IOP Publishing European Journal of Physics

Eur. J. Phys. **39** (2018) 063001 (30pp) https://doi.org/10.1088/1361-6404/aadf9b

Review

Evidence for Maxwell's equations, fields, force laws and alternative theories of classical electrodynamics

Max Tran

Department of Mathematics, Kingsborough Community College, Brooklyn, NY 11235-2398, United States of America

E-mail: mtran@kbcc.cuny.edu

Received 20 June 2018, revised 29 August 2018
Accepted for publication 7 September 2018
Published 15 October 2018

Abstract

The set of equations known today as Maxwell's equations along with a few constitutive equations lie at the heart of classical electromagnetism. A common misconception held by many is that Maxwell's equations are essential, and that classical electromagnetic theory is settled science and is no longer an active field of investigations. We will review the four Maxwell's equations and related equations, their supporting experimental evidence, the field concept, and the Lorentz and Ritz force laws. We will give a brief outline of two approaches to classical electromagnetism which bypass Maxwell's equations, the propagated potential approach and the direct action approach which bypasses even the field concept. We will also give a few indications of some current research being done in comparing the fitness of the Lorentz and Ritz force laws, and point out some area of research yet to be explored concerning the direct action approach to classical electromagnetism using the Ritz force formula.

Keywords: Maxwell's equations, Ritz force law, Lorentz force law, electrodynamics, classical electromagnetism

1. Introduction

In his 1861 article 'On Physical Lines of Force', James C Maxwell constructed a mental representation, a type of mechanical model, to explain what was known of electricity and magnetism at that time before the electron's discovery [1]. He based his model on a stationary elastic medium called the ether and sought to explain electromagnetic (EM) phenomena in terms of stresses, strains, and disturbances of this ether. In his model, the ether was a continuous dielectric substance pervading everywhere, with electrical charges being its discontinuities and all currents being neutral and closed. Each charge interacts with this ether only, and only indirectly with each other through the ether, in essence, a local contact action theory as opposed to an instantaneous action at a distance theory. In his later works, the model faded into the background leaving the equations developed from it front and centre. These equations were either expressed in component form with no clear indication of which component is associated with a particular vector, or in the form of quaternions, making them difficult to comprehend. It was left to Oliver Heaviside to recast these equations into vector forms that are widely used today.

As new EM phenomena were discovered, modifications to Maxwell's equations or new interpretations were used in attempts to explain these discoveries. We must be careful when reading Maxwell's works since his conceptions evolved with no clear indications of change. As Pierre Duhem the French physicist, mathematician, historian and philosopher of science once said of Maxwell's theory [2, p 184]:

> Maxwell's electrodynamics proceeds in the same unusual way already analysed in studying his electrostatics. Under the influence of hypotheses which remain vague and undefined in his mind, Maxwell sketches a theory which he never completes, he does not even bother to remove contradictions from it; then he starts changing his theory, he imposes on it essential modifications which he does not notify to his reader … And yet this strange and disconcerting method led Maxwell to the electromagnetic theory of light!

Duhem also expressed what the writer thinks a physical theory should be [3]:

> A physical theory is not an explanation; it is a system of mathematical propositions deduced from a small number of principles whose object is to represent a group of experimental laws as simply, completely and exactly as possible. The sole test of a physical theory, which allows us to pronounce it good or bad, is the comparison between the consequences of the theory and the experimental laws.

With the above in mind, we take the view that Maxwell's theory is the set of equations named in his honour or their reinterpretation due to later researchers. We will examine the original equations, their reinterpretations, and any supporting evidence. Our purpose is to delineate how various equations and concepts in classical electrodynamics are connected to each other and to supporting experiments at a level accessible to undergraduates with some knowledge of calculus. Some articles or books along similar lines but focussing on different aspects can be found in the [4–8]. For a more in-depth critical examination of different theories and equations of electrodynamics, we recommend the freely available book of O'Rahilly [9].

Since Maxwell's equations were developed from a macroscopic continuous medium model, we will examine them from that perspective and will not go into the microscopic theory of subatomic elementary charged particles, except for some brief statements relating the two. Ultimately any macroscopic electromagnetic framework must be built upon elementary charged particles that are observed to interact electrically. Wilhelm Weber, a contemporary of Maxwell was the first investigator to formulate a microscopic electrodynamic theory explaining electromagnetic phenomena from the forces between electric charges in relative motion. In fact, he did this before Maxwell developed his equations, using a force law that depends not only on the relative displacement between the charges but also on their relative radial velocity and acceleration [8]. Despite its many successes in explaining various experimental data, Weber's theory was broadly abandoned in favour of Maxwell's. Hendrik A Lorentz was the first investigator to base a macroscopic classical electromagnetic theory upon a microscopic one using Maxwell's equations and the force law named in his honour [8, 10]. For more recent derivations look into the works of de Groot and Suttorp and the book of Robinson [11, 12]. Jackson and O'Rahilly also have chapters showing how to build a macroscopic EM theory from a microscopic one. These approaches often employ microscopic forms of Maxwell's equations along with the Lorentz force law, or alternatives force laws like the Weber–Ritz, together with some assemble averaging or smoothing process. Looking at building a macroscopic theory from a microscopic one is beyond the scope of this review, so let us return to the macroscopic Maxwell's equations in their simplest forms, in a vacuum, air or linear isotropic dielectric and magnetic materials using SI units.

In the next section, we will look at the equation modelling charge conservation and its experimental evidence. In the third section, we will review Coulomb's law, its equivalences, the corresponding Maxwell's equation and experiments which established its domain of validity. The section will also briefly touch upon the electric field concept. In the fourth section, we will review Gauss' magnetic law, the second Maxwell's equation, and its experimental evidence. This section will also touch upon the magnetic field concept and its relation to electrical currents and the vector potential. In section five, we will look at the Maxwell–Ampere law, its supporting experimental evidence, and related equations. In section six, we will look at the Faraday induction law and the fourth Maxwell's equation, the Maxwell–Faraday equation, and their experimental backings. In section seven, we will examine the concept of fields in more detail and highlight some of its problems. In section eight, we will look at the propagated potentials approach to classical electromagnetism which does not use the four Maxwell's equations. In the ninth section, we circle back to force laws and their relation to fields. We will examine the Lorentz force formula, some of its difficulties and its supporting evidence. We will also briefly discuss an alternative force law, the Ritz force law, and examine experiments comparing the fitness of the two in explaining electrical phenomena involving relatively low charge velocity. In section ten, we briefly examine the direct action approach to classical electrodynamic theory, which bypass fields altogether, using these two force laws. In the final section, we conclude with some indications of further research pathways.

2. Charge conservation and the continuity equation

We begin our brief survey of electrodynamics with the central concept of electric charge. Charge is one of the fundamental aspects of the physical universe similar to space and time. Similar in the sense that we do not really understand their nature, but we can measure their differences. Charge may be defined as the property of matter which gives rise to electromagnetic interactions. One such interaction is the well-known attraction between unlike charges and repulsion between like charges. We know that there is two kind of charges, called positive and negative. We also know that charges come in an integer multiple of a primary unit charge, and is highly likely to be a conserved quantity. Charge conservation was first proposed by British scientist William Watson in 1746 and American statesman and scientist Benjamin Franklin in 1747 without providing definitive experimental evidence in support. In 1843, Michael Faraday, the leading experimentalist of the nineteen century, conducted experiments to support this proposed law.

In modern time, the best experimental tests of electric charge conservation are searches for subatomic particle decays that would be possible if electric charge is not conserved. One such search is for the spontaneous decay of an electron into neutral particles like photons and neutrinos. A recent paper on electron decay by Klapdor-Kleingrohaus *et al* estimate the mean lifetime of an electron to be greater than 10^{26} years, greater than the estimated age of the Universe [13]. To date, no such decays have been observed, according to the particle data group—an international collaboration of particle physicists that compiles and reanalyses published results related to the properties of subatomic particles and fundamental interactions [14].

Although charge conservation requires the total quantity of charge in the Universe to be constant, charge carriers can be created and annihilated as long as the net charge of the new carriers or the annihilated carriers add to zero. An open question is the quantity of net charge in the Universe. Evidence point to it being zero, implying an equal number of positive and negative charges [15]. As evidence for ordinary matter, the difference in magnitude of charge on positive and negative particles are extremely tiny, differing by no more than a factor of 10^{-21} in the case of protons and electrons [16]. Ordinary matter on average contains equal numbers of positive protons and negative electrons in vast quantities. If there is a slight difference in charge between electrons and protons, all earthly matter would have a significant net electric charge and would be mutually repulsive.

In the science of electromagnetism, charge conservation is modelled by the charge continuity equation:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}, \tag{1}$$

which relates the time rate of change of a charge density, ρ, at a point to the divergence of a current density $\mathbf{J}$, in or out of that point. The divergence of a quantity at a point can be thought of as a measure of how much of that quantity is 'flowing' away from the point. Thus charge either accumulated or dispersed at a point either from a net flow of charge into or away from the point. This equation is for a system of coordinates at rest with respect to an inertial frame in which the current source is stationary. When working with moving reference frames or moving sources, the equation would need to be modified to take into account frame or source motion.

3. Coulomb's law

In the development of electrical science, Charles-Augustin de Coulomb theorised that the force law between two static charges to be inversely proportional to the square of their separation distance. He even conducted experiments to support his theory. To honour his contribution, many called the equation expressing the force between two static charges Coulomb's Law. In today's notation, the force between two stationary charge carriers q and Q separated by a distance r is modelled by

$$\mathbf{F} = \frac{qQ}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}}, \tag{2}$$

where ϵ_0 is a constant called the vacuum permittivity or permittivity of free space. This equation gives the force but does not describe how it is transmitted between the charges. Maxwell and many of his contemporaries theorised that the force was transmitted by a medium called the ether. In the current era, the role of the ether in Maxwell's theory is replaced by fields, with charges emitting electric and magnetic fields. These fields propagate at a finite speed and interact with other charges when they are reached. The electric field generated by Q is defined to be the force divided by the second charge, the test charge, as it goes to zero:

$$\mathbf{E} = \lim_{q\to 0} \frac{\mathbf{F}}{q} = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}}.$$

In practical terms, the test charge should be small enough so that its effects on the primary charge carrier can be neglected. In defining the electric field this way, two charges are required, a charge to 'generate' the field and a vanishingly small test charge to 'detect' or 'measure' the field at each point in space where it is said to exist. In this sense, an electric field can be thought of as force per unit charge. This definition in truth does not give us any meaningful picture of what is an electric field, only an operational way to measure it using force and charges. In situations where the field of many charges is required, we use the linear superposition principle—that the field from each charge can just be added together to get the total electric field. This principle can be used for fields described by equations linear in the quantities generating the fields. In the case of an electric field, the equation is linearly in the charge variable, Q. The superposition principle, in essence, says that charged particles do not cause reactive effects on each other in generating a field, that there is no decrease in a charge's ability to affect distant charges as a result of its actions imparted to intervening charges. This is certainly an idealisation whose domain of validity or limits of applicability is unclear to the author. According to Jackson, linear superposition has been tested at the 0.1% accuracy level on the macroscopic and atomic scale but did not give any reference to performed experiments [17]. Jackson went on to say that departures from linear superposition can be sought for in the subatomic domain. Supposing that we are in the domain of validity of the linear superposition principle, the total field from a distribution of charge is a sum of fields from each charge. For a continuous charges distribution ρ, the sum is generalised to an integral to get the total electric field:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\mathbf{r}_s)\hat{\mathbf{R}}\mathrm{d}V}{R^2}, \tag{3}$$

where $\mathbf{r}$ is the field point or location of a test charge, $\mathbf{r}_s$ the position of a source charge, $\mathbf{R} = \mathbf{r} - \mathbf{r}_s$ the displacement of field and source points, with $R = |\mathbf{R}|$ being the separation distance between them and V the volume containing the charge density.

Using (3) to find an electric field requires knowledge of the charge distribution, which in practice is often difficult to obtain. Often alternative equations are used, especially when the charge distribution is highly symmetrical. One such equation is called Gauss' law, after Carl F Gauss, a mathematician and physicist, who formulated it. This law relates the flux of an electric field through a surface Σ to the enclosed net charge in the volume defined by the surface, V:

$$\oint_\Sigma \mathbf{E}\cdot\mathbf{ds} = \int_V \frac{\rho\mathrm{d}V}{\epsilon_0}. \tag{4}$$

Using the above equation and the divergence theorem, which equates the flux of a vector field through a surface to the divergence of the same field integrated over the volume defined by the surface, we obtain Gauss' law in differential form:

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_o}. \tag{5}$$

This process requires certain smoothness assumption on the $\mathbf{E}$ field, which will not be elaborated here but is related to the scale of the model being sufficiently large in comparison to the size of individual charged particles. One way to interpret this law is that positive charge acts as a source of electric fields while negative charges act as a sink. This equation does not determine **E** uniquely. Helmholtz' vector field decomposition theorem tells us the curl of **E** is also needed [18]. The curl of **E**, represented by the expression $\nabla \times \mathbf{E}$, turns out to be zero when **E** is given by (3). The curl of a vector field being zero implies that the vector field can be defined as a gradient of a scalar function. For the case of the **E** field, the scalar function is called the electrostatic potential or just the scalar potential, denoted by ϕ. For the situation of static charges, their relationship is expressed by the equation:

$$\mathbf{E} = -\nabla\phi, \tag{6}$$

where the negative sign is just an adopted convention. This equation defines the scalar potential in terms of the electric field, but more often it is used to find the field once the potential function is determined. For the electric field defined by Coulomb's law, (3), the scalar potential is given by the equation below:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\mathbf{r_s})\mathrm{d}V}{R}. \tag{7}$$

A common physical interpretation of the scalar potential is in terms of the work required to move the test charge 'q' from far away to the position **r**, closer to the charge density ρ. This equation is important its own right, since it can be used to develop electromagnetic theory in a different way then what is usual done, as we shall see later.

When investigating electric phenomena in the presence of material bodies, it was found useful to define an electric displacement field **D**, to model how material bodies realign their internal structures in response to an impressed electric field. Useful in the sense that the **D** field can be related to free charges by an equation, while the **E** field requires the total net charge which includes free and bounded charges. The constitutive equation which defines the **D** field in terms of the electric field **E** in linear isotropic dielectric materials is given by

$$\mathbf{D} = \epsilon\mathbf{E}, \tag{8}$$

where ϵ is a constant, called the permittivity of the material. The constant ϵ, which has to be measured for each substance, reduces to ϵ_0 when there is no dielectric material presence. If the materials are not linear or isotropic, the constants ϵ may sometime be replaced by higher dimensional expressions to indicate that the two fields are no longer aligned in the same direction. In other cases, it is not possible to relate the two fields together using a simple equation. In actuality, the assumption of linearity and isotropy is satisfied only under certain conditions, within a certain range of temperature, pressure, or strength of the impressing electric field.

The first Maxwell's equation is Gauss law, (5), but we will call it Coulomb's law to differentiate it from 'Gauss law' for magnetism. In [19], Maxwell expresses this law in the form

$$e + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z} = 0,$$

which in more modern notation is written as

$$-\rho_f + \frac{\partial D_1}{\partial x} + \frac{\partial D_2}{\partial y} + \frac{\partial D_3}{\partial z} = 0,$$

and more compactly in vectors notation as

$$-\rho_f + \nabla \cdot \mathbf{D} = 0, \qquad \text{equivalently} \qquad \nabla \cdot \mathbf{D} = \rho_f, \tag{9}$$

where $\mathbf{D}$ is the electric displacement and ρ_f is the free charge density, both at the point (x, y, z, t). The negative sign in the modern notation may be due to the fact that we use divergence instead of convergence, as used by Maxwell.

Let us examines the experimental evidence for (2)–(5), and (9). Since they are mathematically equivalent when there are no dielectric materials present, evidence for one is evidence for all of them. Moreover, many experiments of various degree of accuracy and different scale regimes have been performed throughout the years, going back to Coulomb and Cavendish in 1772. The results of experimental tests of Coulomb's law are expressed in one of two forms [17, p 5]:

(i) Assume that the force varies as $1/r^{2+\epsilon}$ and give a value or an upper bound for ϵ.
(ii) Assume that the electrostatic potential has the 'Yukawa' form, $r^{-1}\mathrm{e}^{-\mu r}$, and give a value for μ or μ^{-1}. Since $\mu = m_\gamma c/\hbar$, the result is sometimes expressed as an upper bound on m_γ the assumed mass of the photon.

Coulomb's experiment used a torsion balance to measure the force between two charged spheres and gave a rough indication that the inverse square law held. Cavendish original experiment in 1773 used two concentric metal shells and gave a bound of $|\epsilon| \leqslant 0.02$ [20]. Around one hundred years later, Maxwell performed a similar experiment and obtained a limit of $|\epsilon| \leqslant 5 \times 10^{-5}$. Plimpton and lawton in 1936, obtained $|\epsilon| \leqslant 2 \times 10^{-9}$ [21]. An experiment by Crandall in 1983 yield $|\epsilon| \leqslant 6 \times 10^{-17}$ and $m_\gamma < 8 \times 10^{-51}$ kg. For a more in-depth discussion of the experiments, see [22] where these figures were obtained. All these were Earth-bound laboratory experiments involving parameters that could be individually varied and tested.

Another approach, called astronomical methods, is based upon measurement of the magnetic fields of planetary bodies like Earth, Jupiter, or the Sun and even more extensive systems. These measurements are used to get smaller upper bounds on the mass of a photon. This approach involves many factors, assumptions and interpretations that are hard to verify, so cannot be said to provide substantial evidence. For instance measurements of the Earth's magnetic field on its surface and at distances from the surface via satellite, along with specific assumptions, give a calculated limit of $m_\gamma < 4 \times 10^{-51}$ kg.

The laboratory experiments and geophysical calculations show that on length scales of order 10^{-2}–10^{7} m, the inverse square law model is extremely accurate. Thus even if photons have a non-zero rest mass, Coulomb's law can be used as an excellent approximation for laboratory scale measurements. Since a non-zero rest mass for photons have consequences not just to the forms of Coulomb's law and the Ampere–Maxwell law but also to many other models in physics, like gravitational deflection and the large-scale structures of the Universe, the experimental science of 'weighing' photons is developing many different approaches, a brief survey of which can be found in [22, 23]. A non-zero photon rest mass can be incorporated into electromagnetic theory by modifying Maxwell's equations. In fact, one such modification, the Proca equations, are used to get estimates on photons rest masses.

A final point before we proceed to the next Maxwell's equation, data and inferences from experiments of alpha or beta particles scattering off of atomic nuclei are often used in support of Coulomb's law down to subatomic scales. Since these experiments involve moving charges, which is outside the scope of Coulomb's law, it is more logically consistent to use these experimental data to support generalisation of Coulomb's law.

4. Gauss' magnetic law

Early investigators of electric phenomena were able to separate two kinds of electrical charges, even though they may not know their true natures. In magnetism, even though two type of poles were identified, they were always observed to occur together. This observation can be modelled in term of an equation, called Gauss' magnetic law:

$$\nabla \cdot \mathbf{B} = 0, \tag{10}$$

which states that magnetic induction fields, **B**, are solendial, or circuital. That is to say, there are no magnetic charges, often called magnetic monopoles, from which a magnetic field emanate. Maxwell first used monopoles in his theorising, but he envisioned it as one pole of a dipole with a large separation distance between the two poles. Some fifty years later, Dirac theorised that the existence of just one monopole would explain the discrete nature of electric charges [17].

The evidence for this law is relatively strong. As yet, according to the particle data group, no monopole has been conclusively found or detected, despite intensive searches [16]. If monopoles are ever found then the above law would still hold in regions free of monopoles; otherwise, it would need to be modified to include magnetic charges should they be present in the region of interest. According to Jackson, if all particles in nature have the same ratio of magnetic to electric charge, the fields and sources could be redefined so that (10) is true. In this sense, it is a matter of convention to say that magnetic monopoles do not exist [17].

When investigating magnetic phenomena in the presence of material bodies, it was found useful to define another field called the magnetic intensity **H**, to model how various material reacted to an impressed magnetic induction field **B**. Useful in the sense that the **H** field can be related to the free current by an equation, while the **B** field requires the total net current which includes free and polarisation currents. Informally, both are called magnetic fields. The constitutive equation which defines the **H** field in terms of the **B** field in linear isotropic magnetic materials is given by

$$\mathbf{H} = \frac{\mathbf{B}}{\mu}, \tag{11}$$

where μ is the magnetic permeability, a constant that has to be measured for each material. If the materials are not linear or isotropic, the constants μ may sometime be replaced by higher dimensional expressions to indicate that the two fields are no longer aligned in the same direction. In other cases, it is not possible to relate the fields together in a simple way, as it is with an important class of ferromagnetic material, like iron, used in making magnets. In actuality, the assumption of linearity and isotropy is satisfied only under certain conditions, within a certain range of temperature, pressure or strength of the magnetic field that the material is under. When no magnetic material is present, μ is taken to be μ_0, the magnetic permeability of the vacuum. In terms of the **H** field, Gauss' magnetic law is $\nabla \cdot \mathbf{H} = 0$.

Gauss' magnetic law along with Helmholtz's decomposition theorem [18] from vector calculus implies that magnetic fields can be found from taking the curl of a vector potential function **A**, in symbols:

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{12}$$

This equation in a sense defines the vector potential, but often it is used the other way around —to obtain **B** from **A**. In fact, electromagnetism can be developed from just the vector potential and the scalar potential functions, as we shall see later.

If there are no magnetic charges, then what gives rise to magnetic fields? Let us examines history for the answer. In the early days of electromagnetic theory, the science of magnetism and electricity were developed independently of each other, with some connection being theorised. Hans Christian Oersted confirmed one connection with his discovery in 1820 that electrical currents would deflect magnetic needles [8]. After Oersted discovery, Andre-Marie Ampere conducted a series of experiments on neutral currents in closed circuits to determine the force law between thin wires. He obtained a formula that was highly praised in his time but is rarely presented in modern textbooks. In 1822, Ampere arrived at the following force law between two current elements, $I\mathbf{ds}$ and $I'\mathbf{ds}'$:

$$\mathrm{d}^2\mathbf{F} = -\frac{\mu_0 II'}{4\pi r^2}[2(\mathbf{ds}\cdot\mathbf{ds}') - 3(\hat{\mathbf{r}}\cdot\mathbf{ds})(\hat{\mathbf{r}}\cdot\mathbf{ds}')]\hat{\mathbf{r}}.$$

This law is not the most general law that can represent the results of his experiments, but was accepted by Ampere because it satisfies Newton's third law in the strong form—equal and opposite forces along the same line. In representing his experimental results, Ampere never formulate them in terms of fields, a concept conceived later by Michael Faraday. In any case, his force formula cannot be expressed in terms of fields in a straightforward manner. Ampere went on to theorise that magnetism in its many forms arose from electric current, charges in some sort of relative motion, a theory that is accepted by many and could explain the various magnetic phenomena on planetary and galactic scales existing in the Universe. If all magnetic interactions arose from electric currents, Gauss' magnetic law would undoubtedly hold. Whether all magnetic phenomena can be explained in terms of electric current is still a topic of investigation as evidenced by the continuing search for monopoles. In any case, a complete theory of ferromagnetism requires the use of quantum mechanics, the structures and inherent properties of electrons and atoms, and their relative motions and alignments; see [24, 25].

In his paper of 1845, Hermann G Grassmann expressed the force on the current element $I\mathbf{ds}$ due to the current element $I'\mathbf{ds}'$ as follows:

$$\mathrm{d}^2\mathbf{F} = I\mathbf{ds} \times \frac{\mu_0}{4\pi r^2}(I'\mathbf{ds}' \times \hat{\mathbf{r}}). \tag{13}$$

Grassmann apparently did not conduct any experiments, and only formulated this law in response to some perceived flaws in Ampere's formula, with the most important one being the improbable change in the force from attraction to repulsion at a critical angle between the directions of the current elements and the straight line connecting them; see Assis' and Chaib's book for a thorough history of Ampere's electrical experiments leading to his formula, and Grassmann's criticism of Ampere's laws which lead to Grassmann's formula [26]. Grassman force law gives the same result of Ampere force when integrated around one current that is steady, closed and neutral, but it does not satisfy Newton's third law of equal and opposite forces for the two current elements. Grassman's force formula is the one often given in modern textbooks, under the name of the Biot–Savart law, since it can be expressed in terms of fields. Around the same time as Ampere, Jean-Baptist Biot and Felix Savart conducted a few experiments on forces and torques between wires and magnets. They theorised that wires became magnetised when an electric current flows through them and obtained a force law that gives the same result as Ampere's force law for closed circuits of neutral currents. Part of their (Grassman's) force formula now served as the basis for defining a magnetic field generated by a current, even though they never mentioned magnetic fields in their research. The magnetic induction field, $\mathbf{B}$, at a point $\mathbf{r}$ generated by a closed, neutral, steady current I is defined by the Biot–Savart law as

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\oint_C \frac{I\mathbf{ds}\times\hat{\mathbf{R}}}{R^2}, \tag{14}$$

where the integral is over the closed current loop C, and R is the distance between $\mathbf{r}$ and a point of the loop. This is just the integral of the second factor in the Grassman force formula (13). For a steady current density $\mathbf{J}$, the Biot–Savart formula becomes

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_V \frac{\mathbf{J}\times\hat{\mathbf{R}}}{R^2}\mathrm{d}V, \tag{15}$$

where the integral is over the current density volume and R is the distance between $\mathbf{r}$ and a point in a volume element. The vector potential in this case is given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_V \frac{\mathbf{J}}{R}\mathrm{d}V. \tag{16}$$

These formulas will later be generalised to time-varying currents and are given here to serve as a basis for comparison.

5. Maxwell–Ampere law

The previous two Maxwell's equations were for situations of stationary charges or steady current densities; conditions often called quasistatic. The next two equations are for more dynamic situations, involving accelerating charges and time-varying current densities. These dynamic situations generate electromagnetic waves of various kinds crucial to our modern communication systems like radios, televisions, mobile phones, and related technologies like wifi, radar and GPS. But before we get to these dynamic situations, we examine Ampere's magnetic circuital law for situations involving steady currents in closed circuits:

$$\nabla\times\mathbf{B} = \mu_0\mathbf{J}. \tag{17}$$

This equation states that the curl of the magnetic induction field, $\nabla\times\mathbf{B}$, at the point (x, y, z, t) is proportional to the current density of free charges, $\mathbf{J}$, at the same point. This law was never written down, much less derived by Ampere. It was first derived by Maxwell in his 1865 paper [8]. Maxwell added another term to this law in situations where $\mathbf{J}$ is not in a closed circuit, such as in a circuit containing a parallel-plate condenser with a sufficiently large plate separation. With the addition of this term, Maxwell completed the closed circuit in his model and obtained the equation below, often called the Maxwell–Ampere law:

$$\nabla\times\mathbf{B} = \mu_0\left(\mathbf{J} + \epsilon_0\frac{\partial\mathbf{E}}{\partial t}\right). \tag{18}$$

This equation states that the curl of the magnetic induction field at the point (x, y, z, t) is related to the total current density $\mathbf{J}$ at the same point plus an additional term $\partial\mathbf{D}/\partial t = \epsilon_0\partial\mathbf{E}/\partial t$, called the displacement current. It is seen by some as Maxwell's major contribution to electromagnetic theory. The interpretation of this term and its significance are subject to much scientific debate; see [7, 27] for some insightful analysis and references on the issue. In his own words in his final work on EM, Maxwell used displacement for the related concept of polarisation [28, p 65]:

> The amount of the displacement is measured by the quantity of electricity which crosses a unit of area while the displacement increases from zero to its actual amount. *This therefore is the measure of the electric polarisation …* The variations of electric displacement evidently constitute electric currents.

Whether he meant polarisation whenever he used 'displacement' is unclear, but it is likely that the modern conception of electric displacement is not the same as Maxwell's conception.

The displacement current term is essential in Maxwell's equations for two main reasons. It is required by Gauss' law and the Maxwell–Ampere law to be consistent with the continuity equation (1). It is also needed by Maxwell's medium electromagnetic theory to arrive at a wave equation for the electric and magnetic fields when its containing equation is combined with the previous two Maxwell's equations and the Maxwell–Faraday equation to follow. These wave equations for the electric and magnetic fields lead to the prediction of electromagnetic waves travelling at the speed of light. Consequently, Maxwell theorised that electromagnetic waves and light are the same thing, an idea that is readily accepted today.

A standard interpretation of the Maxwell–Ampere law is that a time-varying electric field creates a magnetic field, especially when there is no current present, but the equation just expresses a correlation between the two fields and does not express a causality relationship between the two. Since the quantities on both sides of the equation are evaluated at the same point in space and at the exact moment in time, one cannot be the cause of the other if we accept the 'causality principle'—a cause must precede its effect. What some researchers think actually give rise to these two fields are charges and currents that are not in the local picture. We will return to this point in the section on alternative approaches to electromagnetism.

Experimental evidence for the Maxwell–Ampere equation is not as extensive as for the previous two laws. Our literature search turned up one experiment. In 1985, Bartlett and Corle conducted an experiment to measure the magnetic induction field inside a parallel-plate capacitor as it was being charged. Their experiment agreed to within 5% of the theoretical prediction that the **B** field, in accordance with (18), varies linearly with distance when a uniform displacement current flows between the plates. The author also made an interesting observation that the prediction would be the same as using the Biot–Savart formula (14), with only the conduction current. Thus the use of the displacement current is not necessary to explain their result; see their paper [29] for more details. In later attempts, the investigators tried to observe the displacement current directly but without success [30].

6. Faraday induction law

It is hard to imagine what our world would be like today without generators, turbines, transformers, and motors, all of which use electrical induction in their operations. This important phenomenon was discovered by Michael Faraday after many years of electrical experimentation. He was inspired by the various discoveries of Oersted, Ampere, and Biot–Savart, and conducted a series of investigations between 1820 and 1831, leading to his statement of the law of electrical induction. More specifically, he observed that an electric current is induced in a circuit whenever the strength of a nearby current is altered, or when a magnet is brought near to the circuit, or when the circuit itself is moved about in the presence of another current or magnet. The most common statement of Faraday's law in modern language is the following statement:

> The induced electromotive force in any closed circuit is equal to the negative time rate of change of the magnetic flux enclosed by the circuit.

Faraday never expressed the law of induction in equation form. This was first done by Franz E Neumann in 1845 if the historians are to be believed. In equation form, this law is expressed as

$$V = -\frac{\mathrm{d}\Phi_B}{\mathrm{d}t}, \tag{19}$$

where V is the electromotive force, and Φ_B is the magnetic flux. The quantity V is measured in volt and is the potential difference which set an electric current into motion in a conductor. The magnetic flux may be thought of as the 'amount' of magnetic field 'flowing' through the surface whose boundary is the closed circuit. In terms of the induced electric field **E** and the inducing magnetic field **B**, this law is often expressed by the equation

$$\oint_C \mathbf{E} \cdot \mathbf{ds} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B} \cdot \mathbf{dA}, \tag{20}$$

where C is any closed path in space and S is any open surface bounded by C. The path and surface may be moving arbitrarily or be stationary. For a moving circuit, the **E** field is at rest with respect to the circuit and must be expressed in terms of fields at rest with respect to the measuring instrument for correct results. This equation summarised the experimental discoveries of Faraday on thin wires, in closed circuit with neutral currents. Despite its relative importance, this equation is not strongly supported by experimental evidence. Faraday did perform a series of experiments which indicate the phenomenon of induction, but these were rather crude experiments when compared to the sensitivities of modern experimental setups. Laboratories' electrical measurements throughout the world based upon the equation agree with each other to a high degree, but there are no experiments on the level of accuracy as those of Coulomb's and Gauss' laws that we are aware of. This circumstantial evidence indicates that Faraday's equation serves as a good enough model of electrical induction encountered in closed circuits, involving neutral currents, and relatively slow moving charges and slowly varying currents, such as in our electrical generators and motors.

The integral Faraday's equation (20) is often used to derive the fourth Maxwell's equation, usually called the Maxwell–Faraday equation. Many textbooks do not delineate the assumption used to derive the fourth Maxwell's equation, so it may be instructive to derive the equation more carefully. We begin with some simplifying assumptions, used by Maxwell implicitly since his model was based upon a stationary ether and stationary sources:

(i) The circuit or path is non-deforming and is at rest in the frame where the **B** field and measuring instrument are at rest so that the induced **E** field is defined in the same frame.
(ii) The surface is non-deforming so that different parts are not moving relative to each other.

We apply Stoke's theorem to the left-hand side of (20), which requires a certain smoothness assumption on the **E** field, to change the line integral to a surface integral, giving the equation:

$$\oint_C \mathbf{E} \cdot \mathbf{ds} = \int_S \nabla \times \mathbf{E} \cdot \mathbf{dA}. \tag{21}$$

Using a theorem from vector calculus which gives the total time rate of change of an integral over a surface moving with a velocity field **v**, we get the equation

$$-\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B} \cdot \mathbf{dA} = -\int_S \left[\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) + \mathbf{v}\nabla \cdot \mathbf{B}\right] \cdot \mathbf{dA}. \tag{22}$$

Equating the two right-hand sides of equations (21) and (22), and bringing all terms to one side under one integral gives

$$\int_S \left[\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) + \mathbf{v}\nabla \cdot \mathbf{B}\right] \cdot \mathbf{dA} = 0.$$

This is true for any arbitrary surface S bounded by C, thus implying the integrand is zero and yielding the equation

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) + \mathbf{v}\nabla \cdot \mathbf{B} = \mathbf{0}. \tag{23}$$

Under the simplifying assumption (ii) $\mathbf{v} = \mathbf{0}$ the last two terms of the above equation become zero, yielding the Maxwell–Faraday equation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{24}$$

This equation equates the curl of an electric field, $\nabla \times \mathbf{E}$ at a point (x, y, z, t) to the negative time rate of change of an accompanying magnetic field, $\mathbf{B}$, at the same point. A common interpretation of this equation is that a changing magnetic field generates an electric field, but the equation just expresses a correlation between the two. Since the quantities on both sides of the equations are evaluated at the same point in space and at the exact moment in time, one cannot be the cause of the other if we accept the causality principle. This equation also focuses on induction associated with a time-varying magnetic field, and neglect induction associated with the relative motion of a circuit with another circuit or magnet. To model motional induction within Maxwell's theoretical framework requires removing the two simplifying assumptions and dealing with the arbitrary motion of the current loop, along with using equations to express the fields in one frame of reference in terms of another frame of reference. The equations to express fields in one frame in terms of another frame are usually obtained from a relativity theory, with Albert Einstein's special theory of relativity being the one most used for uniform motion in one direction.

Experimental evidence for the Maxwell–Faraday equation is not as solid as for the other Maxwell's equations. Some of Faraday's experiments come to mind, but there are none on the level of accuracy as those to test Coulomb's and Gauss's magnetic law that we are aware of. Since the Maxwell–Ampere and Maxwell–Faraday equations are used to obtain wave equations for the electromagnetic fields, many point to electromagnetic waves like light and radio waves as definitive evidence for Maxwell's equations. Since other theories like the Weber–Ritz' ballistic theory also predicts electromagnetic waves, detections of such waves can be used to support these other theories as well as Maxwell's [9].

Before leaving this section, we will derive an important equation, which is often used to get an electric field from the two potential functions. The argument leading to the Maxwell–Faraday equation can also be done with the vector potential since by definition $\mathbf{B} = \nabla \times \mathbf{A}$. Doing so for a circuit at rest gives a series of equations relating an induced electric field to the vector potential of the associated magnetic field:

$$\nabla \times \mathbf{E} + \nabla \times \frac{\partial \mathbf{A}}{\partial t} = \nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = \mathbf{0}.$$

The curl of a vector field being zero implies that the vector field can be obtained from a gradient of some scalar function, ϕ. By convention, a negative sign multiplies ϕ, and for modelling EM phenomena it is often identified with the scalar potential. In equation form, we have

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla\phi.$$

Isolating the $\mathbf{E}$ field, we arrived at an important equation which defines an electric field $\mathbf{E}$ in terms of the vector and scalar potentials and which can be used to develop an electromagnetic theory without using Maxwell's equations:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{25}$$

7. What exactly is an electric or magnetic field?

We have seen how electric fields and magnetic fields are measured or calculated. To summarise, in terms of force the electric field is just force per unit charge, while the magnetic field is force per unit current. What evidence is there for the existence of fields independent of forces? Since there are those who question the reality of electric and magnetic fields, let us briefly look at their various definitions or conceptions. Ribeiro, Vannucci, and Assis in their paper summarised the various definitions of the electric or magnetic field in the following list [31].

(i) A region of space in which charged particles or electric currents are under the influence of electromagnetic forces (the original meaning of the concept, scarcely used nowadays).
(ii) A vectorial quantity, defined by mathematical functions, $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$.
(iii) A group of lines of force either, electric or magnetic, in a certain region.
(iv) A state of the space around charges, magnets, and currents, that is a condition of space.
(v) Immaterial matter, susceptible to suffer the action of charged particles and electric currents.

Definition (i) of the fields as regions of space was used by the likes of Faraday and Maxwell when the field concept was just being developed by them, and has long been discarded. It is only included here as an illustration that not all concepts or definitions of the past are correct or of any use today. Definition (ii) of the fields as vectorial functions is certainly something most people would accept since they are often used to model electromagnetic phenomena. Definition (iii) is a pictorial model and is equivalent to definition (ii) when expressed mathematically as vector fields. These two definitions give us no sense as to the nature of fields and beg the question of what are these lines of force or what is the physical reality that we are modelling with these vector fields? Definitions (iv) and (v) give a better sense of electric and magnetic fields and share some characteristics of the theorised ether, a concept that is often dismissed today. Yet it has reappeared under a different name, with some envisioning it as a polarisable sea of virtual subatomic charged particles called the quantum vacuum. Robert B Laughlin, Nobel Laureate in Physics, had this to say about the ether in contemporary theoretical physics [32]:

> The word 'ether' has extremely negative connotations in theoretical physics because of its past association with opposition to relativity. This is unfortunate because, stripped of these connotations, it rather nicely captures the way most physicists actually think about the vacuum … Relativity actually says nothing about the existence or nonexistence of matter pervading the Universe, only that any such matter must have relativistic symmetry … It turns out that such matter exists. About the time relativity was becoming accepted, studies of radioactivity began showing that the empty vacuum of space had spectroscopic structure similar to that of ordinary quantum solids and fluids. Subsequent studies with large particle accelerators have now led us to understand

> that space is more like a piece of window glass than ideal Newtonian emptiness. It is filled with 'stuff' that is normally transparent but can be made visible by hitting it sufficiently hard to knock out a part. The modern concept of the vacuum of space, confirmed every day by experiment, is a relativistic ether. But we do not call it this because it is taboo.

The ether still appears in classical electrodynamic theories, mainly as a standard of rest for the velocity appearing in various force formulas.

In modern theories, electric and magnetic fields are viewed as different aspects of one field—an electromagnetic field, with macroscopic fields being a statistical averaging of microscopic fields. In terms of quantum theory, microscopic electromagnetic fields are identified with photons, generally thought to be massless particles with definite energy, momentum, and spin (angular momentum). So macroscopic EM fields may be conceived as functions modelling particle streams of photons. Without going into the realm of quantum theory, what are the evidence for the reality of electric or magnetic fields on a macroscopic scale? The evidence for the reality of fields apart from forces is nonexistence. Percy W Bridgman, an experimental physicist and 1964 Nobel Laureate, expressing his critique of the field concept had this to say [33]:

> The reality of the field is self-consciously inculcated in our elementary teaching, often with considerable difficulty for the student. This view is usually credited to Faraday and is considered the most fundamental concept of all modern electrical theory. Yet in spite of this I believe that a critical examination will show that the ascription of physical reality to the electric field is entirely without justification. I cannot find a single physical phenomenon or a single physical operation by which evidence of the existence of the field may be obtained independent[ly] of the operations which entered into the definition … I do not believe that the additional implication of physical reality has justified itself by bringing to light a single positive result, or can offer more than the pragmatic plea of having stimulated a large number of experiments, all with persistently negative results … The electromagnetic field itself is an invention and is never subject to direct observation. What we observe are material bodies with or without charges—including eventually in this category electrons—their positions, motions and forces to which they are subject.

If macroscopic fields are mere mathematical descriptions, they will not rotate with their sources. So any experimental evidence indicating that the field of a rotating magnet also revolves along with the magnet would support the reality of fields. In fact, there have been many experiments going back to Faraday that were conducted to detect such field rotation. These experiments often fall under the label 'unipolar induction', with some arguing that the field rotates and others that it does not. Two recent experimental papers provide a brief history of this 200-year-old question and indicate that the field does not rotate [34, 35].

Richard Feynman proposed an apparent paradox which requires that fields have angular momentum to resolve [36, section 17.4]. In his words:

> Imagine that we construct a device … There is a thin, circular plastic disc supported on a concentric shaft with excellent bearings, so that it is quite free to rotate. On the disc is a coil of wire in the form of a short solenoid concentric with the axis of rotation. This solenoid carries a steady current I provided by a small battery, also mounted on the disc. Near the edge of the disc and spaced uniformly around its circumference are a number of small metal spheres

> insulated from each other and from the solenoid by the plastic material of the disc. Each of these small conducting spheres is charged with the same electrostatic charge Q. Everything is quite stationary, and the disc is at rest. Suppose now that … the current in the solenoid is interrupted, without, however, any intervention from the outside. So long as the current continued, there was a magnetic flux through the solenoid more or less parallel to the axis of the disc. When the current is interrupted, this flux must go to zero. There will, therefore, be an electric field induced which will circulate around in circles centred at the axis. The charged spheres on the perimeter of the disc will all experience an electric field tangential to the perimeter of the disc. This electric force is in the same sense for all the charges and so will result in a net torque on the disc. From these arguments we would expect that as the current in the solenoid disappears, the disc would begin to rotate. If we knew the moment of inertia of the disc, the current in the solenoid, and the charges on the small spheres, we could compute the resulting angular velocity.
>
> But we could also make a different argument. Using the principle of the conservation of angular momentum, we could say that the angular momentum of the disc with all its equipment is initially zero, and so the angular momentum of the assembly should remain zero. There should be no rotation when the current is stopped. Which argument is correct?

In his lecture on field energy and momentum, Feynman said that when the current is turned off the disc should spin because energy and angular momentum is being transferred to it from the field [36, section 27.6]. But the rotation of the disc in the paradox could also be explained by the direct action force formula of Weber, as it was done by Fukai in [37]. In any case, arguments like these cannot really determined if fields and their associated momenta are real or not, only crucial experiments can decide the issue.

In summary, the evidence supporting the reality of macroscopic electric or magnetic fields as physical entities independent of forces are inconclusive at best. This lack of solid evidence has significant implications for models in theoretical physics which use magnetic field lines breaking and re-connecting, so-called magnetic reconnection. As O'Rahilly has shown in his book, a classical electrodynamic theory capable of explaining induction and radiation can be developed using just force laws without resorting to fields [9].

If macroscopic fields are not physically real what is electromagnetic radiation? What is the cause of radiation pressure and what is responsible for the transfer of energies in electromagnetic waves? When we switch on an incandescent lamp, it lights up almost instantly, which gives us a sense of how fast electromagnetic interactions can be. It is certainly not the moving electrons in the wire which transfer the energy to the device, since their drift velocities are too slow, on the order of 0.01 cm s^{-1} for direct currents of 10 amps in copper wires with cross-sectional area of 3.14×10^{-2} cm^2, while for alternating currents the electrons just oscillate back and forth. So what is the physical mechanism that transfers energy to the lamp? Many would say that an electromagnetic field moving along the wire in waves is the carrier of the energy. But other theorised it could just be the propagation of induction or polarisation along the conductor or insulator which is transferring energy. We will return to these questions after a review of other approaches to electrodynamics and of force laws.

8. Alternative approaches to electrodynamics I—propagated potential

We have examined briefly the four Maxwell's equations and their evidence or lack thereof. Now let us look into other approaches of electrodynamics which bypass Maxwell's equations. One such approach generalises the two potentials from quasistatic situations to more dynamic ones and assumes that the effect of the potentials propagates at a finite speed, usually the speed of light. The idea of propagated potential was first proposed and developed by Bernhard Riemann, Ludvig Lorenz and Carl Neuman [9]. In modern times, Oleg D Jeffimenko called this line of development 'electromagnetism according to the causality principle'. Jeffimenko expressed this principle very clearly in [38], so we will just quote him to explain the principle:

> According to this principle [of causality], all present phenomena are exclusively determined by past events. Therefore equations depicting causal relations between physical phenomena must, in general, be equations where a present-time quantity (the effect) relates to one or more quantities (causes) that existed at some previous time. An exception to this rule are equations constituting causal relations by definition; for example, if force is defined as the cause of acceleration, then the equation $\mathbf{F} = m\mathbf{a}$, where $\mathbf{F}$ is the force and $\mathbf{a}$ is the acceleration, is a causal equation by definition.
>
> In general, then, according to the principle of causality, an equation between two or more quantities simultaneous in time but separated in space cannot represent a causal relation between these quantities. In fact, even an equation between quantities simultaneous in time and not separated in space cannot represent a causal relation between these quantities because, according to this principle, the cause must precede its effect. Therefore the only kind of equations representing causal relations between physical quantities, other than equation representing cause and effect by definition, must be equations involving 'retarded' (previous-time) quantities.

In the main line of development, the retarded time potentials formulas, first proposed by Lorenz in 1867, are used as the basis for modelling electromagnetic phenomena:

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0}\int_V \frac{\rho(\mathbf{r}', t_r)}{R}\mathrm{d}V', \qquad \mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi}\int_V \frac{\mathbf{J}(\mathbf{r}', t_r)}{R}\mathrm{d}V', \tag{26}$$

where t_r represent the retarded time, and t the present time. The $\mathbf{E}$ and $\mathbf{B}$ fields are then computed using the following equations, seen in the previous section:

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla\times\mathbf{A}. \tag{27}$$

A lengthy series of algebraic manipulations yields a generalisation of Coulomb's formula for the electric field:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0}\int_V \left[\frac{\rho(\mathbf{r}', t_r)}{R^2}\hat{R} + \frac{\dot{\rho}(\mathbf{r}', t_r)}{cR}\hat{R} - \frac{\dot{\mathbf{J}}(\mathbf{r}', t_r)}{c^2R}\right]\mathrm{d}V', \tag{28}$$

and a generalisation of the Biot–Savart formula for the magnetic field:

$$\mathbf{B}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int_V \left[\frac{\mathbf{J}(\mathbf{r}', t_r)}{R^2} + \frac{\dot{\mathbf{J}}(\mathbf{r}', t_r)}{cR} \right] \times \hat{R} \mathrm{d}V'. \tag{29}$$

These time-dependent generalisations of the Coulomb and Biot–Savart formulas appear to have been first published by Jefimenko in 1966 in section 15-7 of his textbook [39] and are often called the Jefimenko's equations. These formulas should be compared to (3) and (15). Nivaldo A. Lemos aptly summarised the usual derivation, saying [40]

> The standard argument for retarded potentials runs as follows. Consider the contribution to the potential at point $\mathbf{r}$ from the charges and currents in the volume element $\mathrm{d}V'$ about point $\mathbf{r}'$. Since electromagnetic influences travel at speed c, the potentials at $\mathbf{r}$ at time t must have originated in the charges and currents presented in $\mathrm{d}V'$ at an instant t_r previous to t so providing electromagnetic influences with the time interval $t - t_r$ to propagate from $\mathbf{r}'$ to $\mathbf{r}$. The distance between $\mathbf{r}$ and $\mathbf{r}'$ is the magnitude of the vector $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, that is $R = |\mathbf{r} - \mathbf{r}'|$. Thus, the retarded time t_r is determined by $c(t - t_r) = R$, or
>
> $$t_r = t - \frac{R}{c}.$$

Physical intuition suggests that the potentials be given in terms of their sources by the same expressions valid in electrostatics and magnetostatics except for the replacement of the charge and current densities by their values at the retarded time.

Lemos went on in the same paper to obtain the time-dependent generalisations of Coulomb and Biot–Savart equations without using the retarded potentials. He applied the same reasoning to the fields equations (3) and (15) directly. There seem to be several related ways to derive the Jefimenko's equations, which have been shown to be equivalent to other equations differing in appearance derived by Panofsky and Phillips [41, section 14.3].

Let us examine these equations in more detail and interpret the meaning of all the symbols contained within them. The generalised Coulomb's formula, (28), gives three possible sources for an electric field, ρ, $\dot{\rho}$ and $\dot{\mathbf{J}}$, respectively the charge density, the time rate of change of the charge density and the time rate of change of a current density. In essence, an electric field arises from charges at rest, charges with some non-zero velocity and charges accelerating. The first two sources, ρ and $\dot{\rho}$, are along the direction of the displacement between the field and the source point, $\hat{\mathbf{R}}$, while the last source $\dot{\mathbf{J}}$ is along a generally different direction, the direction of the changes in current density. There is also a negative sign in front of the last source indicating its contribution to the field is in the opposite direction to the changes in current density. Jefimenko has argued that this last source term is the one responsible for the induction of an electric current in a conductor from a changing electric current in another. For a wire, the direction of this field contribution will be along the direction of the current, and only last as long as the current is changing. On a nearby conductor, this field will cause electric charges to move parallel or anti-parallel relative to the direction of the current; in essence, a dragging force not an attracting or repelling force due to charges. The generalised Biot–Savart formula (29) gives two sources for a magnetic field, $\mathbf{J}$ and $\dot{\mathbf{J}}$, a current density and the time rate of change of the same current density. Since an electric and magnetic field share a common source when the current density varies with time, they will always occur together in these dynamic situations, which would explain the correlation between the two fields in the Maxwell–Ampere and Faraday equations.

In practice, the correction due to the retardation effect in the near-field region is minimal and is experimentally undetectable among the uncertainties of measurement. For instance, let us suppose that the average time variation in the charge and current distribution of our experiment is t_a. Let $R_a = ct_a$ be the distance travelled by a light ray in that time. The near-field region is any distance R from the charges and current much less than R_a, $R \ll R_a$. In the near-field region the difference between the retarded time and the actual time is relatively small when compared to t_a:

$$\frac{|t - t_r|}{t_a} = \frac{R}{R_a} \ll 1.$$

To get an estimate of the correction due to the retardation effect, we expand the charge and current density in a Taylor series expansion centred at the current time t, then put these expressions into the generalised Coulomb and Biot–Savart formulas and simplify. For the near-field region, these corrections are taken to be zero, and the basic Coulomb and Biot–Savart formulas can be used as an approximation to their generalisations.

Let us look at the fields in the far field region, defined as the region where $R \gg R_a$. The dominant contributions to the fields are the terms which vary like R^{-1}, giving the approximation

$$\mathbf{E} \approx -\frac{1}{4\pi c^2}\int \frac{1}{R}\frac{\partial(\mathbf{J} - (\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}})}{\partial t}\mathrm{d}V', \tag{30}$$

and

$$\mathbf{B} \approx -\frac{\mu_0}{4\pi c}\int \frac{1}{R}\frac{\partial(\mathbf{J} - (\mathbf{J}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}})}{\partial t} \times \hat{\mathbf{R}}\mathrm{d}V', \tag{31}$$

obtained with the use of the following series of equations:

$$\dot{\rho}(\mathbf{r}', t_r) = -\nabla \cdot \mathbf{J}(\mathbf{r}', t_r) = \frac{\partial \mathbf{J}(\mathbf{r}', t_r)}{\partial t} \cdot \frac{\hat{\mathbf{R}}}{c} + O(1/R^2).$$

For charges and currents localised in the vicinity of $\mathbf{r}_b$, let $\mathbf{R}_b = \mathbf{r} - \mathbf{r}_b$ and $R_b = |\mathbf{R}_b|$. With the assumption that the region containing the charges and currents have extension much less than R_b, the distance between the field point and a source point $\mathbf{r}'$ in the region will approximately be R_b. Thus the $\mathbf{E}$ and $\mathbf{B}$ fields are approximately given by

$$\mathbf{E} \approx -\frac{1}{4\pi c^2 R_b}\int \frac{\partial(\mathbf{J} - (\mathbf{J}\cdot\hat{\mathbf{R}}_b)\hat{\mathbf{R}}_b)}{\partial t}\mathrm{d}V', \tag{32}$$

and

$$\mathbf{B} \approx -\frac{\mu_0}{4\pi c R_b}\int \frac{\partial(\mathbf{J} - (\mathbf{J}\cdot\hat{\mathbf{R}}_b)\hat{\mathbf{R}}_b)}{\partial t}\mathrm{d}V' \times \hat{\mathbf{R}}_b. \tag{33}$$

These approximate $\mathbf{E}$ and $\mathbf{B}$ fields are mutually perpendicular to each other, $\mathbf{E}\cdot\mathbf{B} = 0$, and satisfy the ratio $E/B = c$, characteristics of fields in electromagnetic waves radially moving away from their source. Thus the $\mathbf{E}$ and $\mathbf{B}$ fields generated by bounded charges and currents, as found in a transmitting radio antenna, in the far field region are just identified with electromagnetic waves. It is easy to see how the notion arose that time-varying electric and magnetic fields in radio waves and light beams ‘generate’ each other as they propagate away from their true sources, electric charges and currents, especially when their sources are not in the local picture. Some would describe the situation as the fields detaching from their source and becoming independent.

The assumption that electromagnetic interactions propagated at a finite speed, which gives rise to retardation effects, seems reasonable and is based upon our experiences with light beams. Still, the possibility exists that some interactions may be instantaneous or much faster than light. This possibility needs to be ruled out since quantum entanglement experiments indicate that instantaneous connections are possible over distances greater than what is coverable by a light signal in the time scale of the experiments. Many believed that crucial experiments already exist which indicate electromagnetic interactions can be no faster than light. We would need to examine the various experiments in the domain of relativistic theories.

Let us relate (26), (28) and (29) to Maxwell's equations. Assuming the continuity equation (1), the two constitutive equations (8) and (11), and the generalised Coulomb and Biot–Savart formulas (28) and (29), Jefimenko was able to obtain all of Maxwell's equations and a few others [38]. The two constitutive equations are not needed if Maxwell's equations are expressed only in terms of the $\mathbf{E}$ and $\mathbf{B}$ fields. Jose A Heras mathematically proved that if there are two retarded fields equations with the same form as the Jefimenko's equations with regard to the sources $\rho(\mathbf{x}, t)$ and $\mathbf{J}(\mathbf{x}, t)$, with the sources satisfying the continuity equation, then these fields satisfy equations similar in form to Maxwell's equations. For ρ and $\mathbf{J}$ identified with the charge and current densities of electromagnetism, the equations are exactly Maxwell's equation [42]. In fact, O'Rahilly reproduced an argument by Levi-Civita in 1897 of deriving Maxwell's equations from just the retarded potentials along with the continuity, constitutive equations and the equations which relate the fields to the potentials. O'Rahilly went on to show that the propagated potentials contain more info than Maxwell's equations, by using the potentials to derive the Lineard-Schwarzschild potential [9, chap VII]. All of these derivations show that other approaches to electromagnetic theory based upon equations other than Maxwell's are possible. They also show that Maxwell's equations are not fundamental and may be bypassed along with their controversies. An important point to note, Heras theorem indicated that Maxwell's equations, the Jefimenko's equations, and the continuity equation are mathematical relationships between abstract entities. The only time that physics comes into play is when we use these abstract entities to model phenomena that we can measure, like forces, charges, and currents.

9. Force laws: Maxwell, Lorentz, Weber, Ritz

To measure electric or magnetic fields, we need to relate them to forces on a test charge. The force laws are independent of Maxwell's equations and must be assumed and independently verified by experiments. True, the first Maxwell's equation is just the Coulomb force law which holds for static charges, so there is still a need for a more general force law. In recent time, the most commonly used equation relating fields to force is the Lorentz force law, which has the following form for a charge q moving with velocity $\mathbf{v}$:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{34}$$

The first term $q\mathbf{E}$ is called the electric force, while the second term $q\mathbf{v} \times \mathbf{B}$ is called the magnetic force. The theoretical physicist H A Lorentz derived the force formula named in his honour as follows, with our comments in brackets:

> It is got by generalising the results of electromagnetic experiments. The first term represents the force acting on an electron in an electrostatic field $[\mathbf{F}_1 = e\mathbf{E}]$ … On the other hand, the part of the force expressed by the second term may be derived from the law according to which an element of a wire

> carrying a current is acted upon by a magnetic field [$\mathbf{F}_2 = I\mathbf{ds} \times \mathbf{B}$] … Now, simplifying the question by the assumption of only one kind of moving electrons with equal charges and a common velocity, we may write [$I\mathbf{ds} = e\mathbf{v}$] … After having been led in one particular case to the existence of the force [$\mathbf{F}_1 = e\mathbf{E}$] and in another to that of the force [$\mathbf{F}_2 = e\mathbf{v} \times \mathbf{B}$], we now combine the two in the way shown in the equation, going beyond the direct result of experiments by the assumption that in general the two forces exist at the same time [43].

Two questionable leaps of logic can be pointed to in this derivation. One, it combines two force laws from incompatible situations, the electrostatic one with charges at rest and the magnetic one with charges moving. Two, experiments on *neutral* currents were used, yet the derivations are for a charge in motion, contradicting charge neutrality. More recent derivations are not much better. Perhaps the best way to resolve these difficulties is to forget the derivations and just take this formula as a hypothesis to be tested with experiments, which may have been Lorentz' intention. Another possibility is to take it as the definition of electric and magnetic forces as it is done in some textbooks. Let us explore some history of this law and examine some of its difficulties.

A force law that is superficially similar to the Lorentz force appeared as equation (77) in Maxwell's 1961 treatise, as equation (D) in 1965 and as equation (B) in vol 2 of his final work, but in a form hardly recognisable as being similar:

$$
\begin{aligned}
P &= \mu\left(\gamma\frac{\mathrm{d}y}{\mathrm{d}t} - \beta\frac{\mathrm{d}z}{\mathrm{d}t}\right) - \frac{\mathrm{d}F}{\mathrm{d}t} - \frac{\mathrm{d}\psi}{\mathrm{d}x},\\
Q &= \mu\left(\alpha\frac{\mathrm{d}z}{\mathrm{d}t} - \gamma\frac{\mathrm{d}x}{\mathrm{d}t}\right) - \frac{\mathrm{d}G}{\mathrm{d}t} - \frac{\mathrm{d}\psi}{\mathrm{d}y},\\
R &= \mu\left(\beta\frac{\mathrm{d}x}{\mathrm{d}t} - \alpha\frac{\mathrm{d}y}{\mathrm{d}t}\right) - \frac{\mathrm{d}H}{\mathrm{d}t} - \frac{\mathrm{d}\psi}{\mathrm{d}z}.
\end{aligned}
$$

Referencing page 486 of Maxwell's 1965 treatise, the components of 'electromagnetic momentum' are (F, G, H). The components of 'magnetic intensity' are (α, β, γ) and the components of 'electromotive force' [per unit charge] is (P, Q, R). From [19], Maxwell had the following explanation after the equation:

> The first term on the right-hand side of each equation represents the electromotive force arising from the motion of the conductor itself. This electromotive force is perpendicular to the direction of motion and to the lines of magnetic force; and if a parallelogram be drawn whose sides represent in direction and magnitude the velocity of the conductor and the magnetic induction at that point of the field, then the area of the parallelogram will represent the electromotive force due to the motion of the conductor, and the direction is perpendicular to the plane of the parallelogram. The second term in each equation indicates the effect of changes in the position of the strength of magnets or currents in the field. The third term shows the effect of the electric potential ψ. It has no effect in causing a circulating current in a closed circuit. It indicates the existence of a force urging the electricity to or from certain definite points in the field.

In the modern era, electromagnetic momentum is identified with the vector potential $\mathbf{A}$, and magnetic intensity is identified with the magnetic field $\mathbf{H}$. Identifying Maxwell electric potential ψ with what is called the scalar potential ϕ, the equation can be written in an easier to read form as

$$\mathbf{f} = \mu(\mathbf{v} \times \mathbf{H}) - \frac{\partial \mathbf{A}}{\partial t} - \nabla\phi,$$

where $\mathbf{f}$ is the force per unit charge. It is not clear if identifying ψ with ϕ is justifiable in all situations, since in his 1861 paper and his last treatise, Maxwell says that ψ is indeterminate but can be determined once all the circumstances of a problem are known. A more detailed analysis of what Maxwell meant by ψ can be found in [6]. To put the above equation in a more recognisable form, with the proviso just mentioned, multiply both sides of the equation by the value of the charge e and use (11) and (25) to obtain the force on the charge. The important term for our discussion is the term involving the velocity, the magnetic force term. Because of this formula, the Lorentz force law is sometime called the Maxwell–Lorentz force law in the literature.

A common problem with many presentations of the Lorentz force formula is the ambiguity in the standard of rest of the velocity. In a lab experiment located on Earth with the test charge 'q' in motion relative to the lab, an observer, sources of magnetic fields, all moving with different velocities relative to each other, $\mathbf{v}$ is one of these velocities or some combination of these velocities added together. Since Earth-based laboratories also partake of the Earth orbital and rotational velocities, these velocities may need to be taken into account in the design and analysis of experiments, like ones involving distant starlight.

Maxwell's explanation immediately after his electromotive force equation indicates the velocity in respect to the magnetic field, which in his model is a state of a stationary ether. This velocity has undergone some reinterpretations over the years as one theory has gained favour over another. We will give a brief summary of these various interpretations, which is covered in more detail in [44, appendix A]. According to Whittaker the first to arrive at the magnetic force were J J Thomson in 1881 and O Heaviside in 1889 [8]. In Thomson's 1881 paper, on p 248, he says, 'It must be remarked that what we have called the actual velocity of the particle is, in fact, the velocity of the particle relative to the medium through which it is moving … medium whose magnetic permeability is μ.' In Heaviside's 1889 paper, 'On the electromagnetic effects due to the motion of electrification through a dielectric', he did not make a clear statement on what is the velocity $\mathbf{v}$. Given the title of his paper, it is likely to be the velocity of 'q' relative to the medium of magnetic permeability μ and dielectric constant ϵ. A more in-depth discussion of the works of these two researchers can be found in Buchwald's book [5].

In 1895 Lorentz presented the expression for the force law that now bears his name. Assis has this to say in his summary of electromagnetic theory [44]:

> Lorentz did not comment on the works of Thomson and Heaviside, and arrived at the magnetic part of this expression from Grassmann's force, substituting $q\mathbf{v}$ for $I\mathbf{ds}$, although he did not mention Grassmann's work as well … Lorentz did not specify in (34) what is the object, medium or system relative to which the velocity $\mathbf{v}$ of the charge q is to be understood. As Lorentz still accepted Maxwell's ether … it is natural to suppose that for him this was the velocity of the charge q relative to this ether, and not relative to any other medium or observer. In support of this statement we have Lorentz's own words on this same page 14: 'Now, in accordance with the general principles of Maxwell's theory, we shall consider this force as caused by the state of the ether, and

> even, since this medium pervades the electrons, as exerted by the ether on all internal points of these particles where there is a charge.' [45, p 14]

After Albert Einstein's special theory of relativity became widely accepted, many people adopted his interpretation of $\mathbf{v}$ in the Lorentz force as the velocity relative to an observer or measuring instrument. This interpretation was introduced in Einstein's 1905 special relativity paper, where he applied Lorentz's transformation on the Coulomb force on a charge in the rest frame of the charge to obtain the force as it would appear in a moving frame. A quote from p 54 of his paper might give a sense of what Einstein did:

(i) If a unit electric point charge is in motion in an electromotive field, there sets upon it, in addition to the electric force, an 'electromotive force' which … is equal to the vector product of the velocity of the charge and the magnetic force … (old manner of expression).

(ii) If a unit electric point charge is in motion in an electromagnetic field, the force acting upon it is equal to the electric force which is present at the locality of the charge, and which we ascertain by transformation of the field to a system of coordinates at rest relatively to the electrical charge (new manner of expression).

If the force in the rest frame of the charge q is $q\mathbf{E}$, then in a frame moving with velocity $\mathbf{v}$ relative to the charge the force appears to be $q(\mathbf{E}' + \mathbf{v} \times \mathbf{B}')$, where the $\mathbf{E}'$ and $\mathbf{B}'$ fields are obtained from applying Lorentz's transformation on $\mathbf{E}$. From the perspective of the moving frame, the charge could be viewed to be in motion while the frame is at rest. Thus, in a manner of speaking the field $\mathbf{E}' + \mathbf{v} \times \mathbf{B}'$ is the perceived electric field $\mathbf{E}$ experienced by the charge q within its own rest frame.

In the final analysis, Maxwell, Thomson, Heaviside, Lorentz, and Einstein use different terms for the velocity standard of rest in the Lorentz force law, but they are mainly saying velocity with respect to some inertial reference frame, whether it is called the ether or an observer. In most Earth-based experiments, the lab frame is often the approximate inertial frame of choice, where the detector or observer is at rest with respect to planet Earth and partakes of its orbital motion. In particular, to find the path of a charged particle in a lab, we often use an inertial frame where Newton's second law of motion, relating the applied forces to the charge's acceleration, can be expressed without fictitious forces. In the past, especially in association with observations involving starlight, the lab frame at rest with respect to the Earth was called the Earth-convected ether. Conventionally, these frame-based velocities are sometimes called 'absolute' velocities, since the velocities of different parts of an experimental setup are not relative to each other, but to some frame of reference which serves as the standard of rest.

The ambiguity with the velocity in the Lorentz force law is not quite as serious as the next difficulty. The Lorentz force does not satisfy Newton's third law in the strong form—equal and opposite reaction along the same line, needed to conserve linear momentum. This feature of the law is most apparent when it is used to calculate the forces between two isolated moving charges. To explain this feature of the Lorentz force law, many theorists reasoned that when the fields' momenta are taken into account, the laws of conservation of linear and angular momentum still hold for isolated systems. As pointed out by Henri Poincaré, a highly regarded mathematician and physicist, it is in fact very hard to show that the law of conservation of linear momentum is violated under laboratory conditions when there are electromagnetic waves involved [46]:

> To demonstrate experimentally that the principle of reaction is indeed violated in reality, as it is in Lorentz's theory, it is not sufficient to show that the device

> producing the energy recoils, which would already be very difficult, it is necessary to also show that the recoil is not compensated by the movement of the dielectric, and in particular the motion of the air traversed by the electromagnetic waves.

Despite the difficulties, some researchers have designed experiments to test the nonconformity of the Lorentz force law to the law of conservation of momentum. As of this date, no one has conducted these experiments, since the feeble effects predicted by these experiments are hard to detect; see, for instance, the interesting book of Junichiro Fukai [37].

If the Lorentz force law is taken as a hypothesis to be tested and confirmed by experiments, we run into the problem of insufficient data. In fact, direct supporting experimental data are hard to come by, with many journal articles saying that high-speed particles experiments confirm this law to a high degree without giving citations. Alternatively, they may cite the experiment of Bailey *et al* as evidence. But the experiment of Bailey *et al* was never designed to test the Lorentz force law, which was assumed to hold so that time-dilation effects can be measured. Other researchers think that it has not been tested adequately in the realm of high speed and have designed experiments to meet their objections; see [47] for a more detailed summary and a more exhaustive bibliography of articles investigating the Lorentz force law in the relativistic region. In any case, a detailed analysis of these experiments involves a theory of relativity of some kind and will be done in a future article.

In the realm of low speed, we do have more theoretically accessible experiments examining different force laws. Before we get to these experiments, we need to look at a force law that will share the spotlight in upcoming discussions. The Weber–Ritz force law between two charges, up to the first order of the factor $1/c^2$, is given by the formula

$$\begin{aligned}\mathbf{F} = &\frac{q_1 q_2}{4\pi\epsilon_0 r^2}\left[1 + \frac{3-k}{4}\left(\frac{v}{c}\right)^2 - \frac{3(1-k)}{4}\left(\frac{\mathbf{v}\cdot\mathbf{r}}{c^2}\right)^2 - \frac{r(\mathbf{a}\cdot\mathbf{r})}{2c^2}\right]\hat{\mathbf{r}} \\ &- \frac{q_1 q_2}{4\pi\epsilon_0 r^2}\left[\frac{k+1}{2c^2}(\mathbf{v}\cdot\mathbf{r})\mathbf{v} + \frac{r\mathbf{a}}{c^2}\right], \end{aligned} \tag{35}$$

with $\mathbf{r}$, $\mathbf{v}$, and $\mathbf{a}$, being the radial separation, relative velocity and relative acceleration of the two charges respectively and k some number to be determined by experiments. Weber's law is obtained when $k = -1$ and the last term in the direction of the acceleration is omitted. Since Weber and Ritz force formulas give the same result when integrated over a complete circuit, with all the terms having the k factor cancelling, the parameter k can only be determined from experiments that do not involve closed circuits. Such experiments would involve charged particle streams in open circuits. O'Rahilly and later Smith *et al* [48] explained that the results of the Kaufmann–Bucherer experiment could be interpreted to imply that $k = 3$. The Weber–Ritz force law does not satisfy Newton's third law when the acceleration term is non-zero. This is to be expected since the acceleration term has been shown to be responsible for radiation of energy, making the reaction force unequal to the action force.

R T Smith, F P M Jjunju and S Maher conducted a series of experiments with electron beam deflections across a solenoid to compare the fitness of the Maxwell–Lorentz and the Weber–Ritz force laws [48]. They concluded,

> Our results indicate that, within the limits of experimental error, both Weber–Ritz and Maxwell–Lorentz theories correlate with measurements for the longer solenoids. However, in the case of the shortest solenoid, the lack of uniformity of the magnetic field, leads to significant error in the calculation of beam

> deflection by the Lorentz force. By contrast in a Weber–Ritz calculation a precise value of beam deflection is obtained by equating the impulse of the non uniform beam force to the vertical momentum change of the electron …

In another series of experiments comparing the force laws of Lorentz, Weber, and Ritz, R T Smith and S Maher gave the following conclusion [49]:

> In the deflection of electron beams by a long, straight wire carrying direct current, the electrodynamic theories of both Weber and Ritz have been tested against the field-based approach of Maxwell–Lorentz. For 'non-relativistic' beams all three theories give identical results which agree with experimental observations. It is of interest that field theory, direct action and emission theories of electrodynamics, each based on essentially different assumptions, give the same result.

Their conclusion clearly indicates that it is the formulas derived from the theories that experiments can test and is of any significant, not the various conceptions behind the theories. As O'Rahilly succinctly put it, 'What is algebraically ineffective must be regarded as scientifically irrelevant' [9, p 464]. These researchers continued with an interesting statement:

> In the case of high-speed electrons, it is well known that electron trajectories must be corrected, according to the Lorentz factor, $\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2}$. In typical crossed-field experiments for high-speed electrons, it is shown that Weber–Ritz provides the same result as the established theory …

To be clear, these investigators followed O'Rahilly's analysis to get the Weber–Ritz result to agree with the Lorentz factor but only up to the second order of v/c, when the Lorentz factor is approximated by a Taylor series expansion. As far as we know, a detailed analysis of the fitness of the Weber–Ritz force formula has not been done for high-speed charged particles where v/c approaches 1 and higher order terms of v/c in the series expansion cannot be neglected.

We have briefly examined the Lorentz force law as a hypothesis to be tested with experiments, but more often it is used as the definition of the electric and magnetic fields. Using the Lorentz force law to define the fields makes testing it independently of the fields impossible, but it can be tested for consistency with other formulas for the electric and magnetic fields, in particular, the fields given by the generalised Coulomb and Biot–Savart formulas (28) and (29). The fields must be independently measured in some way and are usually done by some kind of force measurement. The various experiments to test these forces or fields use instruments that have been calibrated using one of these force laws. One can begin to see the field definitions circling back to forces and the need for a force formula independent of fields.

10. Alternative approaches to electrodynamics II—direct action theories

Earlier we mentioned some alternative approaches to electrodynamic theories, some even completely dispensing with fields. Let us explore these direct action theories in more depth. The direct action approach to electrodynamics was used before the concept of fields was developed and was used by the likes of Ampere, Gauss and Weber. As mentioned in the introduction, Wilhelm Weber, a contemporary of Maxwell, was the first investigator to formulate a microscopic electrodynamic theory modelling electromagnetic phenomena from the forces between electric charges in relative motion. He did this before Maxwell developed his equations, using a force law that depends not only on the relative displacement between the charges but also on their relative radial velocity and acceleration [8]. In 1857, Weber and Kirchhoff independently obtained the wave equation for an electromagnetic signal utilising Weber's force formula. Despite its many successes in modelling various electromagnetic interactions, Weber's theory was broadly abandoned in favour of Maxwell's because of several flaws that were not addressed until Walter Ritz revitalised this approach in 1908, with his derivation of (35). Ritz had this to say about the formulation of his force law [50]:

> Experience has shown that actions are not instantaneous; also it hasn't revealed any trace of a medium which could exist in materially empty space. I therefore felt I could restrict myself to give to the law of propagation of these actions, a very simple kinematic interpretations borrowed from the emanative theory of light and satisfying the principle of relativity of motion. Fictitious particles are constantly emitted in all directions by electric charges; they keep on moving indefinitely in straight lines with constant speed, even through material bodies. The action undergone by a charge depends uniquely on the disposition, velocity, etc, of these particles in its immediate surroundings. The particles are therefore simply a concrete representation of kinematic and geometric data.

Because Ritz used the emission of fictitious particles in deriving his force law, some researchers might classify Ritz' EM theories as an emission theory, similar to quantum electrodynamics, wherein EM forces are mediated with the exchange of particles. We will classify it as a direct action theory since these fictitious particles just represent kinematic and geometric info. Preceding Ritz, Schwarzschild and others not mentioned here used a direct action approach with a force formula similar to Ritz', but one involving absolute, frame-based velocities instead.

Reproducing the works of Schwarzschild and Ritz, O'Rahilly did a mathematical analysis comparing the Lorentz force and the Ritz force laws in explaining electrical phenomena occurring at relatively low speed. To summarise, O'Rahilly used the retarded time potentials of Riemann–Lorenz, transformed the integrals into sums, by finding the limiting ratio of two volume elements, to obtain the Lienard–Wiechert potential for discrete charges. This step was needed to transition from a macroscopic scale to a microscopic one. Then he used the Lienard–Wiechert potential to find the electric and magnetic fields generated by a single charge for substitution into the Lorentz force law to obtain the Lienard–Schwarzchild force formula for the force between two charges in term of their frame separation, velocity and acceleration at the retarded time. He then approximated the Lienard–Schwarzchild force formula with a Taylor series expansion, so that distance, velocity, and acceleration are evaluated at the present-time variable. The approximation neglects all terms with order higher than one of the factor $1/c^2$, and resulted in a force formula free of fields for microscopic charges moving at speed much less than the speed of light. Using this microscopic force law and an averaging process over aggregations of charge complexes, O'Rahilly was able to explain most macroscopic electromagnetic phenomena, including electrical induction, electromagnetic waves, and radiation pressure, *when the velocity and acceleration of the charges are taken with respect to the lab frame.* He was also able to explain the same set of macroscopic electromagnetic phenomena using the Ritz force law (35), *without using the concept of electric or magnetic fields.* By 'explain' we mean that with the use of either force laws O'Rahilly was able to obtain the equation that model induction (19), and the equation which express the radiation pressure, equations that are backed by experiments to some degree without using potentials or fields of any kind. Another researcher, Junichiro Fukai, building upon the work of Assis [44], argued that Weber's force law is capable of explaining most of classical electrodynamics including radiations when certain assumptions on the vacuum are used. The relevant part of his arguments for EM radiation is given below [37]:

> If an electrical disturbance occurs at a point of space that is readily polarisable, we can imagine a charge separation formed immediately in the vicinity of the source. If we assume, for simplicity, that the disturbance creates a charge separation in the shape of a coaxial cable for an infinitesimal distance of say dx … The vacuum may be considered as an ideal 'dielectric' substance and yet without resistance since it is filled with pairs of positive and negative [charged] particles with negligible masses. Thus, we consider electric signals propagating in the vacuum medium that has an equivalent circuit consisting of a series of LC components (per dx) connected to infinity … The result is the telegrapher's equation that describes the signal propagation with speed $v = 1/\sqrt{LC}$, where we can calculate L and C for a coaxial configuration. We obtain, for either Weber's theory or Maxwell's theory, $L = (\mu_0/2\pi)\ln(b/a)$ and $C = 2\pi\epsilon_0/\ln(a/b)$ where constants a and b are arbitrary. The propagation speed is readily seen as $v = 1/\sqrt{\epsilon_0\mu_0} = c$, the speed of light in free space … Therefore, it is possible to add the propagation of electrical signals in space according to Weber's theory by implementing some aspects of modern view in the vacuum.

Note that in Ritz' and Weber's theories, radiation, the transfer of energies or electromagnetic waves are interactions of charge complexes. In these theories, radiation can only happen when at least two charges are interacting. If there is a lone charge with no corresponding receiver charge, no electromagnetic waves will be emitted. The advantage of O'Rahilly approach, really Ritz' approach, is that no properties of the vacuum are assumed.

If we accept O'Rahilly's or Ritz' detailed mathematical analysis as correct, we are forced to accept that both electromagnetic theories, one based upon the Lorentz force law and one based upon Ritz force law, are capable of explaining most of classical electrodynamics for relatively small charge velocities. Returning to the unanswered question on what transfers energy to a lamp when an electrical switch is turned on, we can use electromagnetic fields, or we can use forces that cause polarisation of the ether to explain the transfer of energy. We can use anything that we want as long as the explanation yield the equations that have been tested by experiments. In any case, experiments can only test the fitness of our mathematical propositions and equations in modelling physical phenomena and not the various ideas used to obtain the equations. If our conception alters previous formulas or come up with new ones, we may be able to test the fitness of the resulting equations in modelling physical phenomena, but not the conceptions used to construct those formulas. This does not imply that all ideas or models of classical electromagnetism are equal. One may be easier to understand by beginning students, one may be more economical in the number of assumptions used, or one may be more useful than another in modelling specific sets of phenomena. The only way to be sure is to find the limits or domain of validity of these theories.

11. Closing remarks

Our investigation into the limits of Maxwell's equations revealed that the two most important ones, the Maxwell–Ampere and the Maxwell–Faraday laws, are not well supported by experimental evidence on the same level of accuracy as the first two. Stronger evidence for these two laws would be of interest, but not crucial, since explorations of other approaches to classical electrodynamics showed that Maxwell's equations are not fundamental and that classical electromagnetism may be developed without them. In fact, the Maxwell–Ampere and the Maxwell–Faraday equations can be derived from the first Maxwell's equation and the Biot–Savart law, the continuity equation and the assumption of the finite speed of propagation of electromagnetic interaction, all of which are backed by experimental evidence. We also uncovered indications that the field concept may be dispensed with, at least in macroscopic classical electromagnetism since most of classical electromagnetism in the realm of relatively low speed can be build up from two competing force formulas—one based upon frame velocities found in the Maxwell–Lorentz theory, and one based upon relative velocities between charges found in the Weber–Ritz theory. Our investigation so far indicates that electrodynamic experiments involving speed much less than that of light are inconclusive as to which theory better fits experimental data.

The scope of our review became more extensive than expected and will have to await further investigations to complete the process. One such investigation would focus on experiments involving high-speed particles and light itself, which requires a critical examination of various theories of relativity and their supporting evidence. Another investigation or series of investigations would focus on the limitations and area of useful applications of the Weber–Ritz force formula, its domain of validity. In particular, it would be of interest to see if the Weber–Ritz force law can explain the radiation spectrum from accelerating charged particles moving very close to light speed within synchrotron accelerators—a domain said to support the Lorentz force law very well. It would also be of interest to see if the Weber–Ritz force law can be usefully incorporated into plasma physics or quantum theory. Due to the tight focus of our review, our coverage of the history of classical electrodynamics has been superficial, those interested in a more thorough history up to recent times are directed to read Darrigol and Whittaker [7, 8]. O'Rahilly's book is also highly recommended for a critical examination of various electromagnetic theories including those of Maxwell, Lorentz, Weber, Ritz, and Einstein.

ORCID iDs

Max Tran https://orcid.org/0000-0003-2719-6675

References

[1] Maxwell J C 1861 On physical lines of force *London, Edinburgh Dublin Phil. Mag. J. Sci.* Fourth Series, Part I, II, III, IVAccessible online at https://en.wikisource.org/wiki/On_Physical_Lines_of_Force
[2] Duhem P 2015 *The Electric Theories of J Clerk Maxwell* (Basel: Springer) Les Theories Electriques de J. C. Maxwell (Paris) (https://doi.org/10.1007/978-3-319-1851-6)
[3] Manville O 1928 L'oeuvre Scientifique de Pierre Duhem (Paris-Bordeaux)
[4] Arthur J W 2008 The fundamentals of electromagnetic theory revisited *IEEE Antennas Propag. Mag.* **50** 19–65
[5] Buchwald J Z 1985 From Maxwell to microphysics *Aspects of Electromagnetic Theory in the Last Quarter of the Nineteenth Century* (Chicago, IL: University of Chicago Press)
[6] Darrigol O 1993 The electrodynamics of moving bodies from Faraday to Hertz *Centaurus* **36** 245–60
[7] Darrigol O 2000 *Electrodynamics From Ampere To Einstein* (Oxford: Oxford University Press)
[8] Whittaker E T 1960 *A History of the Theories of Aether and Electricity* (New York: Harper)
[9] O'Rahilly A 1965 *Electromagnetic Theory, A Critical Examination of Fundamentals v 1–2* (New York: Dover)

[10] Lorentz H A 1902 The fundamental equations for electromagnetic phenomena in ponderable bodies, deduced from the theory of electrons *Proc. R. Acad. Amsterdam* pp 254–66 (http://dwc.knaw.nl/DL/publications/PU00014180.pdf)
[11] de Groot S R and Suttorp L G 1972 *Foundations of Electrodynamics* (Amsterdam: North-Holland)
[12] Robinson F N H 1973 *Macroscopic Electromagnetism* (Oxford: Pergamon)
[13] Klapdor-Kleingroththaus H V, Krivosheina I V and Titkrova I V 2007 A new experimental limit for the stability of electron *Phys. Lett.* B **644** 109–18
[14] Particle Data Group 2010 Tests of conservation laws *J. Phys. G: Nucl. Part. Phys.* **37** 89–98
[15] Orito S and Yoshimura M 1985 Can the Universe be charged? *Phys. Rev. Lett.* **54** 2457–60
[16] Particle Data Group 2016 The review of particle physics *Chin. Phys.* C **40** 100001
[17] Jackson J D 1998 *Classical Electrodynamics* 3rd edn (Hoboken, NJ: Wiley)
[18] Heras R 2016 The Helmholtz theorem and retarded fields *Eur. J. Phys.* **37** 1–11
[19] Maxwell J C 1865 A dynamic theory of the electromagnetic field *Phil. Trans. R. Soc.* **155** 459–512
[20] Cavendish H 1879 *Electrical Researches of Henry Cavendish* ed J C Maxwell (Cambridge: Cambridge University Press) pp 104–13
[21] Plimpton S J and lawton W E 1936 A very accurate test of Coulomb's law of force between charges *Phys. Rev.* **50** 1066–71
[22] Tu L, Luo J and Gillies G T 2005 The mass of the photon *Rep. Prog. Phys.* **68** 77–130
[23] Goldhaber A S and Nieto M M 2010 Photon and graviton mass limits *Rev. Mod. Phys.* **82** 939–79
[24] Timm C 2015 *Theory of Magnetism* (*International Max Planck Research School for Dynamical Processes in Atoms, Molecules and Solids*)
[25] Matthis D C 2006 *The Theory of Magnetism Made Simple* (Singapore: World Scientific)
[26] Assis A K T and Chaib J P M C 2015 *Ampere's Electrodynamics* (Montreal: C Roy Keys Inc)
[27] Arthur J W 2009 An elementary view of Maxwell's displacement current *IEEE Antennas Propag. Mag.* **51** 58–68
[28] Maxwell J C 1873 *A Treatise On Electricity and Magnetism* vols 1–2 (Oxford: Claredon Press)
[29] Bartlett D F and Corle T R 1985 Measuring Maxwell's displacement current inside a capacitor *Phys. Rev. Lett.* **55** 59–62
[30] Bartlett D F 1990 Conduction current and the magnetic field in a circular capacitor *Am. J. Phys.* **58** 1168–72
[31] Ribeiro J E A, Vannucci A and Assis A K T 2008 The multiple definitions of 'field' in the context of electromagnetism *Proc. VI Taller Internacional 'ENFIQUI 2008'* pp 1–4
[32] Laughlin R B 2005 *A Different Universe: Reinventing Physics from the Bottom Down* (New York: Basic Books) pp 120–1
[33] Bridgman P W 1927 *The Logic of Modern Physics* (New York: Macmilan) pp 57–136
[34] Chen K, Li X J and Hui Y X 2017 An experimental study on unipolar induction *Acta Phys. Pol.* A **131** 271–3
[35] Muller F J 2014 Unipolar induction revisited: new experiments and the 'edge effect' theory *IEEE Trans. Magn.* **50** 1–11
[36] Feynman R, Leighton R and Sands M 1963 *The Feynman Lectures on Physics* vol 2 (Pasadena, CA: California Institute of Technology)
[37] Fukai J 2003 *A Promenade Along Electrodynamics* (Pueblo West, CO: Vales Lake Publishing)
[38] Jefimenko O D 2004 Presenting electromagnetic theory in accordance with the principle of causality *Eur. J. Phys.* **25** 257–74
[39] Jefimenko O D 1996 *Electricity and Magnetism* (New York: Appleton-Century-Crofts) Sec 15–7; same section in the 2nd edn, (Star City, WV: Electret Scientific)
[40] Lemos N A 2008 Physical intuition and time-dependent generalizations of the Coulomb and Biot–Savart laws arXiv:physics/0508047v1 [physics-ed-ph]
[41] Panofsky W K H and Phillips M 2005 *Classical Electricity* 2nd edn (New York: Dover)
[42] Heras J A 2007 Can Maxwell's equations be obtained from the continuity equation? *Am. J. Phys.* **75** 652–7
[43] Lorentz H A 1909 *The Theory of Electrons* (Leipzig: Teubner) p 14
[44] Assis A K T 1994 *Weber's Electrodynamics* (Norwell, MA: Kluwer Academic Publishers)
[45] Lorentz H A 1915 *The Theory of Electrons* 2nd edn (Leipzig: Teubner)
[46] Poincaré H 1900 The theory of Lorentz and the principle of reaction *Archives néerlandaises des Sciences Exactes et Naturelles series 2* **5** 252–78
[47] Huang Y-S 1991 Has the Lorentz-covariant electromagnetic force law been directly tested experimentally? *Found. Phys. Lett.* **6** 257–74

[48] Smith R T, Jjunju F P M and Maher S 2015 Evaluation of electron beam deflections across a solenoid using Weber–Ritz and Maxwell–Lorentz electrodynamics *Prog. Electromagn. Res.* **151** 83–93
[49] Smith R T and Maher S 2017 Investigating electron beam deflections by a long straight wire carrying a constant current using direct action, emission-based and field theory approaches of electrodynamics *Prog. Electromagn. Res.* B **75** 79–89
[50] Ritz W 1908 Critical researches on general electrodynamics *Ann. Chim. Phys.* **13** 145 Translated from *Recherches critiques sur l'Electrodynamique generale*