth noting in passing that, of the other five elements transforming the above to a travelling wave, α_{23} and α_{31} also lead to wave solutions of the general equation $dG=0$, and these may correspond more closely to elements of electron-like and positron-like solutions. The remaining three possibilities may also be associated with light-like and particle-like solutions and may indeed be the primary- initial or lightest- solutions. Curiously, the dual bivector pair $(\alpha_{10}$ and $\alpha_{20})$ do not lead by themselves to a magnetic monopole-like but also to electric monopole-like fields, as can readily be verified by substitution and expansion. Magnetic monopoles may be described, however, by introducing more complicated terms involving a product with the pre-factor A . The α_{123} case may be associated with a precursor to the electron-positron pair in the creation process as it resembles a twisted-mode solution, the solution obtained by overlapping counter-propagating right-right or left-left circularly polarised light. This is the configuration for the creation of a particle-antiparticle state at spin zero at sufficiently high energy. It is also possible (by choosing an appropriate pre-factor and/or the relative propagation direction of space and time) to associate these with the primary photon-like and electron-like solutions. This opens up the possibility that the other set, which may introduce mass through the scalar term, may be involved in the description of the weak interaction. The further development of these speculations will be left to future work.

In popular expositions of relativity, one often talks of rulers and clocks, these often being held by idealised "observers". In reality, there are no observers, only the emitting and absorbing particles or systems and the intermediating photons themselves. Confusion is often introduced in arguments by ascribing "knowledge" to an external observer that it could not possible have.

Consider an idealised system in three frames: an emitter, an absorber and an intermediating photon. Each frame has its own scale of space, time (\mathcal{T}) and frequency. Its own “rulers” and ”clocks” and its own scale of energy or, equivalently, frequency. In a general “event” where a photon is exchanged between two particles, the particles may be in very different Lorentz frames. These frames, and hence their scales, will also change due to the effect of the exchange. Let the scale-change with respect to the centre of momentum frame be denoted by a single factor which sums up the relativistic transformation R . For example, a photon in one frame (e.g. the emitter) may have a particular energy, frequency and wavelength. In the (e.g. blue shifted absorber) frame where the energy (and hence the frequency) increases by R , the wavelength decreases by $1/R$. It is enlightening to write this scale-factor R in terms of the usual relativistic β and γ factors:

$$R = \sqrt{\frac{1+\beta}{1-\beta}} = \gamma(1+\beta), \quad 1/R = \sqrt{\frac{1-\beta}{1+\beta}} = \gamma(1-\beta), \quad \omega' = \omega R, \quad \lambda' = \lambda/R \quad (18)$$

With:

$$\beta = v/c = \frac{R^2 - 1}{R^2 + 1}, \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{2}\left(R + \frac{1}{R}\right) \quad (19)$$

Note that the last relation above means that the gamma factor is, physically, the average of the increase in energy of the light travelling against the motion, with that travelling with the motion. One may conclude that the Lorentz scaling of the mass of material particles is just that of the energy of light in a light-tight box.¹¹ Further, provided the magnitudes of the electric and magnetic field components are equal (as they are for propagating free-space electromagnetic waves in general and for photons in particular), so that $|E| = |B|$, they transform relativistically as:

$$E' = \gamma(E + \beta B) = RE(= RB), \quad B' = \gamma(B + \beta E) = RB(= RE) \quad (20)$$

That is, for light, the fields transform in the same way as does the frequency and energy: linearly with R . Relativity of space and time is that transformation forcing both integral energy and local field to add linearly.

The formalism to write down a new, fully relativistic solution to the first-order Maxwell equation in free space ($\mathcal{D}_4 F_6 = 0$) is now complete. For the simple case of a propagating free-space electromagnetic wave, forcing z and t to take the proper form α_3 and α_0 respectively, a single photon solution of a left circularly polarised electromagnetic wave, travelling in the the $+z$ -direction and transmitting a quantum of energy \mathcal{E} in the centre of mass frame may be written:

$$F_L = H_0 U_F R \mathcal{E} (\alpha_{10} + \alpha_{31}) e^{\frac{\mathcal{E}}{\hbar} R (\alpha_3 z - \alpha_0 t) \alpha_{012}} = F_0 R (\alpha_{10} + \alpha_{31}) e^{R (k \alpha_3 z - \omega \alpha_0 t) \alpha_{012}} = \mathcal{F} \mathcal{W} \quad (21)$$

This has a pre-factor part representing the initial (or final) field configuration $\mathcal{F} = F_0 R (\alpha_{10} + \alpha_{31})$, and a hypercomplex exponential wave-function part $\mathcal{W} = e^{R (k \alpha_3 z - \omega \alpha_0 t) \alpha_{012}}$. Note again the units of z here are seconds.

The real-number constants c are the (scalar) speed of light and $\mathcal{E} (= \hbar c k = \hbar \omega)$ the (scalar) quantum of energy transmitted in the centre-of-momentum frame respectively. U_F is a universal constant, taking the same value for all photons of energy $\hbar \omega$, converting to field units. H_0 is a distribution function representing the spread of field or energy over phase, whose square integrates to unity. This is an invariant and is the same in all frames, right up to the limit of light-speed where the integrated energy goes to zero. The single parameter R is that factor which determines the scales of energy, frequency, length and time for the same photon in any Lorentz frame. Taking as a reference that R is unity in the centre of momentum frame, then \mathcal{E} determines the proper magnitude of the energy-momentum transmitted. The factor $\mathcal{E} R$ is then the energy in each relevant frame. The factor of R in the exponent pertains to the relativistic transformation of “rulers” and “clocks”. The proper reference “ruler” and “clock” for any given photon exchange event scales with the centre of momentum energy. This is just the wavelength and frequency of the photon in the proper (centre of momentum) frame. Note that, in equation (21), it is not the time that is primary, but the energy. This corresponds more properly to an inverse time, the frequency. Inverse time is in the same direction as time, as in equation (2), but with inverse magnitude and units. For a photon exchange over appreciable distance, the energy is fixed, the time over which this occurs,

the number of cycles in phase, is indeterminate and flexible. There is a sense in which, at least to a constant and a factor of 2, energy *is* frequency. The factor of R in the pre-factor corresponds to the proper relativistic transformation of fields in the emitter and absorber frames. Both field and frequency scale linearly with R . In the centre of momentum frame the proper frequency is the energy divided by Planck's constant $\omega_0 = \frac{\mathcal{E}}{\hbar}$. Note that the energy density in the field is proportional to field squared. Explicitly, $D(\mathcal{E}) = \frac{1}{2}(E^2 + B^2) = \frac{1}{2}\epsilon_0(\mathcal{E}^2 + c^2\mathcal{B}^2)$. This means the wave-function in equation (21) may be converted to square-root energy density units by the simple expedient of defining another universal factor $U_{\sqrt{\mathcal{E}}}$ in place of U_F . This, then, is a fully relativistic wave-function giving the root-energy “probability density” in any desired frame. This raises the question of why the factor for energy appearing in the pre-factor should be \mathcal{E} and not $\sqrt{\mathcal{E}}$. The reason is that, experimentally, both energy and field must add linearly. It is then space and time themselves which must deform, relativistically, to accommodate these deeper principles of linearity. That is exactly what the Lorentz transformation achieves, as has been sketched above. The Lorentz transformation is exactly that transformation which linearises the addition of both energy and field in the allowed solutions. If the scaling factor is 2, for example, the relativistic transformation is such that the field doubles, the energy density quadruples, but the length of the wave train in the new frame halves (due to the observed deformation of space relativistically), giving a linear increase in the total energy with frequency overall, as is observed experimentally. It is the number of cycles of phase which remains the same in all frames-not the length or the time. It is the stringent constraints of linearity of energy and of field, together with the variation of space and time in special relativity, that force allowed solutions to be only those where the frequency in the exponent corresponds to the proper magnitude of the energy in the pre-factor. One can fill in any value of energy from radio-waves to high energy gamma photons to the wave-function above - but this must affect both exponent (frequency) and pre-factor (field) proportionately. Although absolute relativity was taken as an ansatz here, it may be seen as a consequence of the deeper requirements of the linearity of energy and field.

The wave-function \mathcal{W} alone is not, strictly, a solution of the conventional Maxwell equations, as it contains terms transforming as an invariant mass as well as terms transforming as fields. It is, however, a solution of the more general set of equations discussed above such that $\mathcal{D}_4\mathcal{W} = 0$. For the pure-field cases observed to exist physically, in the product $F_L = \mathcal{F}\mathcal{W}$, the mass terms cancel leaving pure field alone, so that the whole expression is then a solution of the free-space Maxwell equations $\mathcal{D}_4F_L = 0$. What this means, physically, is that, \mathcal{W} may be an element of the proper relativistic wave-function of the (rest-massive) emitter and absorber. If so, such physical functions may combine with initial and final fields such that they effectively propagate a pure field packet of arbitrary total energy. Such fields are propagated at light-speed only if the initial (emitter) and final (absorber) fields have equal and perpendicular magnetic field components and are such that the total energy in their frames (note carefully that these are usually different) is proportional to the local wave-function frequency ν_l (where the suffix l denotes the local frame under consideration). Denoting emitter e absorber a and photon p , one needs to consider the concept of locality for the same photon in three frames. This is a coherent tri-locality. The locality of the emitter, the locality of the absorber and the locality of the exchanged photon. The photon emerges from the emitter with energy $\mathcal{E}_e = \hbar\nu_e$. For the absorber, the same photon arrives with energy $\mathcal{E}_a = \hbar\nu_a$. For the co-moving (at least nearly lightspeed) frame of the photon, the energy $\mathcal{E}_p = \hbar\nu_p$ tends to zero, the wavelength tends to infinity, but the transformations of absolute relativity place each phase value in H_0 at (very nearly) the same point in space-time for both the emitter and absorber. Locality, for the photon, may span vast tracts of space and time. The absorber “sees” a packet of positive energy arriving from its past. The emitter “sees” a packet of negative energy leaving into its future. This is the same photon. Positive energy backwards in time is the same thing as negative energy forwards in time. In any event, both happen. Causality is in the direction of the energy transmitted. From emitter to absorber. That fixes the “arrow of time” (more properly the “arrow of frequency”) for each event. The picture is of a resonant, coherent, smooth exchange of a packet of mass energy characterised by the local frequency alone. This is exactly what is observed experimentally.

Elements of the application of the principle of absolute relativity have appeared seven times in constraining the form of equation (21). Space has been inserted with its proper form (z here) in the direction of propagation direction as α_3 . Time appears associated with its proper unit vector α_0 . The proper “direction” of the angular momentum around the propagation direction α_3 is α_{012} . The electromagnetic field must obey strict constraints in order for equation (21) to be a solution: the starting fields must be perpendicular, E_x and B_y here in proper “directions” α_{10} and α_{31} and they must be of equal magnitude; otherwise the rest-mass terms do not cancel to

leave a zero rest-mass pure field capable of light-speed propagation. The (scalar) energy must then appear twice to be consistent with the relativistic laws: once in the pre-factor (expressing the linearity of field under a Lorentz transformation with R) and once in the exponent (expressing the linearity of energy and its proper transformation relativistically). If all of these conditions are satisfied, the resulting propagating part is necessarily quantised, according to equation (21).

Equation (21) is a universal, relativistic, wave-function describing photons of any energy. Varying \mathcal{E} from zero to infinity one obtains any photon of any energy, all with the same angular momentum. Conversely, the same photon, viewed from different frames characterised by relativistic scaling factor R , will appear to have different frequencies and energy (right down to the zero-energy limit). In each frame the angular momentum is the same, as the energy scales up with R , the sizes scale inversely. One can, as a photon emitter or absorber, be in any arbitrarily blue or red-shifted frame, the wave-function takes the same form, differing only by the change in the relative scale of frequency, energy and length as described by the relativistic scaling factor R . In every frame, however, these properly relativistic wave-functions are necessarily quantised with a characteristic frequency ν_l and an integral energy given by $\mathcal{E}_l = h\nu_l$. The new, fully relativistic, quantised wave-function equation (21) is the main result of this paper.

Usually one looks for solutions of the second-order equations eliminating one of the fields in favour of the other. Equation (21) is a solution of the first-order equation $\mathcal{D}_4 F_L = 0$ directly. It describes both electric and magnetic fields and the transformations between them. The development of these in time and space parallels the field transformations expressed by the Maxwell equations.

It is also worth noting that, in contrast to conventional wave-functions of the form of equation (15), the development of the wave-function in space and time in equation (21) is different. In space one has a wave, in time one has a rotation. As one moves in space, there is an alternation between electric and magnetic field - just as described by the first-order Maxwell equations. Sitting at one point in space, and allowing time to pass, one has a rotation of the field vectors. The new wave-functions, in and of themselves, parallel the physical development of the field components in the Maxwell equations more closely than do conventional solutions.

To make the connection with the form encountered in standard textbooks,² equation (21) may be readily expanded in any particular frame. For the conditions corresponding to experimentally observed photons, the non field (scalar and quadri-vector) terms in the exponential part cancel. Setting $F_1 = H_0 U_F R \mathcal{E}$ and $ck = k = \omega = \frac{\mathcal{E}}{\hbar}$ on expanding one obtains:

$$F_L = F_1[(\alpha_{10} + \alpha_{31}) \cos(kz - \omega t) + (\alpha_{23} - \alpha_{20}) \sin(kz - \omega t)] \quad (22)$$

This describes electric (α_{i0}) and magnetic (α_{ij}) fields rotating in time in a plane perpendicular to the direction of momentum transport and transforming in space from magnetic to electric and vice-versa. The resultant field configuration is that shown in Fig. 1. It is identical to that found in any elementary textbook on electromagnetism for a left-handed circularly polarised wave. This is comforting: provided the constraints are satisfied, despite its apparent complexity, the new wave-function expands to that form measured in experiments and familiar from elementary textbooks.

4. DISCUSSION

The new wave-function is consistent with the experimentally-observed field pattern for a photon. The extent to which it truly describes a light quantum, a photon, is now discussed.

Firstly, consider that if equation (21) is a solution, then the linearity of field addition and the condition of energy conservation require that light should come in “lumps”. Proof: consider an emission-absorption event in the frame of the photon. Now consider what happens if one superimposes a second such solution, where both overlap precisely in phase and wave-train length, such that twice the energy is transmitted at the same frequency in a single overlapping event. In this case the fields add everywhere. Since the energy density goes as the field squared, this would give four times the energy density everywhere and hence four times the energy transmitted. This violates energy conservation and hence such a process is, at least to first order, “not allowed”. This is similar to the argument proposed in earlier work to explain the origin of the exclusion principle.¹³ Viewing the same photon from a different frame scales energy density by R^2 , but length by R^{-1} leading to a linear increase in energy

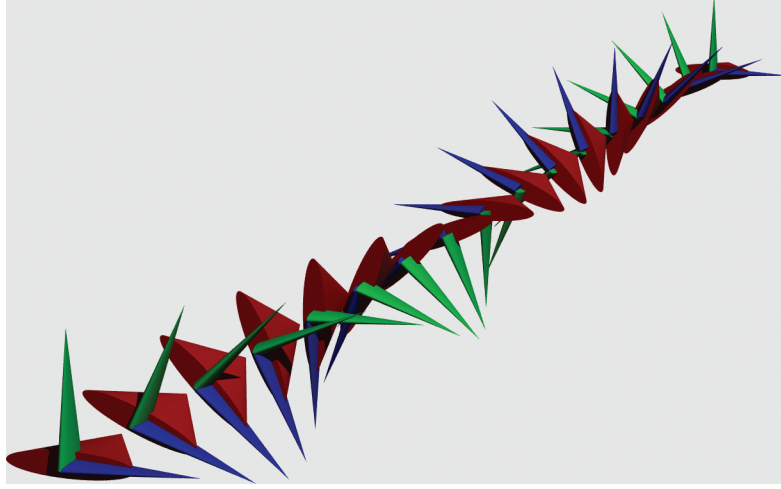


Figure 1. Representation of a single wavelength of a circularly polarised photon of equation (21). The electric field direction is represented using green arrowheads, the magnetic field blue and the momentum density($E \times B$) red.

overall. The energy may be increased in a single event, self-consistent with both the linearity of energy and field and with relativity, is to increase the frequency. This is tantamount to varying the factor “ R ” in equation (21), affecting the frequency, the field and the overall energy by the same factor. This process gives a linear increase in both energy and field, as is required by the relativistic transformation of the solution and by experiment. This is the primary reason why the new first order, relativistic expression of equation (21) is necessarily physically quantised. The quantisation of light, far from being in conflict with relativity, is here a consequence of it. Note that, conversely, relativity itself may be viewed as that transformation required to ensure the linearity of both energy and field as expressed by equations 18, 19 and 20 above. Manifestly, for such wave-functions, the energy scales with frequency ν as this appears both in exponent and pre-factor. Clearly, in this relation h must be identified with the constant of Planck such that $\mathcal{E} = h\nu$. In other words, the Planck constant defines the scale of length for any given photon wave-function of the form of equation (21) with characteristic energy \mathcal{E} . For a given proper energy, it sets the scale of rulers and clocks for that event. One may have different wave-lengths, but then one must also have correspondingly different energies - just as is observed. This is the key result of this paper: the quantisation of allowed solutions of the continuous theory is a consequence of the experimentally-observed conservation of energy and the linearity of field. It is worth noting that, to second order it may be possible to have double or triple photons. These would then have the quantisation rule $\mathcal{E} = hn^2\nu$, with n a natural number. Such photons (Phat photons) may already have been observed, as has been discussed by Williams.¹⁷

Secondly, note that though the fact of quantisation of the kind of solution represented by equation (21) has not, here, required the introduction of a differential operator, a calculation of the value of the constant of proportionality between energy and frequency (Planck’s constant) does. Charge appears at the level of the vector, and the field at its differential, bringing in a factor of R such as that in the pre-factor of equation (21). The question is then: can an expression be found relating the value of the elementary charge and that of Planck’s constant in the present formalism? Such an estimate requires a study of the internal dynamics of the emitting and absorbing system and this is beyond the scope of the present paper. In earlier work, however, a consideration of a simple semi-classical model of the electron as a localised photon did lead to such a relation.¹⁶ This leads to an estimate for Planck’s constant in terms of the elementary charge in that model of $\hbar = 1.27 \times 10^{-34} Js$ which is, at least, of the right order of magnitude. The quantisation of light does not arise from the field equations, but from the nature of the emitter and absorber - usually atoms or molecules. This opens up the possibility that light with no quantisation, or different kinds of quantisation may exist, as will be discussed in the sequel.

Thirdly, note that the field development and transformation parallel the Maxwell equations more closely than do more conventional solutions. The microscopic development of the field components is not merely a rotation.

In equation (21), as one progresses forwards in space, the field elements in the solution transform back and forth between electric and magnetic field components, just as in the case of equation (8) and equation (9). Although equation (22) looks just like a conventional simple electromagnetic wave-function, the underlying origin of the elements of electric and magnetic field, as described by eq. (21) is back and forth between each other, just as described by the Maxwell equations. Nonetheless, for a fixed position in some frame, as time progresses the field components appear to rotate (or oscillate), just as is the case for physical sources such as incandescent lights or a transmitter. In these respects, the new solutions match not only what is observed, but also parallel more closely the underlying field transformations of the Maxwell equations.

Fourthly, consider the relativistic transformations of space, time and field. From the underlying axiomatic construction, the internal elements of space and time retain their proper relative form. This leads to the correct transformation properties of all components under a general Lorentz transformation.¹² In any other frame the proper relative transformation of the spatial and temporal field components is ensured in that they are constructed in such a way as to differ by the proper unit element from each other. In these solutions space and time transform properly with respect to each other. The result is that the transformed solution remains a solution in any proper Lorentz frame, right up to the limit of lightspeed. Put simply: all elements in the solution, both exponent (energy) and pre-factor (field) must scale linearly with R . Conversely, the Lorentz transformation is that transformation which ensures the linearity of both energy and field addition. Demanding that both energy and field should add linearly requires the introduction of the principle of absolute relativity. Although the new principle was taken here as an ansatz, it appears that the principle of relativity of space and time is required in order to be consistent with the deeper principles of the conservation of energy and the linearity of the field. Absolute relativity is then not a postulate, but a requirement.

Fifthly, in any extension where ω and k are not precisely equal, the separation between the space and time oscillations allows an identification with the two-phase harmony of de Broglie which lies at the root of quantum mechanics.^{14, 15} This corresponds to an extension to quantum particles with rest mass, as has been discussed elsewhere.^{9, 13, 16}

Sixthly, if we demand further that the exponent should be consistent with a scalar wave-function, at least at the points of emission and absorption, such that it matches such wave-functions (of the form of equation (15), for example), then to achieve this, the factors in the exponent ($\alpha_3 kz - \alpha_0 \omega t$) and α_{012} must commute. In particular, this requires that α_{012} commutes with the factor for the wavenumber k . That is: to match simultaneously a wave function constrained by the principle of absolute relativity such as equation (21), to a more conventional one where the exponent acts as a scalar factor such as equation (15), then the exchange process as a whole must be black-body quantised. This is only the case if that wavenumber corresponds to an integral number of half-wavelengths. This extra condition corresponds then to that of black-body quantisation.

Seventhly, consider simple transformations of the solution proposed in equation (21). Changing the sign of one component of the pre-factor alone, for example $(\alpha_{10} + \alpha_{31})$ to $(\alpha_{10} - \alpha_{31})$ has an interesting effect. This is no longer a left-handed solution for a wave propagating in the positive z direction, but a solution for a right-handed photon travelling in the negative z direction. That is, such a transformation matches precisely the physical process of reflection: the handedness of the field with respect to one another matches the direction of momentum transport. A change in the relative handedness of the electric and magnetic field components reverses the propagation direction (and flips the helicity). In other words, just as observed in experiment, the relative handedness of the electric and magnetic field components determines the direction of propagation. As discussed in the case of black-body quantisation above, changing the order of the unit angular momentum factor in the exponent from $(\alpha_3 kz - \alpha_0 \omega t)\alpha_{012}$ to $\alpha_{012}(\alpha_3 kz - \alpha_0 \omega t)$ is not a solution to the Maxwell equations. Indeed, for $k = \omega$ it is a frequency doubled oscillation. Though this is not a solution for the Maxwell equation, in the case of the electron-positron annihilation it does correspond precisely to the internal zitterbewegung frequency of the fermions as described by the Dirac equation.^{3, 8} It is tempting, then, to identify this double-frequency solution with an electron wave-function. This is not so, the solution remains too simple. For a description of a massive particle the solution must, at the very least, follow periodic boundary conditions such as those described in the simple semi-classical model considered in earlier work.¹⁶ If this is done, this may give a description of purely electromagnetic charged particles with half-integral spin.¹³ Simply changing the sign of the exponent remains a solution, but has the physical effect of transforming from left-handed to right-handed or vice-versa. Thus a

linearly polarised photon may be represented as a sum or difference of such solutions. For the particular solution proposed, these are “x” polarised and “y” polarised respectively. Elliptical polarisations may be obtained from a linear combination in the usual way.

Eighthly, the expansion to a full 4-dimensional wave-function introduces, necessarily, a limited extent perpendicular to the propagation direction. In an obvious extension, replacing the exponent with $(\alpha_1 kx + \alpha_2 ky + \alpha_3 kz - \alpha_0 \omega t)\alpha_{012}$ leads to the perpendicular x and y components having an expansion in terms of cosh and sinh instead of cos and sin. Explicitly, these are $F_0(\alpha_P \cosh(kx) - \alpha_{20} \sinh(kx))$ and $F_0(\alpha_P \cosh(ky) + \alpha_{10} \sinh(ky))$ respectively. These terms do not describe propagating solutions. Propagation is supported only along a line joining emitter and absorber and not transverse to the photon path. Further, both sinh and cosh functions increase exponentially in magnitude for larger lateral values, a clearly unphysical condition. $\mathcal{D}_4 F = 0$ only if the expansion in the transverse direction is zero or constant. This confines the lateral, non-propagating dimension of the wave-function to lie close to the axis. In particular, some longitudinal components of field may be completely suppressed, since the sinh function is zero on axis. This may help to explain why the field of physical photons is primarily transverse. The scalar component, however, has a finite minimum on axis, and may supply a constant term. This term may prove to express the scalar mass-energy transferred by the photon from emitter to absorber.

Finally, the wave-function in equation (21) describes a temporal rotation in real space. This means the lateral extent in the photon frame should not exceed a rotation horizon imposed by the speed of light. This imposes conditions on the angular momentum of allowed solutions. The concept was used in previous work to lay bare the physical origin of the anomalous magnetic moment of the electron as a localised photon.¹⁶ For a given frequency the limit imposed by the speed of light on rotation about the photon axis, the rotation horizon, is just $r_h = \frac{c}{\omega}$. Introducing the photon momentum observed in experiment, $\vec{p} = \hbar\omega/c$, gives a limit on the integral allowed angular momentum of the solutions of $r_h \times \vec{p} = \hbar$. This sets the intrinsic scale of unit angular momentum for solutions such as that described by equation (21). The form demanded by equation (17) and manifested in equation (21) is not merely descriptive, it is strongly proscriptive. Demanding the principle of absolute relativity, manifested in the form of equation (17), places strong restrictions on allowed solutions, over and above those required by the Maxwell equations alone. Such a solution allows the energy to go to the very large (gamma) or very small (sub-radio) smoothly, while retaining the same integral angular momentum.

In conclusion, the properties discussed above mean that the new construction does not describe many non-physical combinations of propagating field, though it does allow those observed in nature. Equation (21) is only a solution if a strongly-constrained set of physical conditions are met. Crucial is that *both* the rate of change of phase *and* the field magnitude scale with R in both space and time. The magnitudes of the electric and magnetic field components must be equal, but are otherwise arbitrary. To be a solution of the free-space Maxwell equations, ω must equal k . The signs of ω and k must match the handedness of the field components and the scale of frequency must match the scale of energy in the pre-factor. Further, a factor corresponding to a unit angular momentum is required in the exponent to transform the proper relative form of the spatial and temporal elements to a travelling wave solution with pure fields alone. In the logical extension of the 2D case to the 4D case the lateral components do not propagate. Once again, the required physical conditions match those observed in the physical photon.

5. EXPERIMENTAL TESTS

The theory is already consistent with a great number of classic experiments. These include all of those well-explained by the Maxwell equations, as these are encompassed within the new theory. Further, as discussed above, the new theory explains the underlying physical origin of effects such as the quantisation of light itself which must otherwise be taken simply from experiment. Equally, it explains why, though transmitted light appears quantised, local fields remain continuous and are well-described by classical electromagnetism. It addresses both the quantum and the wave nature of light within a single framework.

Although the new theory remains closely consistent with a wide body of existing experiment this does not mean that it does not suggest new avenues for experimental investigation. The fact that the limits on the angular momentum of the photon may arise from the nature of emitter and absorber opens up an interesting set of experimental possibilities. It is possible that photons emitted and absorbed by collective states of matter,

such as in the superconducting or fractional quantum Hall regimes at low energy, or in regimes where fractional charges such as quarks may be present at high energy, may have different constraints on the limits of the photon angular momentum. It may therefore be possible to produce or detect photons with angular momentum a fraction or multiple of \hbar with energy proportional to this fraction or multiple squared. For example, quanta could be emitted with fractional angular momentum only if an absorber with fractional states were present. This would demonstrate that the photon quantisation arises from the emitter and absorber rather than the field itself.

6. CONCLUSIONS

An extended theory of electro-magnetism encompassing the Maxwell equations has been proposed. Solutions to this, such as equation (21), embody the propagation of coupled equal and perpendicular electric and magnetic fields along an axis perpendicular to both, with respect to which they are strongly constrained. The simultaneous imposition of the conditions of linearity of energy, field and the principle of absolute relativity require that the energy transmitted should come in “lumps” and that this energy is proportional to the frequency. The solutions are fully relativistic, describing the photon in the emitter, absorber and any other frame. A single constant, derived from the properties of emitter and absorber, means that all such solutions have the same angular momentum, so that this appears quantised. An extra boundary condition matching emitter, absorber and photon wave-function requires a black-body quantisation. Taken together, it may be argued that these features mean that the new wave-functions better represent the physical photon than do more conventional solutions.

6.1 Acknowledgments

Much of the theoretical basis of the work here has been carried out in collaboration with Martin van der Mark. The authors are grateful to Phil Butler, Tim Drysdale, John Duffield, Richard Gauthier, David Mathes and Michael Mobley for useful suggestions and to Nick Bailey, Nick Green and John Weaver for their constant encouragement. The figure was drawn using Povray. SJL acknowledges financial support from EPSRC.

REFERENCES

- [1] Maxwell, J.C., [A treatise on electricity and magnetism]. 3rd edition, Dover(1954).
- [2] Jackson, J.D., [Classical Electrodynamics]. Wiley, New York, (1975).
- [3] Dirac, P. A. M., [The Principles of Quantum Mechanics], Oxford U.P., London, (1958). 4th ed.
- [4] Lounesto, P., [Clifford Algebras and Spinors], *Lond. Math. Soc. Lecture Notes Series* **239**, (1997).
- [5] Williamson, J. G., “The nature of the photon and the electron”, companion paper to this one for August 2015.
- [6] Mead, Carver A., [Collective electrodynamics], MIT press (2002).
- [7] Penrose, R., [The Road to Reality], Jonathan Cape (2004).
- [8] Hestenes, D., “Zitterbewegung in Quantum Mechanics”, *Found Phys*, 40 (2010).
- [9] Williamson, J. G., “On the nature of the electron and other particles”, paper presented at The Cybernetics Society 40th Anniversary Conference, (2008).
- [10] Williamson, J. G., “A new theory of light and matter”. Paper presented at FFP14, Marseille, France, (2014).
- [11] M.B. van der Mark and G.W. ’t Hooft, “Light is heavy”, *Van A tot Q*, NNV (november 2000).
- [12] Leary, S.J., “Investigation of electromagnetism in a real Dirac algebra”. PhD Thesis, University of Glasgow, (2007).
- [13] Williamson, J. G., 2012 “Fermions from Bosons and the origin of the exclusion principle”, *Proceedings of MENDEL* (2012).
- [14] de Broglie, L., “Waves and Quanta”. *Nature* **112**, 540 (1923).
- [15] de Broglie, L., “Recherches sur la théorie des quanta”, *Ann. Phys. Ser. 10* **3**, 22 (1925).
- [16] Williamson, J.G., & van der Mark, M.B. “Is the electron a photon with toroidal topology?”, *Ann. Fondation L. de Broglie* **22**, 133 (1997).
- [17] Williams, Pharis E., “Phat photons and Phat lasers”, *Proc. SPIE 8832, The Nature of Light: What are Photons?* V, 88320D (2013).