Reinterpreting the Solutions of Maxwell's Equations in Vacuum

V.A.I. Menon , M.Rajendran, V.P.N.Nampoori
International School of Photonics, Cochin Univ. of Science and tech., Kochi 680022, India.

The authors show that the solutions of Maxwell's equations in vacuum admit electromagnetic waves having oscillations not only in electromagnetic field but also in spatial displacement. The fact that the existence of such a spatial component of the electromagnetic wave has remained hidden from observations can be directly related to the gauge invariance. It is shown that the existence of the spatial oscillations can be attributed to a new basic field which appears to contribute major part of the energy of the electromagnetic wave with the electromagnetic field accounting for the remaining. The authors hint that the new field may be identified with the Higgs field.

PACS numbers: 41.20jb, 42.25Bs, 03.65-w. 03.50 De

1. Introduction

We know that the beauty of the Maxwell's equations is that while they are very simple linear equations, all aspects of electromagnetism can be explained by them. Since we propose to confine ourselves to the study of the transmission of electromagnetic waves in vacuum, we shall study Maxwell's equations in vacuum given by

$$(i) \quad \nabla\cdot\xi = 0, \qquad (ii) \quad \nabla\times\xi = -\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{t}}.$$
$$(iii) \quad \nabla\cdot\boldsymbol{B} = 0\,, \qquad (iv) \quad c^2\nabla\times\boldsymbol{B} = \frac{\partial\xi}{\partial t} \tag{1}$$

We know for the electric field these equations have solutions in the form **[1]** (see Annexure)

$$\xi = \xi_o \, sin(\omega t - \boldsymbol{k}.\boldsymbol{r}) \ . \tag{2}$$

Similarly, for the magnetic field, the solutions may be written as

$$\boldsymbol{B} = \boldsymbol{B_o} \, sin(\omega t - \boldsymbol{k}.\boldsymbol{r}) \ . \tag{3}$$

Note that the magnetic field will always be perpendicular to the electric field and both will in turn be perpendicular to the direction of propagation. Another important point to be kept in mind is that the solutions represent not a single wave, but a wave front that has electric and magnetic fields as constants at any instant in the transverse direction.

2. Gauge Invariance and the Spatial Amplitude

This constancy of the electric field in the transverse direction (field in the direction of propagation is zero) at any instant results in its zero divergence at any point. In fact the electromagnetic wave has to be seen as progressing as a wave front, and not as a single wave progressing along a straight line. The progression of the wave may be represented by successive wave fronts in the direction of motion. Note that the wave front is assumed to extend to infinity in the transverse plane. This concept of the wave front and the energy contained by it cannot be understood clearly in classical terms. The picture becomes clear when we take up the quantum mechanical interpretation according to which the wave front is constituted by the points on paths at any given instant that the wave could occupy. This means that although there is only one wave, its progression can only be represented by succession of wave fronts in time.

An interesting aspect of this wave propagation is that it does not alter the picture even if we attribute spatial oscillations to the wave along with the oscillations in the electric (also magnetic) field. The idea that the electromagnetic wave has oscillations not only in the electromagnetic field, but also in space by way of displacement along the transverse direction

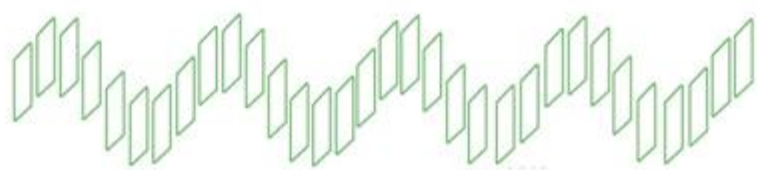

Fig.1. This shows the propagation of the plane wave in terms of successive wave fronts in the transverse direction.

may appear rather naive at first. But we shall show that such an idea is quite logical. We should keep in mind that the introduction of the transverse spatial oscillations will just shift the wave front slightly in the transverse plane which will have no effect on the electromagnetic aspects of the wave propagation as the wave front is assumed to have infinite spread in the transverse plane. If we take a wave front frozen at any instant, the electric field on it in any direction would remain constant. This would mean that the divergence of the electric field will be zero at any point on that wave front. Therefore the introduction of the spatial amplitude to the electromagnetic wave will still be consistent with Maxwell's equations.

It is reasonable to assume that the spatial oscillations propagating at luminal velocities constitute the basic wave that transports the electromagnetic oscillations and that both oscillations will be either in phase or with opposite phases. To put it differently, the electromagnetic wave could be pictured having its electric and the magnetic oscillations riding on the spatial wave. If we denote the spatial amplitude by $\boldsymbol{\eta_o}$, then we may represent the spatial wave as

$$\boldsymbol{\eta} \; = \pm\, \boldsymbol{\eta_o} \sin(\omega t - \boldsymbol{k.r}) \, . \tag{4}$$

Note that the introduction of the spatial wave above does not alter the situation represented by the solutions given in (2) and (3). The electromagnetic wave represented by (2) and (3) continues to represent the electric and magnetic fields in the transverse planes and their variation with time. In other words, the spatial wave defined by (4) is consistent with the Maxwell's equations.

We shall now relate the spatial amplitude of the electromagnetic wave to the currently understood gauge invariance of Maxwell's equations. We know that Maxwell's equations possess the gauge freedom in terms of vector potential and scalar potentials which satisfy the relations

$$A` \;=\; A + \nabla\chi \quad and \quad \varphi` \;=\; \varphi - \partial\chi/\partial t \tag{5}$$

where χ satisfies the wave equation

$$\frac{\partial^2\chi}{\partial^2 x} - \left(\frac{1}{c^2}\right)\frac{\partial^2\chi}{\partial^2 t} \;=\; 0 \tag{6}$$

Needless to say, (6) is the direct result of the introduction of the Lorenz gauge condition [2]

$$\nabla.\boldsymbol{A} - \left(\frac{1}{c^2}\right)\frac{\partial\varphi}{\partial t} = 0 \quad . \tag{7}$$

It can be easily shown that the values of the electric field and the magnetic field given by

$$\boldsymbol{\xi} = -\nabla\varphi - \frac{\partial\mathbf{A}}{\partial \mathrm{t}} \quad ; \quad \mathbf{B} = \nabla \mathrm{x}\mathbf{A} \tag{8}$$

are independent of χ. In fact, till recently it was presumed that only $\boldsymbol{\xi}$ and **B** represent whatever is observable in an electromagnetic field. The vector potential **A** and the scalar potential φ were not considered to be observable entities. However this assumption had to be changed after the discovery of Bohm-Aharonov effect [3].

We know that just like χ, the spatial displacement η also follows the wave equation. If we now relate η to χ by the relation

$$\eta = C\,\chi \tag{9}$$

where C is a constant, then we need not invent a new gauge freedom to explain the invariance of solutions of Maxwell's equations to the introduction of the spatial amplitude to the electromagnetic wave. Since "eχ" has the dimension of action, C will have the dimension of (LeA^{-1}) where L and A stand for the dimension of length and action respectively. We shall now try to obtain a new interpretation for the gauge invariance.

Let us study the case in more detail. If we represent the electric field of the electromagnetic wave by (2), then to comply with (8), the scalar and the vector potentials may be expressed as

$$\boldsymbol{A} = A_{\boldsymbol{o}}\, cos(\omega t - \boldsymbol{k.r}) \quad ; \quad \phi = \phi_{\boldsymbol{o}}\, cos(\omega t - \boldsymbol{k.r}) \tag{10}$$

Substituting (2) and (10) into (8), we obtain

$$\boldsymbol{\xi_o} = -(\boldsymbol{k}\phi_o - \omega \boldsymbol{A}_o) \tag{11}$$

In order to maintain compatibility with (5), (8), (2) and (10), we may express χ as

$$\boldsymbol{\chi} = \pm \boldsymbol{\chi_o}\, sin(\omega t - \boldsymbol{k.r}) \tag{12}$$

Based on this expression for the χ-wave, we obtain the desired result that the gauge transformation introduced in (5) modifies only the amplitudes of **A** and φ without altering their dynamical properties. This is what is expected in a gauge transformation. Now when we compare (12) with (4), we obtain the relation (9). The positive and negative signs on the right hand side of (12) mean that the oscillations of the electric field in the electromagnetic wave have a phase difference of either 0 or 2π with those of χ (and also η). It will be shown in a separate paper that these two values for the phase difference are of immense importance in determining nature of the electric charges of the electron and positron.

3. Structure of Photon

Initially when the concept of photon was introduced by Einstein, it was treated as the particle aspect of the energy of the electromagnetic waves. The wave nature becomes relevant when we take a large group of the photons which can be studied classically. But the structure of photon itself has remained an enigma. The classical picture of the wave and the quantum picture of the particle have remained irreconcilable. There is one more reason for this incompatibility. The classical picture deals with only the propagation of the wave front. The idea of a single wave train is not properly defined in the classical approach. This is more so because it is not clear what the spatial spread of a single wave is in the transverse direction although as a matter of convenience it is generally taken as zero. If in the photon has to be treated as a wave train, then it will have to be constituted by such wavelets having zero spatial spread in the transverse direction. But since photon is a circularly polarized system with unit spin, it becomes still more difficult to represent it in terms of such waves with zero spatial amplitude.

With the introduction of the spatial amplitude as explained in the previous section, attributing an internal structure to photon becomes much simpler. We know that the equation given in (9) along with the gauge invariance of Maxwell's equations hides the existence of the spatial amplitude of the electromagnetic wave. The existence of the spatial amplitude allows us to treat photon as a wave train formed by helical waves. If we assume that the basic form of the electromagnetic wave is helical, then, it is clear that the electric vector will ride on the helical spatial wave, making it circularly polarized. We shall now show that the helical structure of the

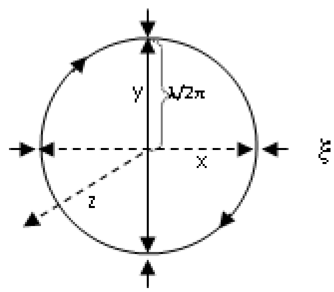

Fig.2. This is the projection of a circularly polarized wave on a transverse plane. The vertical and the horizontal lines stand for two spatial waves having a phase difference of $\frac{1}{2}\pi$. Here we have also shown arrows pointing inward representing the electric field at these points having a phase difference of π with spatial oscillations.

electromagnetic wave will force η_o to possess only a certain value. We know that the helical nature of the electromagnetic wave may be represented by

$$\eta_x = \eta_o \sin[\omega t - kz] \quad and \quad \eta_y = \eta_o \cos[\omega t - kz]\,. \tag{13}$$

Here we assume that the wave is progressing along the z-axis. If we take the projection of such a wave on to the transverse x-y plane, then it is obvious that we will obtain a circle (figure 2). We have to now estimate the velocity of the circular motion. The simplest case of the circular motion can be obtained if we assume that it is executed at the velocity of light. Any other velocity will involve introduction of a new attribute to the electromagnetic wave. Since the wave is also progressing with velocity c along z-axis, it is obvious that by the time the wave travels one wave length, it would have executed one full circle in the transverse directions. Therefore, we may conclude that the radius, R of the circle will be given by

$$R = \lambda/2\pi = 1/k\ . \tag{14}$$

Here we have a clearer idea of the term helicity used in the case of the electromagnetic wave because now we have the case of the spatial amplitude of the wave spinning around an axis. This allows us to understand the concept of spin of the electromagnetic wave classically on the basis of the helical structure of the wave. Since a point on the electromagnetic wave executes rotational motion in the transverse direction, we may take the magnitude of its angular momentum for this motion to be

$$S \;=\; |\boldsymbol{r} \times \boldsymbol{p}| \;=\; (\lambda/2\pi)(h/\lambda) \;=\; \hbar \tag{15}$$

This picture of the wave allows only one value for the spatial amplitude which is $\lambda/2\pi$. This in turn means that χ_o can take only the value $\lambda/(2\pi C)$. Note that only χ_o is taking a constant value. The phase of χ can take the entire range of values from 0 to 2π. On a deeper study of equations (9) to (12) we find that the spatial wave is the most fundamental wave. We may assume that the potentials **A** and φ given by (10) are waves riding on this fundamental spatial wave with a phase difference of $\frac{1}{2}\pi$. The $\boldsymbol{\xi}$-wave given by (2) also can be taken as riding on the spatial wave with a phase difference of 0 or π.

On the basis of the above discussion one may be tempted to presume that the gauge freedom allowed to the amplitude of the spatial wave is not actually put to use as it takes only one value which is $\lambda/2\pi$. On detailed scrutiny we observe that this is not true. We know that if we consider two plane polarized waves of same frequency travelling in the same direction in phase, the resultant amplitude will be twice that of the individual ones. In this manner, it is possible to construct plane polarized waves having any amplitude and all these combinations will satisfy Maxwell's equations due to the property of the gauge freedom discussed above.

In the above approach we have considered Maxwell's equation in vacuum. In fact, it can be easily shown that the above approach will hold good even when we take the case with electric charges.

4. Discussion

We saw that the Maxwell's equations in vacuum accept plane polarized electromagnetic waves as the solution. In this context it is interesting to note that waves moving forward in time and those moving backward in time, both are solutions of the Maxwell's equations. Therefore, even when we take the circularly polarized photon as the fundamental state of the electromagnetic wave, its corresponding state in reverse time also will be a solution of the Maxwell's equation. But we know that on reversing the time coordinate the helicity as also the direction of the electric (and the magnetic) field will get reversed. In fact, we have to take the linear combination of these helical waves of opposite helicity as the solution of the Maxwell's equation. Here it is pertinent to recall that photon is its own antiparticle which means that a photon travelling backward in time will be indistinguishable from one travelling forward in time. Therefore, the solutions of Maxwell's equations will have to be represented by a linear combination of the waves moving forward and backward in time. This is the reason why in Feynman diagrams photon is treated as an entity having no direction in time. Here in our case, the linear combination of two circularly polarized helical waves with opposite helicity will give us a plane polarized wave both in electromagnetic and spatial oscillations. Obviously such a wave will have zero spin angular momentum. It becomes clear now why cosine and sine functions provide a good representation of the electromagnetic wave. The plane polarized spatial wave will not be observable due to the gauge invariance discussed in section 2.

The introduction of the spatial amplitude to the electromagnetic wave implies the existence of a new field which causes these spatial oscillations. It will be shown in a separate paper that a particle like electron can be attributed the structure of confined electromagnetic wave. The generation of the rest (mass) energy of the particle is seen to be the direct result of the localization of the energy of the electromagnetic wave having spatial oscillations. It is observed that the fine structure constant, α is the ratio of the energy of the electromagnetic field of the electron to its rest mass. Since the fine structure constant represents the strength of the electromagnetic interactions, we are prompted to conclude that the rest mass of electron obtains its main contribution from some other field. This leads to the suggestion that the energy of the spatial oscillations belongs to new field. We should keep in mind that the existence of the Higgs field has been mooted for quite some time as the field that creates mass. We shall not identify the spatial oscillations to the Higgs field right now [4]. We shall examine the possibility of identifying the new field with the Higgs field later. We shall show in the next paper that mass could be created by the confinement of the electromagnetic wave and a major part of the energy involved in its creation could be attributed to the energy of the spatial oscillations. If this proposition is to hold good, the electromagnetic wave has to be treated as a composite wave having oscillations in the new field as well as in the electromagnetic field.

5. Conclusion

The idea that the electromagnetic wave possesses spatial oscillations may appear fanciful as the properties of the electromagnetic wave have been extensively studied and understood. However, we observe that this spatial amplitude has remained hidden under the disguise of the gauge invariance and therefore it does not go against the experimentally observed properties of the electromagnetic wave. In a series of papers we shall show that we could treat electron as a confined electromagnetic wave and the spatial amplitude of the electromagnetic wave plays a crucial role in explaining its spin and electric charge. Therefore, it is a comforting thought that the introduction of the spatial amplitude will in no way affect the results validated by the conventional approach. The idea that the major part of the energy of the electromagnetic wave is constituted by the oscillations in the new field, if found acceptable, may result in a completely new way of looking at the basic structure of particles.

References:

[1] E.D.Jordan, K.G.Balmain, Electromagnetic Waves and Radiating Systems, (second Edition) Prentice-Hall of India, New Delhi (2008), p.113-26.

[2] Richard Feynman, The Feynman Lectures on Physics (The Definitive Edition), Vol.2, Dorling Kindersley (India) Pvt. Ltd.,(2009). Ch.18-11.

[3] J.J.Sakurai, Advanced Quantum Mechanics, Addison-Wesley, New Delhi (Second Indian reprint)(1999), p.16-19.

[4] K.Moriyasu, An Elementary Primer For Gauge Theory, World Scientific Publishing Co. (P) Ltd, Singapore, (1983), p.96-97..

Annexure

Maxwell's equations in vacuum is given by

$$
\begin{aligned}
&(i) \quad \nabla \cdot \xi = 0, \qquad &&(ii) \quad \nabla \times \xi = -\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{t}}. \\
&(iii) \quad \nabla \cdot \boldsymbol{B} = 0, \qquad &&(iv) \quad c^2 \nabla \times \boldsymbol{B} = \frac{\partial \xi}{\partial t}
\end{aligned}
\qquad (\text{I})
$$

We shall now solve these equations using Feynman's insightful approach to understand how the concept of the electromagnetic wave emerges from them **[1]**.

We know that the first equation in (I) could be expanded to obtain

$$\nabla \cdot \xi = \frac{\partial \xi_x}{\partial x} + \frac{\partial \xi_y}{\partial y} + \frac{\partial \xi_z}{\partial z} = 0 \qquad (\text{II})$$

Here we assume that there are no variations in the field variables with respect x and y, so that the first two terms could be taken as zero. Hence, we have

$$\frac{\partial \xi_z}{\partial z} = 0 \qquad (\text{III})$$

This means that ξ_z is a constant in the z-direction. If we study Maxwell's equation (1.iv), assuming that just as in the case of the electric field, the magnetic field also has no variation in x and y directions, then it can be seen that E_z is also a constant in time. Such a field could be conveniently taken as zero as we are interested in only dynamic fields. Therefore we may take $E_z = 0$. In other words, the electric field exists only in the x and y directions. Now as a first step, for the sake of simplicity, we may assume that the electric field has a component only in the x-direction and obtain a solution on that basis. Later we may take up the case where the electric field has a component only in the y-direction and get the corresponding solutions. Then, the general solution could always be expressed as the superposition of the two cases.

Let us take the Maxwell's equation (1.ii) and express the components along the three coordinate axes as

$$(\nabla \times \xi)_x = \frac{\partial \xi_z}{\partial y} - \frac{\partial \xi_y}{\partial z}, \quad (\nabla \times \xi)_y = \frac{\partial \xi_x}{\partial z} - \frac{\partial \xi_z}{\partial x}, \quad (\nabla \times \xi)_z = \frac{\partial \xi_y}{\partial x} - \frac{\partial \xi_x}{\partial y} \qquad (\text{IV})$$

Here $(\nabla \text{x} \boldsymbol{\xi})_z$ will be zero because the derivatives with regard to x and y are zero. Note that from (II) we have already taken ξ_x as a constant while ξ_y is taken as zero. $(\nabla \text{x} \boldsymbol{\xi})_x$ is zero because the first term which is a derivative of ξ_z is zero while the second term is zero for reasons already stated. The only component which is not zero is $(\nabla \text{x} \boldsymbol{\xi})_y$ which is equal to $\partial\xi_x/\partial z$. Setting the three components of **(∇xξ)** equal to the corresponding components of $-\partial \mathbf{B}/\partial t$, we obtain

$$\frac{\partial B_z}{\partial t} = \mathbf{0}; \quad \frac{\partial B_x}{\partial t} = 0; \quad \frac{\partial B_y}{\partial t} = -\frac{\partial \xi_x}{\partial z}. \qquad (\text{V})$$

Since the z and x components of the magnetic field have zero time derivatives, they represent constant fields. Such a field could be conveniently taken as zero as we are interested in only dynamic fields. Therefore, we may take $B_z = B_x = 0$. The last equation in (V) shows that the electric field has only the x-component while the magnetic field has only the y-component. This means $\boldsymbol{\xi}$ and $\mathbf{B}$ are perpendicular to each other.

Let us now take the last Maxwell's equation whose components along x, y and z directions could be written as

$$\boldsymbol{c^2}\left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}\right) = \frac{\partial \xi_x}{\partial t} \quad ; \quad \boldsymbol{c^2}\left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}\right) = \frac{\partial \xi_z}{\partial t}; \quad \boldsymbol{c^2}\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) = \frac{\partial \xi_y}{\partial t} \qquad \text{(VI)}$$

On the left hand side of these equations, except for $\partial B_y/\partial z$ all other terms are zero. Therefore we have

$$-c^2 \frac{\partial B_y}{\partial z} = \frac{\partial \xi_x}{\partial t} \ . \qquad \text{(VII)}$$

Now taking partial differentiation with regard to t and using the last equation in (V), we obtain the wave equations

$$\frac{\partial^2 \xi_x}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \xi_x}{\partial t^2} = 0\ ; \quad \frac{\partial^2 B_y}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 B_y}{\partial t^2} = 0\ . \qquad \text{(VIII)}$$

Note that the above equations represent waves having polarization in one plane. Similarly, we can obtain the equations for waves having polarization in a perpendicular plane involving only ξ_y and B_x as

$$\frac{\partial^2 \xi_y}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \xi_y}{\partial t^2} = 0\ ; \quad \frac{\partial^2 B_x}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 B_x}{\partial t^2} = 0\ . \qquad \text{(IX)}$$

The solutions for these wave equations can be written as

$$\xi_x = \xi_{xo}\, sin(\omega t - kz)\ ; \ \xi_y = \xi_{yo}\, sin(\omega t - kz) \qquad \text{(X)}$$

$$B_y = B_{yo}\, sin(\omega t - kz)\ ; \ B_x = -B_{xo}\, sin(\omega t - kz) \qquad \text{(XI)}$$

where ω is the angular frequency and k is the wave vector. Actually we could have as well taken cosine function or even a complex function of the type "ξ_o exp[-i(ωt-kz)]". It is a matter of convenience. However, the fact that the sine function could be expressed as a linear combination of two waves, one travelling forward in time and the other travelling reverse in time is an added advantage as the wave equation given in (IX) possesses functions representing both waves as its solutions. Combining both, the wave equation in a general direction will be given by

$$\boldsymbol{\xi} = \boldsymbol{\xi_o}\, sin(\omega t - \boldsymbol{k.r})\ . \qquad \text{(XII)}$$

Similarly, we may obtain the wave equation for the magnetic component also which may be written as

$$\boldsymbol{B} = \boldsymbol{B_o}\, sin(\omega t - \boldsymbol{k.r})\ . \qquad \text{(XIII)}$$

where $\boldsymbol{B_o}$ will always be perpendicular to $\boldsymbol{\xi_o}$