Maxwell's electromagnetic theory and special relativity

By Graham Hall*

*Department of Mathematical Sciences, University of Aberdeen,
Aberdeen AB24 3UE, UK*

This paper presents a brief history of electromagnetic theory from ancient times up to the work of Maxwell and the advent of Einstein's special theory of relativity. It is divided into five convenient periods and the intention is to describe these developments for the benefit of a lay scientific audience and with the minimum of technical detail.

Keywords: Maxwell; electromagnetism; special relativity

1. Introduction

This paper is written in commemoration of the 150th anniversary of James Clerk Maxwell's taking up, in 1856, the chair of natural philosophy in Marischal College, Aberdeen (later, in 1860, to merge with King's College to form the University of Aberdeen). It is the text of a lecture given by the author at a meeting held in Maxwell's honour in Aberdeen in September 2006. It should be added that this latter year is also the 175th anniversary of his birth, on 13 June 1831, in Edinburgh.

Maxwell's theory of the electromagnetic field is one of the main pillars of modern theoretical physics and, of course, played a key role in the formulation and the development of Einstein's special theory of relativity. The purpose of this article is to give a brief description of the events that led up to Maxwell's (1865) paper and his 1873 Treatise and which gave to the world the conclusions of one of the most fundamental pieces of scientific research ever undertaken. It will take the form of a brief history of the developments of electricity and magnetism from ancient times until 1905 when Einstein published his special theory of relativity. In keeping with the intention that the lecture and subsequent paper should be for a lay scientific audience, the mathematical details will be kept to a minimum. However, some mathematical language will be needed to aid coherency, especially in the later stages, but it is hoped that the paper can be read, with some profit, by those who wish or have to disregard such details.

The theory of the electromagnetic field, especially in its more advanced stages, has a complex history and many interesting sideshows must necessarily be omitted. However, the main developments will be included and, it is hoped, set in the correct context. The history of this subject, and thus the content of this

*g.hall@maths.abdn.ac.uk

paper, will be divided into five parts: first, the knowledge of the ancients, culminating in the work of Gilbert; second, the experimental period from Gilbert to the beginning of the nineteenth century; third, the short but exciting experimental and theoretical period from Oersted to the conclusion of the researches of Faraday; fourth, the period of Maxwell; and fifth, the period leading from Maxwell's great treatise to Einstein's special theory of relativity.

2. Early days

The original source of awareness in matters, electrical or magnetic, no doubt comes from natural causes such as the discharge of lightning or the pale glow sometimes seen on the sharp ends of certain objects such as ships' masts or even warriors' spears, during stormy weather. However, as a topic of scientific debate, it appears to have come from two naturally occurring materials, that is, *amber* and *loadstone*. Amber, a resin from a type of pine tree, often found in the appearance of stone washed up in certain coastal locations, was known to have the special property that, when rubbed, it had the power to attract a variety of light objects, such as pieces of straw or fur placed in its vicinity. Amber and its properties were known to the ancient Greek and Roman civilizations. Loadstone (or lodestone) is a naturally occurring magnetized piece of the iron ore called magnetite, whose ability to attract iron was also well known to these cultures. Thales of Miletus apparently experimented with amber and loadstone in *ca* 600 BC. The ancient Chinese civilization may well have known of loadstone much earlier than the Europeans. It has also been known, probably for at least 1000 years, that a long thin piece of loadstone will, if rested on a piece of wood which is then floated on water, turn so that the tips of its long axis point north and south. Thus, the mariner's compass was born and was known to the Chinese, Muslim and north European sailors. Although the respective properties of amber and loadstone contained many similarities, their basic difference, that amber attracted many different types of object whereas loadstone attracted only iron, seemed sufficient to keep them separated.

There is little of significance to report further on the knowledge of amber and loadstone until the scientific renaissance came to Europe. Then in the middle of the sixteenth century, the Italian, Jerome Cardan, promoted further the case for the difference in amber and loadstone by carefully listing and contrasting their respective properties. He stressed, as mentioned above, that amber would attract anything sufficiently light whereas loadstone attracted only iron. He also made the interesting remarks that the attractive ability of amber was destroyed when some material is interposed between it and its victim whereas that of loadstone was unaffected under similar conditions, and that, when loadstone attracted iron, the iron was also attracted by the loadstone whereas no such reciprocal action was observed with amber. Finally, he made the interesting observations that loadstone attracted towards its ends (poles) whereas no such 'polarity' was observed in amber and that, whereas friction increased (or was necessary for) the attractive powers of amber, no similar effect was observed with loadstone except, of course, by cleaning the poles. Cardan's work was followed by what is usually regarded as the most significant early piece of research on electricity and magnetism (mainly on magnetism) by William Gilbert (1544–1603). Gilbert, who

was born in Colchester, England, and educated at Cambridge University, published the influential book *De Magnete* (*On the magnet*) in 1600. In this book, he showed himself to be an exhaustive experimenter, finding a large number of materials which possessed the properties of amber, including the already known materials, jet (a kind of compacted coal) and diamond and named 'electrics' by him. He also gave a more complete list of which materials would be attracted by his electrics. In the process of these experiments, Gilbert invented what was, in a sense, the first electroscope (called, by Gilbert, the *versorium*); a fine needle of metal (or certain other substances) delicately balanced on a sharp point and which could rotate freely upon it. A rubbed material could then be brought close to it and even very small attractions between material and needle could be observed and recorded. Gilbert tested many hypotheses; in particular, he showed that heating amber is not enough to give it its attractive property. It is only when this heating comes from rubbing that this property is found. He re-examined the effect that one can shield the force of an electric but not that of a loadstone, coming to the same conclusion as Cardan. He also demonstrated that electric and magnetic attraction between two bodies decreased as the distance between them increased and was aware of the fact that, on making contact with an electric, an attracted body would soon fall away. In his experimental work on magnetism, he described what would now be called the *magnetic field* by displaying the familiar patterns around a magnet using iron filings. (Descartes' suggestion of a fluid vortex theory of magnetism could also be interpreted as containing the germ of the magnetic field concept.) Gilbert regarded the Earth as a large magnet (linking it to the rotation of the Earth) and was thus able to explain the directional properties of the compass. He was, in fact, responsible for the terms north and south poles. He was aware of the magnetic dip of the Earth (probably first studied in detail by Robert Norman in 1581) and the difference between the true and the magnetic poles. Gilbert, although he framed several hypotheses in his famous book, relied on the experimental method and, like his contemporaries, Galileo and Francis Bacon, was one of the original exponents of experimental philosophy.

3. Experimental development of electricity

In 1629, Niccolo Cabeo wrote an important work that listed even more materials that could be classified as electrics and also observed *repulsion* of electrics (but interpreted it as a 'bouncing-off' effect). Other later Italian workers observed and interpreted correctly the mutual attraction effect for electrics and experimented to see if the attractive properties of amber were due to the surrounding air by attempting the experiment *in vacuo*. Their attempt was inconclusive but a successful answer was provided by Robert Boyle in England in the 1670s (following the development of a better vacuum pump) when he showed the effect to be still present. In the early eighteenth century, a series of experiments was carried out by Francis Hauksbee in England. He observed the effects of electric attraction on suspended threads and revealed the *electric field* in a similar way to Gilbert's unveiling of the magnetic field. He also 'felt' the 'electric wind' caused by a stream of charged particles emanating from a rubbed electric. In 1729, Steven Gray observed the feature of 'charging by conduction' and, a little later,

that of 'charging by influence'. He was aware of the difference between these in that the latter, unlike the former, even for an insulated body, was present only as long as the influencing body was. Another crucial difference would, of course, be noted later. He also had the concept of electric current. Using a coarse thread as a 'wire', he was able to charge one end and record it at the other end even though the thread was of significant length. He was also aware that this charge transfer would be improved if the thread was wet and also that some materials did not transmit charge. Thus, he had identified and distinguished between 'conductors' and 'insulators'. In a letter to the Royal Society, published in 1734, the Frenchman, Charles du Fay, extended the list of electrics to essentially anything that could be rubbed (metals excluded) and also described the conduction of 'electricity' along threads. Further, he identified the repulsion of an object from the electric to which it had originally been attracted, that this repulsion from the original electric was retained until the object was 'discharged' and that it may, before discharge, be attracted to other electrics. He was the first to formulate the idea of two types of electricities that he called 'vitreous' and 'resinous' (largely dependent on the material being rubbed) and which could neutralize each other and where one (either) type could be created by removing some of the other type. He also developed the concept of 'amount' of electricity in the sense that equal amounts of each type of electricity would exactly neutralize each other. Later, of course, it was discovered that the type of electricity depended on both the rubbed and the rubbing materials. In the middle of the eighteenth century, Benjamin Franklyn recognized only one type of electricity but that it could occur in two forms which he called 'plus' or 'positive' and 'minus' or 'negative' and which corresponded to the vitreous and resinous electricity of du Fay, respectively. Here, as with du Fay, is the suggestion of the well-known law of charge conservation. Franklyn is also known for his experiments with lightning conductors. The eighteenth century also witnessed the development of superior *sources* of electricity for experimentation, such as the Leyden jar, which gave a heavy and short-lived discharge, and the voltaic pile, which gave a more continuous flow of electricity.

Perhaps the only truly quantitative results about electricity and magnetism developed in this period were the so-called Coulomb laws regarding the force between two charges or between two magnetic poles. Using a torsion balance, the French Scientist, Charles Augustin Coulomb, showed in 1785 that the magnitude of the force between two fixed charges or two fixed magnetic poles varies as the inverse square of their distance apart (including the sign in the force for the cases when the charges or poles are of the same or opposite type). The ability to measure, with accuracy, *quantity* of charge was not well developed in Coulomb's day, but later work confirmed that the force between two charges of whatever sign is, in magnitude, also proportional to the product of the quantities of the charges involved. Thus, one has the familiar Coulomb law. This result, for charges, was also known earlier to some extent to Daniel Bernoulli in *ca* 1760, John Robison in Edinburgh in 1769 and Henry Cavendish in 1771. An indirect suggestion that the inverse square law was the appropriate law of attraction was given by Joseph Priestly in 1767 with an appeal to a result in potential theory. Coulomb's law for magnetic poles was probably first given by John Michell in Cambridge in 1750.

In summary, one can say that, by the beginning of the nineteenth century, physicists (or natural philosophers as they may have preferred to have been known) were aware of magnets, electrics and some of their properties. They had elementary 'batteries' and had sent electric currents to significant distances. They knew of insulators and conductors and had reduced electric charge to two types. With Coulomb and others, they had a quantitative law about the mutual effects of charges and of magnetic poles and some concept of electric and magnetic fields. Perhaps, more importantly, they had learned (from, among others, Gilbert, Galileo and Bacon) modern experimental physics in the sense that they realized the importance of performing experiments and recording the results carefully. However, the study of electrics and magnetic materials had been largely kept separate in spite of the similarities between them. It was at the beginning of the nineteenth century that their interrelationships were discovered and that they were merged into a single discipline.

4. Electricity and magnetism

In 1820, the Danish scientist, Hans Christian Oersted, apparently inspired by the effect of a thunderstorm on a compass needle and suspecting lightning as the culprit, performed an experiment to show that, if a compass needle is positioned above or below a current-carrying wire, it attempts to set with its long axis perpendicular to the wire. This was perhaps the first demonstration of a *direct* link between electricity and magnetism, that is, between the electricity in the current and the magnetic effects on the compass needle, rather than the revealing of another similarity between electric and magnetic effects. (It is appropriate at this point to stress the importance for this, and other similar experiments, of the discovery of the voltaic pile as a provider of a *continuous* source of current.) Very soon after, the French scientists, Jean-Baptiste Biot and Félix Savart, in their attempts to find more qualitative results showed experimentally that the actual force on a magnetic pole due to the current in a long straight wire was proportional to the current in the wire, inversely proportional to the perpendicular distance from the point to the wire and acted in the direction perpendicular to the plane containing the wire and the point. They then investigated the problem theoretically and found the familiar differential form of the law. In this differential form, where the force due to a current-carrying 'element' of wire is written down, the familiar inverse square law is again encountered which then integrates out to give the above-mentioned law. Also, in 1820, André-Marie Ampère undertook further mathematical study of this phenomenon, showing how two parallel current-carrying wires attract or repel each other, depending on whether the currents are in the same or opposite direction. He also wrote down an expression for this force for more general kinds of currents. Since this force is essentially the magnetic field, Ampère's work gives a very general expression for the magnetic field generated by a current. His work was of crucial importance for what was to follow several decades later (and Maxwell was to refer to him as the 'Newton of electricity'). Thus, not only could a current-carrying wire affect a magnet, as Oersted had showed, but it could also affect another current-carrying wire. But if a current-carrying wire (i.e. charges in motion) could create a magnetic field, could a magnetic field create a current?

In 1831, in a paper read to the Royal Society, the English scientist, Michael Faraday, described a most remarkable experiment (see Faraday 1839–1855). He wound two separate insulated coils of wire onto an iron ring with one coil attached to a galvanometer and the other to a battery. When he switched the battery current on, the galvanometer needle flickered, as it did again when he switched it off. In the intervening stationary period, and although current still flowed in the second (battery) wire, the galvanometer recorded no current in the first wire. Thus, he demonstrated that a *changing* magnetic field, caused by the current *starting up or closing down* in the second wire, created a current in the first (galvanometer) wire. He also showed that the same current creation could be achieved by simply *moving* a magnet (again creating a *changing* magnetic field) inside a wire coil which was attached to a galvanometer. No such effect on the galvanometer was recorded if the magnet was nearby but stationary. Thus, a changing magnetic field was capable of creating a current in a conductor. (This phenomenon was also apparently known to the American, Joseph Henry, but he failed to publish before Faraday.) This remarkable result of Faraday was to be put to many practical uses including the electric motor and the dynamo for generating electricity.

Faraday's ideas rested not on mathematics, for Faraday had no mathematical training, but on his concept of *lines of force*. (However, Faraday's thinking was, in its nature, mathematical (geometrical) and was found to be rather useful in the later developments by the mathematical physicists.) Thus, Faraday imagined such lines emanating out from electric charges (electric lines of force) or from magnetic poles (magnetic lines of force). Indeed, such visual concepts are still used in understanding the theory of electricity and magnetism, even if the mathematics that was ultimately to describe this theory had no need of them. Faraday imagined that the space in which a magnetic field operated was in a special state, called *electrotonic* and his lines of force are particularly instructive, geometrically, in understanding, for example, the well-known patterns observed in the vicinity of a magnet (as observed by Gilbert) by covering it with a horizontal piece of paper and sprinkling iron filings on the paper. Lines of force are also useful in many applications of electromagnetic theory; for example, the current observed in Faraday's experiment can also be created by moving the galvanometer wire in the vicinity of a stationary magnet. Thus, the consequences of Faraday's experiment really depend only on the *relative* motion of the magnet and the wire. So one can *picture* the Faraday experiment by imagining that the current in the wire is created by the *changes* of the lines of magnetic force across the galvanometer wire and this picture is faithful to the fact that only *relative* velocities are involved. Thus, the experiments of Oersted, Ampère, Faraday and others display the interrelations between currents (moving charges) and magnetic fields. (However, as remarked by Einstein in his original paper on special relativity, although the effect is experimentally symmetrical, in the above sense of relative velocity, the conventional way of *describing* it is not. For when the magnet moves, one seeks an explanation for the induced current in the changing magnetic field in the vicinity of the wire whereas if the magnet is at rest one finds the reason for the current in the forces acting on the moving charges in the wire within the magnetic field created by the magnet.)

5. The period of Maxwell

In the middle of the nineteenth century, one could, albeit rather simplistically, regard research into mathematical electromagnetic theory as divided into two camps. The first of these, mostly in continental Europe and following Ampère, Weber and Neumann, were Newtonians in that they formulated theories that were based on action at a distance (as Newtonian gravity was). In Great Britain, Faraday, Kelvin and Maxwell preferred what might be called field theories in which electric and magnetic effects were either transmitted by some medium (an *ether*) or by Faraday's lines of force described earlier. The concept of an ether had been given extra impetus from the demonstration of the wave-like character of light by Huygens, Fresnel and Young. Faraday even suggested that light was a kind of undulatory motion in his lines of force and supported this view by his discovery of the so-called *Faraday effect* in which the plane of polarization of a light beam was rotated by a nearby magnetic field (or current flowing in a wire around the light beam). Thus, he found an effect linking electricity, magnetism and light.

Maxwell's first real step towards his electromagnetic theory was in the paper *On Faraday's lines of force* (in two parts in 1855–1856; Maxwell 1864). In this paper, he proceeded by analogy, as was not atypical of Maxwell, basing his techniques on those of Faraday. He introduced into space a hypothetical incompressible fluid, the flow lines of which could be taken as a geometrical representation of the electric or magnetic field, or the current flow. For example, in the case of electric fields, the *sources* of the flow lines were positive charges and the *sinks* were negative charges. There was difficulty in this geometrical approach in incorporating the existence of an interaction between the electric and magnetic fields and the current. But he was able to provide a mathematical description of several of Faraday's ideas. In his next paper *On physical lines of force* (in four parts 1861–1862), he introduced his famous 'vortices' or 'cells'. This was an attempt to incorporate the known 'rotational' properties of the magnetic field into his model. He envisaged these vortices (perhaps one can think of them as very small and spherical) occupying space (even the vacuum) and to be of small but non-zero mass, and elastic. However, their properties could change depending on the medium they occupied. These vortices were separated by particles (which one can think of as small balls) which acted as idle wheels. Thus, the idle wheels (particles), rotating without translational movement, would, from the rolling condition, gear with the vortices and enable them to rotate in the same sense. However, any translational movement of the particles would generate a differential rotation of the vortices. The rotating vortices are to be thought of as representing the magnetic field, whose magnitude is essentially the rotational speed of the vortex and whose direction is determined by the vortex's axis and sense of rotation. The movement of the balls is interpreted as electric current. With these interpretations appropriately made, Maxwell was able to derive Ampère's circuital law and to provide an explanation of Faraday's law of induction. In Maxwell's model, the balls were free to move in a conductor but not in a dielectric (an insulator). In the third part of his paper, and using his elasticity assumption, he was able to explain the familiar laws of attraction between charges. Further, he was also able to suggest that, in a dielectric medium, an electric field can slightly displace the electrical balls and that, during

the initial phase of this change, when the electric field is changing, this movement of the balls gives rise to what is essentially a transient electric current. This is Maxwell's *displacement current* and his model showed that it was intimately connected to the rate of change of the electric field. In the same part of this paper, and as will be dealt with in more detail later, he pointed out how his model suggested that changes in these electromagnetic fields would be propagated through this elastic medium at a finite, if large, speed and which he could calculate approximately. The value he obtained was rather close to the experimentally accepted value of the speed of light in the supposed 'luminiferous' ether in which light was assumed to travel. He concluded that light consisted of transverse vibrations in the (electromagnetic) ether and suggested identifying the latter with the luminiferous ether. In the fourth part of his paper, he offered an explanation for the Faraday effect mentioned earlier.

Maxwell's introduction of the displacement current modified Ampère's circuital law because now there was an extra current to be included in addition to the conduction current. Now the equations of electromagnetic theory, including the equation of charge conservation, were, prior to the introduction of this extra current by Maxwell, internally inconsistent, at least for the case of a charging capacitor in open circuit. Maxwell's inclusion of the displacement current in Ampère's law corrected this deficiency. This term plays the role, for example, of a current in the gap between the plates of a capacitor during charging.

In the key paper of this series and carrying the title *A dynamical theory of the electromagnetic field* (see Niven 1965), Maxwell attempted to remove from his theory some of the appeals to models and analogies and to lay out a coherent mathematical theory of electromagnetism. He was also able to derive from his equations an electromagnetic theory of light. He had built upon the earlier physical and mathematical developments of Ampère, Faraday, Kelvin, Weber, Green and Stokes and had derived the equations describing the behaviour of the electromagnetic interaction. Maxwell's equations, as they now stood (and in modern notation following the later simplifying work of Heaviside (who removed the potentials in Maxwell's original expressions so that the equations were written entirely in terms of the electric and magnetic fields and their sources), Hertz and Lorentz), are given by

$$\nabla \cdot \boldsymbol{D} = 4\pi\rho, \tag{5.1}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{5.2}$$

$$\nabla \times \boldsymbol{H} = (4\pi/c)\boldsymbol{j} + 1/c \, \partial\boldsymbol{D}/\partial t, \tag{5.3}$$

$$\nabla \times \boldsymbol{E} = -1/c \, \partial\boldsymbol{B}/\partial t, \tag{5.4}$$

where $\boldsymbol{D}$ is the electric (displacement) field; $\boldsymbol{E}$ is the electric field; $\boldsymbol{B}$ is the magnetic (induction) field; $\boldsymbol{H}$ is the magnetic field; $\boldsymbol{j}$ is the conduction current density vector field; ρ is the charge density; and ∇ is the usual differential operator. The quantity c will be described later. These equations would normally be augmented by the so-called constituent relations relating $\boldsymbol{E}$ and $\boldsymbol{D}$, $\boldsymbol{H}$ and $\boldsymbol{B}$ and $\boldsymbol{j}$ and $\boldsymbol{E}$ but they will not be needed here except to say that, *in vacuo*, $\boldsymbol{E}=\boldsymbol{D}$, $\boldsymbol{H}=\boldsymbol{B}$, $\boldsymbol{j}=0$ and $\rho=0$. Equation (5.1) arises from Coulomb's law for the force between electric charges, while (5.2) is the statement of the absence of free

magnetic poles. Equation (5.3) is Maxwell's corrected form of Ampère's circuital law relating the magnetic field to the current and where the conduction current $\boldsymbol{j}$ has been augmented by the Maxwell displacement current $1/4\pi\,\partial\boldsymbol{D}/\partial t$. Thus, the set of equations above was now consistent with the law of charge conservation, $\partial\rho/\partial t + \nabla\cdot\boldsymbol{j} = 0$. Equation (5.4) is Faraday's induction law. The quantity c can be calculated from the electromagnetic properties of the medium under consideration. (In this respect, the work of the German scientists, Wilhelm Weber and Carl Friedrich Gauss, must be mentioned since it was they who were the major forces in the determination of the system of units used in electromagnetic theory and which is intimately related to this problem.) *In vacuo*, where $\boldsymbol{E} = \boldsymbol{D}$ and $\boldsymbol{H} = \boldsymbol{B}$, c turned out to be a constant rather close to the value of the speed of light accepted at the time. In this case, a rather well-known and straightforward manipulation of Maxwell's equations, using some standard vector analysis, then shows that $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfy the relations

$$\nabla^2\boldsymbol{E} = 1/c^2\partial^2\boldsymbol{E}/\partial t^2, \qquad \nabla^2\boldsymbol{B} = 1/c^2\partial^2\boldsymbol{B}/\partial t^2, \tag{5.5}$$

where ∇^2 is the usual Laplacian operator. Thus, the electric and magnetic fields satisfy the familiar wave equation with propagation speed c and which is now interpreted as the speed of light. This was the main supporting evidence that light was an electromagnetic phenomenon. In this sense, Maxwell had succeeded in theoretically unifying electricity, magnetism and light and pressed his claim for the need of only one ether with respect to which c was the speed of light. Maxwell gathered together his work and the work of others into his great mathematical work *A treatise on electricity and magnetism* (1873) and which was mostly written on his estate at Glenlair. In a series of experiments in 1887–1888, sadly 9 years after the death of Maxwell, the German physicist, Heinrich Hertz, experimentally detected Maxwell's electromagnetic waves (essentially radio waves) and confirmed their speed to be of the same order of magnitude as the speed of light. (In fact the English physicist, Hughes, had similarly detected such waves some 7 years earlier than Hertz but became discouraged when it was suggested that his results may be explained by electromagnetic induction, and Hertz published first.) The celebrated experiment of Hertz recorded results that were consistent with a Maxwell-type theory in which effects were propagated with a finite velocity and which were not consistent with 'action at a distance' theories, in which effects were propagated instantaneously. It may be said that, from the date of Hertz's success, Maxwell's theory was essentially fully accepted.

6. From Maxwell to Einstein

Towards the end of the nineteenth century, theoretical physics was dominated by the mechanics of Newton and the electromagnetic theory of Maxwell. In Newton's theory, one has the concept of absolute time and a system of preferred observers, the *inertial* observers. Such inertial observers, or inertial frames of reference, arise from the fact that Newton's theory is based on the concept of force and essentially claims to be able to distinguish between what might be called 'real' forces and 'inertial' (also called 'accelerative') forces. Then Newton's second law, which equates the real force on a particle to the product of its mass and its acceleration, holds only in such privileged inertial frames. Each inertial

observer is, then, a representative of Newton's absolute space and any non-inertial observer will experience accelerative forces as a kind of reaction on him due to his acceleration with respect to (any of) these absolute spaces. The inertial observers, regarded in the usual way as coordinate systems, are connected, mathematically, by the so-called *Galilean* transformations and Newton's relativity principle is the statement that these inertial observers are indistinguishable one from another by mechanical experiments (Galilean invariance). However, Maxwell's theory now introduces an ether and the crucial constant c, the speed of light in the rest frame of the ether. Thus, if one assumes the existence of an inertial observer who is at rest with respect to this ether, this observer measures the speed of light, independently of the latter's direction, to be precisely this number c. However, other inertial observers in relative motion with respect to this ether observer, will clearly not record such a convenient measurement owing to the addition laws for velocity derived from the Galilean transformations in Newtonian mechanics. Thus, an electromagnetic experiment could distinguish one inertial frame from another and the Galilean invariance of Maxwell's theory is lost.

In this period, one of the first duties was to find the relative speed v of Earth and ether and the most celebrated experiment in this direction was performed, following Maxwell's suggestion, by Michelson and Morley in 1887 (following an earlier flawed one by Michelson in 1881). In this experiment, the quantity being measured, that is the above relative speed, occurred in the equations in the combination $(v/c)^2$ and was thus a 'second-order' effect. It was considered impossible at that time to measure such small quantities directly, but Michelson's interferometer, taking advantage of wave interference, was capable of the task. The conclusion was that, assuming the prevailing Newtonian theory and Galilean transformations, no motion of the Earth through the ether was detected. Other similar experiments were also performed but with the same general result. Many explanations were forthcoming to explain these so-called 'null' results, of which perhaps the most famous were the *contraction hypotheses* of Fitzgerald and Lorentz. These hypotheses actually envisaged that a rigid rod moving with respect to the ether would suffer a physical shrinking.

The history of special relativity (for more details, see Whittaker 1951) may be taken to begin with Poincaré in 1899 when he claimed that optical experiments really depended only on the relative velocity of the bodies in question and that absolute motion in this sense was *undetectable*. Thus, Poincaré in 1904 and Einstein in 1905 announced the *principle of relativity* stating that *no* experiment could distinguish one inertial frame from another and hence that *all* physical laws should reflect this *invariance*. Thus, it extended the Newtonian principle of relativity *from mechanical to all experiments*. The true understanding of what is now called special relativity originates with Einstein's paper of 1905. Although it is not entirely clear whether Einstein was aware of the Michelson–Morley experiment at the time, it appears that he probably was. In any case, Einstein had no need of an ether and accordingly rejected it. He argued, roughly speaking, that one should keep the inertial frames of Newtonian theory but should include, in addition, this much stronger principle of relativity, together with the assumption that each inertial observer would measure exactly the same speed of light, independently of its direction with respect to that observer. Clearly, one cannot retain the Galilean transformations, since they are obviously inconsistent with

such an assumption. Einstein then proceeded to calculate the (consistent) transformations between inertial frames needed to replace these Galilean transformations and gave the first real physical derivation of the (Lorentz) transformations (which had, in fact, been used earlier by Voigt and Lorentz and named after Lorentz by Poincaré). It followed from this that Newton's absolute time had to be abandoned as inconsistent with these transformations (but absolute space was essentially as in Newton's theory). Einstein was also able to show that the contraction hypotheses of Fitzgerald and Lorentz arose naturally in his theory and were of a kinematic rather than a physical nature. He further showed that the preservation of time by the Galilean transformations in Newtonian theory was replaced, in his theory, by an expression describing the 'relative' nature of time for inertial observers. Following an earlier and incomplete attempt by Lorentz (of which Einstein was apparently unaware), Einstein was able to use the assumption that Maxwell's equations are invariant in form under the Lorentz transformations (as required by the principle of relativity) to find the relations between the electric and magnetic fields in two inertial frames. It turned out that the electric field for one inertial observer is a particular combination of the electric and magnetic fields of the other observer (the precise nature of which depends on the kinematical relationship between the two observers) and similarly for the magnetic field. It follows that the electric and magnetic fields do not, in this sense, have a separate existence but rather are observer-dependent manifestations of a single *electromagnetic field*. Einstein, in fact, in his 1905 paper referred to the strange asymmetry in the *physical description* of the important experiment of Faraday (as mentioned in §4) as being at variance with the experimental symmetry. This problem arises as a consequence of its being described by a single observer in terms of his electric and magnetic fields and the dependence of these fields on the observer. Thus, Maxwell's work, together with Einstein's special relativity theory, constituted a unification of the electric and magnetic fields into a more meaningful electromagnetic field. Of course Maxwell's equations, as written above, still refer to the electric and magnetic fields separately. The complete mathematical unification of these fields came a little later, from the beautiful work of Hermann Minkowski, when he introduced the *Maxwell tensor* (or *Maxwell bivector*) into the theory in 1908. This mathematical object has six degrees of freedom, representing the three components of each of the electric and magnetic fields and which, owing to its *tensor* nature, could be regarded as being, in a well-defined sense, independent of the observer. However, a particular observer may still retrieve his personal electric and magnetic fields from it by a simple procedure. Maxwell's equations may then be rewritten in a simple and elegant way as differential equations involving this tensor.

Maxwell's electromagnetic theory and Einstein's special theory of relativity have withstood the test of time and stand as remarkable tributes to these scientists. The theory of electricity and magnetism is one of the few areas of physical enquiry that did not attract Newton's attention. It differed in many ways from Newtonian philosophy. For example, the magnetic field due to a current element which is experienced at a certain point does not act along the straight line joining the element and point but rather at right angles to this line. Thus, whereas classical ideas, embodied in the law of inertia, lead to the identification of *straight* lines in the (assumed) Euclidean structure of space, electromagnetic theory identifies *right angles* in this structure.

Maxwell will be remembered for his many contributions to theoretical physics and his work on electricity and magnetism and the consequent Maxwell equations are probably the most important addition to theoretical science in the nineteenth century.

References

Gilbert, W. 1600 De Magnete (English transl. by P. F. Mottelay, published in 1893). London.

Faraday, M. 1839–1855 *Experimental researches in electricity* (vol. I, eds R. Taylor & W. Francis, 1839; vol. II, ed. R. Taylor, 1844; vol. III, eds R. Taylor & W. Francis, 1855). London, UK: R. & J. E. Taylor.

Maxwell, J. C. 1861–1862 On physical lines of force. *Philos. Mag.* **21** (parts I and II) and **23** (parts III and IV).

Maxwell, J. C. 1864 On Faraday's lines of force. *Trans. Camb. Philos. Soc.* **10**, 27–83. (Part I read 1855, part II read 1856.)

Maxwell, J. C. 1865 A dynamical theory of the electromagnetic field. *Phil. Trans. R. Soc. Lond.* **155**, 459–512. (doi:10.1098/rstl.1865.0008)

Maxwell, J. C. 1873 *A treatise on electricity and magnetism*, vol. 2. Oxford, UK: Clarendon Press.

Niven, W. D. (ed.) 1965 *The scientific papers of James Clerk Maxwell*. New York, NY: Dover Publications, Inc.

Whittaker, E. T. 1951 *A history of the theories of aether and electricity*, vol. 2. London, UK: Thomas Nelson and Sons Ltd.