

On Maxwell's discovery of electromagnetic waves and the gauge condition

P. T. Leung

**Department of Physics, Portland State University, P. O. Box 751,
Portland, OR 97207-0751**

Abstract

The original “theoretical discovery” of electromagnetic waves by Maxwell is analyzed and presented in modern notations. In light of Maxwell’s well-known prophetic dedication to the concept of the vector potential, it is interesting to reveal his derivation of the wave equation for this potential *without* the application of any gauge condition. This is to contrast with typical approaches students learn from standard textbooks for the derivation of the wave equation in various forms. It is in our opinion that intuition and insight, rather than logical deduction, must have played a more significant role in Maxwell’s original discovery, as is not uncommon with discoveries made by the pioneers in science.

Introduction

Physics is an experimental science. This is undisputable in the sense that, just like with any discipline in natural science, a discovery must be ultimately substantiated by experimental verification. Moreover, in the history of Physics, one can find examples of extremely important discoveries which were first made completely from theoretical studies, and then lead subsequently not only to beautiful experimental confirmations, but also to a large variety of significant applications going far beyond the imagination of the theorist(s) who first contributed to that particular discovery. These include, for example, the discovery of antimatter, mass-energy equivalence, curved space-time, ... leading to the applications like Positron Emission Tomography (PET), nuclear energy, Global Positioning System (GPS), ... etc. Nevertheless, few people would have disagreed, among these “theoretical discoveries”, that of the electromagnetic waves propagating at the speed of light made by Maxwell [1, 2] would rank one of the most dramatic, fundamental, and significant achievement for all time. Without the knowledge of electromagnetic (EM) waves, many of our daily-life applications ranging from microwave oven to wireless internet would not have been possible.

It is thus a meaningful task to review the original discovery of Maxwell and learn how such a brilliant theoretical discovery was made during the latter part of the nineteenth century. While it is well-known that the existence of EM waves must be a gauge-independent phenomenon and therefore is best substantiated via the wave equations for the fields, Maxwell had actually derived these equations for both the fields and the potentials (mainly the magnetic field and the vector potential) already in his very original work in 1864 [1]. We shall hence contrast Maxwell’s original work with those found in most modern EM textbooks [3] which are based exclusively on the Lorenz gauge condition [4] on the electromagnetic potentials. In fact, the original paper of Lorenz [4] came two years after that by Maxwell [1], and Maxwell could

not have referred to it in his original discovery of EM waves, although it had then become available when his *Treatise* [2] was published.

For completeness, we shall begin with a brief resume of the textbook derivation of the various EM wave equations and then move on to explain how Maxwell obtained the wave equation for the vector potential without the Lorenz condition.

Textbook derivation of EM wave equations

Let us first display the full set of electromagnetic equations in vacuum in modern notations and in Gaussian unit system:

$$\vec{\nabla} \cdot \vec{E} = 4\pi\rho \quad (1)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad (2)$$

$$\vec{\nabla} \times \vec{E} + \frac{1}{c} \frac{\partial \vec{B}}{\partial t} = 0 \quad (3)$$

$$\vec{\nabla} \times \vec{B} - \frac{1}{c} \frac{\partial \vec{E}}{\partial t} = \frac{4\pi}{c} \vec{J} \quad (4)$$

The sourceless pair then enables us to introduce the scalar and vector potentials with

$$\vec{B} = \vec{\nabla} \times \vec{A} \quad (5)$$

$$\vec{E} = -\vec{\nabla}\phi - \frac{1}{c} \frac{\partial \vec{A}}{\partial t} \quad (6)$$

The various wave equations can then be derived in the following two cases:

Case (i): regions with no source

For $\rho = 0$ and $\vec{J} = 0$, Eqs. (1)-(4) lead straightforwardly to the following d'Alembertian

(wave) equations for the fields $\left(\text{with } \square \equiv \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right)$:

$$\square \vec{E} = 0 \quad (7)$$

$$\square \vec{B} = 0 \quad (8)$$

But as for the potentials, the equations obtained from Eqs. (1), (4), (5), (6) take more complicated forms as follows:

$$\nabla^2 \phi + \frac{1}{c} \frac{\partial}{\partial t} (\vec{\nabla} \cdot \vec{A}) = 0 \quad (9)$$

$$\square \vec{A} = \vec{\nabla} \left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} \right) \quad (10)$$

Now if we impose the following Lorenz gauge condition:

$$\vec{\nabla} \cdot \vec{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} = 0 \quad (11)$$

Eqs. (9) and (10) immediately lead to the following wave equations for the potentials:

$$\square \phi = 0 \quad (12)$$

$$\square \vec{A} = 0 \quad (13)$$

Eqs. (7), (8), (12), (13) constitute the complete set of sourceless EM wave equations in vacuum.

Case (ii): regions with source

In the case with $\rho \neq 0$ and $\vec{J} \neq 0$, Eqs. (1) – (4) will imply modifications to the wave equations and (5) and (6) now take a more complicated form with the introduction of appropriate effective source currents as follows [4]:

$$\square \vec{E} = 4\pi \vec{J}_E ; \quad \vec{J}_E = \vec{\nabla} \rho + \frac{1}{c^2} \frac{\partial \vec{J}}{\partial t} \quad (14)$$

$$\square \vec{B} = 4\pi \vec{J}_B ; \quad \vec{J}_B = -\frac{1}{c} \vec{\nabla} \times \vec{J} \quad (15)$$

Correspondingly, Eqs. (9) and (10) now take the following modified forms:

$$\nabla^2 \phi + \frac{1}{c} \frac{\partial}{\partial t} (\vec{\nabla} \cdot \vec{A}) = -4\pi\rho \quad (16)$$

$$\square \vec{A} = \vec{\nabla} \left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} \right) - \frac{4\pi}{c} \vec{J} \quad (17)$$

Again, under the Lorenz gauge condition Eq. (11), Eqs. (16) and (17) lead to the well-known wave equations for the potentials with source:

$$\square \phi = -4\pi\rho \quad (18)$$

$$\square \vec{A} = -\frac{4\pi}{c} \vec{J} \quad (19)$$

A few comments are in order. First of all, the existence of EM waves, of course, must be a gauge-independent phenomenon. Thus, the wave equations for the fields, Eqs. (7), (8), (14), (15) are of course valid for any gauge. However, those for the potentials, i.e. Eqs. (12), (13), (18), (19), are valid only under the Lorenz gauge condition Eq. (11). This is not a problem within classical electrodynamics if one regards only the fields to be physically realistic and uniquely measurable for a given source. Nevertheless, as is well-known, Maxwell had all the time regarded the vector potential as fundamental by describing it as the electromagnetic momentum in his formulation of EM theory [1, 2, 5]; and had derived and given significance to *both* the wave equations for the magnetic field and the vector potential in his pioneering work [1]. Hence the following intriguing question arises: how could Maxwell have derived Eq. (13) *without* the gauge condition of Lorenz, whose work only appeared two years after that of Maxwell? In fact, while Maxwell was for sure fully aware of the gauge freedom of his theory, he would not have formulated a wave equation for the vector potential by giving up the gauge freedom of such an equation even if the Lorenz gauge condition was available to him. We shall provide a brief

digress of the original reasoning of Maxwell in his derivation of Eq. (13) in modern terminology, and then comment on it.

Maxwell's discovery of EM waves

First we note that Maxwell had only considered the sourceless wave equations with $\vec{J} = 0$. In his original paper on the dynamic theory of EM [1], he had essentially taken the curl of the result in the above Eq. (10) and obtained first the wave equation for the magnetic field in Eq. (8). This is straightforward since the two operators, \square and $\vec{\nabla} \times$, commute with each other. Note that Maxwell's original derivation of Eq. (8) can be made more apparent doing this way, though it has been regarded as rather subtle in some previous works in the literature [6]. Following this, he then went on to derive a similar equation for the vector potential, and we shall follow his *Treatise* (Volume 2, Chapter 20, Article 783) [2] to present this derivation. Essentially, Maxwell had obtained in his Eq. (6) in Article 783 the equivalent of our above result in Eq. (17). Rather than imposing the Lorenz condition in Eq. (11) whose justification is based on the relative arbitrariness of the potentials (as long as they give the same fields), Maxwell had regarded the potentials real and physical and would not have "imposed arbitrary conditions" to be satisfied by them. Instead, he carried out the following steps starting from Eq. (17) to arrive at the wave equation:

Step (i) Taking the divergence of (17) and arrive at:

$$-\frac{1}{c} \frac{\partial^2}{\partial t^2} (\vec{\nabla} \cdot \vec{A}) = \frac{1}{c} \frac{\partial}{\partial t} (\nabla^2 \phi) - \frac{4\pi}{c} \vec{\nabla} \cdot \vec{J} \quad (20)$$

Note that Maxwell's definition of the Laplacian operator for this derivation in [2] is in opposite sign compared to the one often used nowadays [see Article 616, Eq. (3)]. Note also that Eq.(20) is equivalent to Eq. (8) of Article 783 in Maxwell's *Treatise* [2].

Step (ii) Assume no free current, $\vec{J} = 0$:

Hence $\nabla^2 \phi \sim \rho$, the free charge density, will be constant in time if there is no flow of charges. Eq. (20) then leads to:

$$\frac{\partial^2}{\partial t^2}(\vec{\nabla} \cdot \vec{A}) = 0 \quad (21)$$

Step (iii) Leaving out *each* of the terms $\vec{\nabla} \cdot \vec{A}$ and $\partial(\nabla^2 \phi)/\partial t$, for periodic disturbances:

Maxwell's argument is that since (21) implies $\vec{\nabla} \cdot \vec{A}$ is at the most linear in time and $\nabla^2 \phi$ is constant in time, one can safely leave each of them out for oscillatory wave motions..

Obviously, steps (ii) and (iii) are not justified since if one applies consistently the result for $\nabla^2 \phi$ from Eq. (16) into Eq. (20), one simply recovers the equation of continuity which implies a constant ρ for $\vec{J} = 0$. Nevertheless, Maxwell was at that time highly motivated to replace the Newtonian "action-at-a-distance" via realistic propagation of interaction through an elastic medium (aether), and was "driven" to arrive at a second degree differential equation for such oscillatory motion for both the field and the potential which he regarded as physically real and fundamental. Furthermore, the derivation presented in [1] appeared slightly different from that in [2] (the above three steps), as the condition $\vec{\nabla} \cdot \vec{A} = 0$ was explicitly stated in [1] which sometimes misled people to conclude that Maxwell had derived the wave equations in the Coulomb rather than in the Lorenz gauge [7]. It is clear that Maxwell had not applied any gauge condition in [2] as explained above, nor in [1] as is evident from reading Section (99) of the original paper in which Maxwell had treated the left-over terms from the "Coulomb-gauged" vector potential. In fact, the approaches in both [1] and [2] are equivalent to each other. All Maxwell was trying to do is to mathematically demonstrate how the wave equation

for the vector potential can be obtained from his electrodynamic equations *without* ever imposing extra constraints for the potential to satisfy.

Conclusion

We have thus clarified that no gauge condition was applied in the original derivation of the wave equation for the vector potential by Maxwell, in contrary to the modern approach which can be found in most textbooks. Although his mathematical steps were not all justified, we owe Maxwell his great intuition which enabled him to draw so many extremely significant conclusions from his wave equations. These include, for example, the plane wave solution with unique transverse polarizations for the fields; the unification of optics and electromagnetism; ...etc. Teachers should be inspired to pass on knowledge to the younger generations not just based on well-formulated modern textbook derivations, but to lead them to appreciate that in many first-hand scientific discovery, intuition and insight are often more significant than pure logical (mathematical) derivations in such a process.

References

- [1] Maxwell J C 1865 A dynamical theory of the electromagnetic field *Phil. Trans.* **155** 459-512
- [2] Maxwell J C 1873 *A Treatise on Electricity and Magnetism, Volume II* (Dover edition, 1954)
- [3] See, e.g., Jackson J D 1998 *Classical Electrodynamics, 3rd edn.* (New York: Wiley)
- [4] Lorenz L V 1867 On the identity of the vibrations of light with electrical currents *Phil. Mag.* **34** 287-301. Note that confusion between this Danish physicist and the Nobel winning Dutch physicist H. A. Lorentz has appeared from time to time in the literature, see, e.g., Nevels R and Shin C S 2001 Lorenz, Lorentz, and the Gauge *IEEE Antennas and Propagation Mag.* **43** 70-72
- [5] Bork A M 1967 Maxwell and the vector potential *Isis* **58** 210-222
- [6] See, e.g., Sengupta D L and Sarkar T K 2003 Maxwell, Hertz, the Maxwellians, and the early history of electromagnetic waves *IEEE Antennas and Propagation Magazine* **45** 13-19
- [7] See, e.g., Wikipedia: Lorenz gauge condition; and the paper by Nevels and Shin in Ref. [4]

Acknowledgments

I would like to dedicate this work to my recently-deceased father, Mr. Leung Chin-Pang, without whose support in my early years, this work would have never been accomplished.