The Fundamentals of Electromagnetic Theory Revisited

John W. Arthur

School of Engineering and Electronics, University of Edinburgh
King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, Scotland
Tel: +44 (0) 131 650 5565; E-mail: john.arthur@ed.ac.uk

Abstract

This tutorial paper deals with several aspects of basic electromagnetic theory that appear to be insufficiently appreciated a century and a half after Maxwell published his well known equations and over half a century since the MKS / SI system of units was introduced. New concepts have not been completely embraced, while older artifacts and anachronisms have lingered on. The main issues include problems stemming from the original theory of magnetism, confusion between key aspects of the fields **B** and **H** and the somewhat puzzling equivalences between characteristically different mathematical models based on poles or currents. While the answers to most of these questions are somewhere or other in the literature, they are often difficult to find and there seems to be a lack of a consistent approach to the fundamentals. This article surveys the problem areas, explores the issues involved and attempts to provide clear answers and understanding through reasoning and commentary. Only simple mathematics has been used and the treatment has been kept strictly in terms of the field quantities. Results and detailed references are given in key areas, and the history of the subject is touched on where relevant.

Keywords: Electrical engineering education; electric fields; electrical forces; electrodynamics; electromagnetic fields; electromagnetic forces; electromagnetic propagation; electromagnetic theory; magnetic fields; magnetic forces; magnetic poles; Maxwell equations; physics; measurement units; relativistic effects.

The full citation for this article is:
Arthur, J.W. "The fundamentals of electromagnetic theory revisited", Antennas and Propagation Magazine, IEEE, Volume 50, Issue 1, Feb. 2008, Pages:19 - 65
Digital Object Identifier 10.1109/MAP.2008.4494503

- page numbers refer as closely as practicable to those of the published article -

1 Introduction

This tutorial paper deals with several aspects of basic electromagnetic theory that appear to be insufficiently appreciated a century and a half after Maxwell published his well known equations and over half a century since the MKS[1] system of units was introduced. New concepts have not been completely embraced, while older artifacts and anachronisms have lingered on.

The main issues are:

- The misleading terminology still in use for the magnetic field quantities
- Definition of magnetic force field in terms of poles
- Incompatibility between the poles defined in SI and emu/Gaussian systems
- The inclusion of magnetic poles within the elementary concepts
- Which of **H** and **B**, if either, is the fundamental force field?
- The casual interchange of quantities such as **H**, **B**, μ**H** and μ_0**H**
- Which of **H**=μ^{-1}**B** and **B**=μ**H** is more consistent with **D**=ε**E**?
- The ambiguous vector character of **H**
- Why **H** often appears to take the role of a fundamental force field
- The correct form and interpretation of the Lorentz force
- The essential *microscopic* and *macroscopic* forms of Maxwell's equations
- The mathematical equivalences and differences between solutions based on poles and circulating currents
- Resolution of the apparent incompatibility between field solutions based on curl and divergence
- Reluctance to employ even simple results from special relativity in establishing the foundations
- Lack of appreciation that magnetism itself is *prima facie* evidence for special relativity
- The need to teach for understanding versus applications and problem solving

- The benefits of treating electricity and magnetism based on a common footing in Coulomb's law.

While the answers to most of these questions are somewhere or other in the literature, they are nevertheless difficult to find with any certainty. Treatments vary, and the emphasis may often be on mathematical technique and applications rather than understanding. Although modern texts tend to take a correct approach, within the literature as a whole a variety of the legacy issues remain. On the other hand, no doubt wishing to put them entirely aside, modern works generally make scant reference to these problems and so, having stumbled over one, it is often difficult to find a ready answer. In addition, there still seems to be scope for clarifying the basic framework of electromagnetic theory, a complex and often mathematically difficult subject which nevertheless has truly simple fundamentals.

---

[1] Now the Systeme International, referred to as SI or the SI system. Where we refer to MKS or MKSA, it is in the historic context, *ca* 1935-1960.

The objective here, therefore, has been to present reasoned answers to the questions above, and to give comment on the issues involved for the benefit of anyone who has an interest in electromagnetics. While the comment is intended to be thought provoking, in places it may appear to be just provoking. However, it is appropriate from time to time to challenge things that are seen or done in a particular way, be it for historical reasons, or because it does not suit a particular view, or simply because a better way cannot be agreed upon.

Only simple mathematics has been used, although a basic appreciation of vector analysis and special relativity is unavoidable. Moreover, the treatment has been kept strictly in terms of the field quantities so as to avoid introducing other notions, such as potentials, that may well be very useful in their own right but simply add another layer to the concepts and techniques that require to be embraced. Once the fundamentals are established, such extensions may be safely introduced, which most textbooks do accomplish very adequately. In spite of this tutorial approach, every effort has been made to state key results precisely and to quote key references wherever possible. The latter task has not been an easy one since many results that are commonly accepted today are often not in their original form, and many sources give them little or no justification, as we shall see. While the history of the subject is very relevant to these issues, within the scope of this article it is only possible to touch on those aspects of it that are particularly relevant. The reader may consult the work of Whittaker [1] for further information.

This introduction provides the background and motivation for the article, Section 2 explores the legacy issues, while 3 attempts to uncover what may be considered to be the true fundamentals of the subject. A review of Maxwell's equations is undertaken in Section 4, examining the free-space form and various macroscopic forms for contrast, including versions which retain magnetic poles. Section 5 covers some basic results from special relativity – transformation of the electric field, deduction of Maxwell's equations from Coulomb's law and the invariance of the speed of light *in vacuo*. Section 6 presents further discussion, in part drawing some of the topics together and in part introducing some further material where comment seems necessary in light of issues that have been raised along the way. After a brief conclusion there are two appendices. The first summarizes the essential electromagnetic equations as a useful basis, and the second discusses some practical examples from electron spin resonance, the Hall effect, and the force on a current carrying conductor in different magnetic scenarios.

Several reference works are cited frequently and therefore wherever possible specific page numbers have been given for each instance to make them easier to find. This is fairly essential as in some instances the relevant information turns on the use of a single term or symbol.

1.1 Historical Background

James Clerk Maxwell effectively established classical electromagnetic theory as a known science by completing its mathematical description [2; 3]. Originally, however, there were two apparently unconnected theories for electricity and magnetism which, in common with gravity, were held to obey an inverse square law. The only distinguishing feature was that the force originated between masses, electric charges or magnetic poles, as appropriate. At the time there was no more idea of a connection between electricity and magnetism than, say, for electricity and gravity. Following the observation that electric currents could produce magnetic fields, as reported by Oersted [4], Ampere [5] and others [6, pp. 91-92], the pursuit of a single integrated theory commenced. If it can be said that Maxwell succeeded in taking the final step in this quest, it cannot be said that the full implications of his unified theory did a great deal to put aside the original theory of magnetism which, for the purposes of magnetostatics at least, persisted.

The premise that all magnetic fields originate from circulating currents alone (or their quantum-mechanical equivalent) has existed since the original work of Ampere [6, pp. 97], but in spite of the famous and oft quoted conclusion from Maxwell's second equation, it has never been conclusively proved that magnetic poles do not exist. It is, however universally, accepted. Occasional conjecture about the possible existence of free poles persists within the field of elementary particles, or special states of matter, but nothing corresponding to the original concept has ever been found [7, p.905].

The non-existence of poles seemed not to deter the proponents of the original magnetic theory. Perhaps this was in full knowledge of the facts, on the basis that the useful working model it provided was too good to be cast aside. Furthermore, it nicely parallels Coulomb's law and the mathematics involved is simpler than for the interaction between currents. The pole description and the old electromagnetic units were consequently allowed to survive. Even in the middle of the last century, the effort to introduce the MKS and MKSA systems of units [8; 9 pp. 16-18] did not displace the pole; it merely redefined it in terms of forces between currents. Today the old electromagnetic units, emu, that are based on pole theory still prevail and the MKS definition of the magnetic pole, while sound enough, merely gives undue credence to the pole concept. There appears to be little awareness that these magnetic poles are quite contrary to the original poles in key respects [9 pp. 241-242; 10, Vol. 1, pp. 179-181], and they have even been referred to as 'induced poles' in order to accommodate this [11, pp. 5-6].

In addition to the undesirable legacy of this process of unifying both of the theories without effectively disposing of the outmoded parts, the nature of the magnetic force, being a higher order effect, is more complex than that of the inverse square law that suffices for both electric charges and magnetic poles. Conceptually we have come to expect symmetry in the electromagnetic equations with **H** and **B** being on a parallel with **E** and **D** respectively [12]. If we did have poles, this would indeed be the case, but it is not so. This issue, together with the 'pole legacy', leaves us with the different systems of units we now have and a persisting misunderstanding - or at least lack of clarity, concerning the magnetic field

quantities **B** and **H**. These sort of issues, some obvious and others subtle, have been carried on through the years in definitions, articles, textbooks, and in much conventional teaching on the subject. While there are numerous *prima facie* examples, there are many more where casual usage is simply at fault. There appears to be a significant remanent tolerance, even sympathy, for the pole description. For example, the magnetic field intensity as defined still refers to **H**, which is often regarded as the primary field, rather than **B**. In addition, **H** often crops up in expressions where we might have expected to find **B,** and we still think and write $\mathbf{B} = \mu_0(\mathbf{H}+\mathbf{M})$ rather than $\mathbf{H} = \mu_0^{-1}\mathbf{B} - \mathbf{M}$ On the other hand, there seems to be no similar problem with the use of **D** and **E** and we all recognize **E** as the primary field – the force field.

Putting aside any debate on systems of units, in order not to confuse matters by having two different sets of equations, a single system of units has been used throughout this article - SI.

1.2 What is Fundamental?

There are a vast number of textbooks dedicated to electromagnetic theory and, being aimed for the most part at standard undergraduate courses, these tend to treat the fundamentals as a formality and then move on quickly to applications and the build up to Maxwell's equations and its applications, such as propagating waves. For example, Coulomb's law for the forces between point charges within a continuous medium of dielectric constant ε may be simply stated in the same form as it applies in a vacuum with the constant ε_o replaced by ε :

In free space
$$F_{12} = \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}^2} \tag{1a}$$

In a material of dielectric permittivity ε
$$F_{12} = \frac{q_1 q_2}{4\pi\varepsilon \, r_{12}^2} \tag{1b}$$

These equations are often given as a pair of definitions [13, pp. 39 and 44], but of the two, only Equation (1a) is fundamental while Equation (1b) may be derived given that it follows from Equation (1a) as a direct result of a linear electric polarization induced on the constituent molecules of the medium. In this context, the term molecule includes atoms, ions *etc*., a meaning that we shall adhere to throughout this article.

We must be careful, therefore, to distinguish what is truly fundamental from what is simply consequential. As a definition, Equation (1b) is valid only as postulate within a phenomenological formulation of electrostatics, that is to say, a description of electrostatics in which we select a set of model equations which appears to describe the observed phenomena without further enquiry as to whether any of these equations can be derived from other principles. We can also use the terms *microscopic* and *macroscopic* to differentiate between Equations (1a) and (1b). Equation (1a) is termed microscopic, as it applies *individually* to all charged particles treated as being in free space, while Equation (1b) is termed macroscopic, as it involves *aggregates* of particles treated as a body, in other words what we refer to as a medium. The scientific fundamentals are based in the microscopic theory from which the macroscopic theory must be later demonstrated in a fully consistent manner.

In principle, therefore, Equation (1b) can be derived directly from Equation (1a) by means of a completely microscopic approach based on the concept of molecular polarisability. The Lorentz-Lorenz description of dielectric media [14, pp. 89-95; 15 pp. 84-87, 100-104; 16, pp. 150-158] takes us as far as relating the dielectric permittivity ε to the molecular polarisability α, which is a measure of the extent to which the charge that is bound within the molecule is displaced, or polarized, by the presence of a Coulomb force field. The added step is the calculation of the net free-space Coulomb force Equation (1b) between the two given test charges *together with all of the polarizable molecules* *affected by their presence*. This represents, both in the physics and mathematics involved, a far from trivial problem. No doubt it has been tackled at some point, but if so, it is now obscure.

Such a head-on approach is not essential. The problem can be tackled in stages, following the development of other concepts such as the electric field, **E**, macroscopic polarization, **P**, and the displacement, **D**. But in many textbooks we would find that Equation (1b) is left to be inferred. At some point we would be able to state that, in the presence of a medium of dielectric constant ε, we can simply write down **D** for the single free charge q_1, then since $\mathbf{E}=\mathbf{D}/\varepsilon$, the force on charge q_2 must be reduced by a factor of $\varepsilon/\varepsilon_0$ with respect to the force in free space. Happily, this does agree with observation. Just as simply, we may infer the replacement of ε_0 by ε, say from the energy stored in a charged parallel plate capacitor. But all the same, these methods are a long way around a direct proof that the general form of Coulomb's law in a dielectric medium reduces to Equation (1b). Rather than actually prove it, we accept Equation (1b) as given based on its apparent self-consistency with other results that fit together to make up the theory. And so it appears to be, as with the development of any subject, that key results such as these are contributed in no particular order from a variety of sources and over a considerable period of time. For every key result a significantly greater amount of ancillary information is generated, filling in the gaps and solving specific applications. Out of this process, however, it is useful to keep firmly in mind the bare fundamentals upon which all else depends, *e.g*. in this case Equation (1a) rather than Equation (1b).

1.3 Difficulties with Phenomenological Models

That a phenomenological description appears to be correct is no guarantee that this is indeed the case. It may behave well as a model, it may be numerically consistent with observed results, but it may also provide an erroneous physical description. And so, in a case where we have two or more different phenomenological models, where by different we mean more than simply having different mathematical representations, it is reasonable to think that only one of them is correct. It is even a frequent outcome that none of the available models may be correct, but unless there are very special circumstances, it is extremely unlikely that two substantially different models are simultaneously correct. Here, by special circumstances we mean something akin to the wave-particle duality in physics where the realization of an equivalence between two characteristically different models represents a major conceptual breakthrough in itself. The more usual situation is the search for and discovery of evidence in an attempt to confirm a preferred model and eliminate the others, and as often as not, different people will have different preferred models! And so here we are concerned with the correct physical model for the magnetic field, and not with which mathematical model is to be preferred.

1.4 Understanding versus Problem Solving

Understandably, since even the modern theory is now quite old, the bulk of textbooks and papers published over the last half a century deal with few fundamentals. Generally they give us the necessary formulation of the basics for the main tasks in hand, for example numerous applications of the theory and the solution of specific problems. Because of the sheer number of these works and their time span, there is a lack of any single seminal work that both fully and accurately encapsulates it all, though Elliott [17] is to be commended. Moreover, we are moving into an age where there seems to be less interest in doing the ‘hard’ mathematics that fills many an advanced textbook on electrodynamics. While the number of people capable of carrying out a *tour de force* in 3-d vector analysis using general orthogonal co-ordinate systems is becoming quite few, on the other hand we now have advanced computer software to do the job for us. But the software cannot do the *understanding* for us. At one time we may have confused understanding with mastery of technical detail, but in this new age we do have an opportunity to redress the balance. Indeed, if we are going to progress by taking advantage of advanced software to solve ever more complex applications, the essentials of a course in electromagnetics will lie more in the understanding of the subject and developing a firm mastery of the fundamentals. The importance of this is, perhaps accidentally, underlined in a technical note accompanying a well-known electromagnetic software package which declared:

“Warning: The Lorentz force does not compute the correct force on objects…where $\mu \neq \mu_0$"

Perhaps this was really intended to mean that it was the software package that was lacking in this respect, not Lorentz. But just how will that statement be understood by someone who does not have a firm grip on the fundamentals? Without the understanding, how can the software be properly applied?

With a mature subject such as electromagnetic theory, there rarely is the opportunity to reorganize the fundamentals and to recapitulate the key results in order to provide a consistent structure that has been built up from the foundations, as happened in the 20th century with mathematics. In this article, however, we can attempt to revisit a few of the fundamentals in order to illustrate just how important a clear understanding of them is. Our approach has been to stay close to the fundamental concepts and to minimize the use of ancillary concepts and terminology (even where these may be useful in their own right), to use examples to illustrate a point rather than resorting to mathematics, and to explain rather than to prove. In particular, we shall attempt to identify and clear up what has become a small legacy of conceptual pitfalls and anachronisms.

1.5 Fundamental but Neglected

There exists a rarely mentioned proof of something essential to our understanding of electromagnetic wave propagation in all matter. Most of us will take this for granted, fundamental though it may be. Fortunately, the proof in question is still accessible. It is the Ewald-Oseen extinction theorem [15, pp. 84-87] that underpins the theory of propagation of electromagnetic waves in real matter, as opposed to theoretical dielectrics. Again, the issue is the difference between the usual phenomenological approach and a more fundamental one. The simplest approach based on Maxwell’s equations provides a phenomenological treatment of the subject which obviously works in that it fits the observations: it shows that waves in isotropic real matter travel as if *in vacuo*, but with the speed of light modified to be $1/\sqrt{\varepsilon\mu}$ rather than $1/\sqrt{\varepsilon_0\mu_0}$. To most of us, this is so basic what more can there be to say? But is it not the case that all we have done is simply to replace ε_0 and μ_0 by ε and μ, just as in going from Equation (1a) to (1b) ? The model itself guarantees the result simply because the two parameters involved, ε and μ, are arbitrary. The weak point is really to do with the usual pair of assumptions, $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. Are they really applicable to what is going on *real* matter, as opposed to a hypothetical macroscopic medium?

Pause for a moment, however, to think of the microscopic picture. For example, a gas at normal ambient temperature and pressure is only about 0.1% filled with matter. It could be argued that if any electromagnetic wave is present it must be propagating in something closer to free space rather than a dense array of molecules that we would intuitively treat as a continuum. Where gas molecules are present, they interact locally with the electromagnetic wave and so scatter some of it. Each molecule sees the incident wave, plus the scattered waves due to all the other molecules, which in this microscopic picture both travel in the intervening space at the speed of light *in vacuo*. How do we deal with this and demonstrate that it still agrees with the phenomenological description as a continuous dielectric medium that we take for granted?

Within any medium, be it solid or liquid as well as gas, the microscopic picture is that every polarizable molecule scatters some of the driving '*in vacuo*' wave as a direct result of the time-dependent polarization that it develops under the wave's influence. Taken as a collection of individual radiators, how does this molecular scattering process come together so that the wave propagates in the one well-defined direction with just the expected velocity? This is not a trivial problem to analyze by any means, but Oseen [18] and Ewald [19] solved it, for isotropic and crystalline media respectively, some 60 years after Maxwell's development of the phenomenological theory that it supports. Effectively they showed that the molecular scattering process generates *two* new waves within the medium, the expected wave, propagating with velocity $1/\sqrt{\varepsilon\mu}$, and yet *another* wave propagating with velocity $1/\sqrt{\varepsilon_0\mu_0}$. This latter wave is everywhere equal in direction and amplitude to the incident or 'driving' wave but exactly 180^{o} out of phase. It therefore completely extinguishes the driving wave, allowing it to be replaced by the refracted wave alone, and hence the name of the theorem. This is an intriguing picture, which also happens to account completely for the processes of reflection and refraction. The theory applies equally to solids and liquids but, importantly, for gases in particular it explains how they can behave electromagnetically just as if they were a continuum.

When we find ourselves in some kind of conceptual quandary over assumptions that we have taken for granted, it is to fundamental proofs such as these that we must resort in order to clarify matters. In addition, there is the obligation to be perfectly clear on the fundamentals of any subject that we make use of in our scientific writing and teaching. The works of Lorenz [20] and Lorentz [21] and in particular Oseen and Ewald may have become neglected because they are detailed and difficult on the one hand yet on the other they only seem to confirm more familiar notions that we seem to grasp intuitively. This is unfortunate since they should be well remembered as providing an essential basis for what we do so much take for granted. They underpin the macroscopic theory, in which matter is

some kind of conceptual continuum, with a more fundamental microscopic one. The fact that theories such as these are 'difficult' should not prevent us from giving them their proper place and citing them in support of simpler ideas. If we should altogether forget these essential underlying theories, at some point we will go astray with the theories with which we are more familiar.

1.6 The Main Issue

As previously stated, the problem that we still encounter in electromagnetic theory is that magnetism has two descriptions, one based on poles and the other on circulating currents. Unlike some other developments in our understanding of the physical world, say the arrival of theory of special relativity, this is not just a question of the new theory extending the old by adding a new layer, as it were. Pre-relativity theory is still valid, both practically and conceptually, for everyday purposes. Nor is it like the wave-particle duality, where we have two apparently different theories which have been found to stem from a common root within quantum mechanics. Here we have two different ideas which produce similar results but which have certain irrevocable incompatibilities. It is all the more fascinating because they are mathematically quite different, yet seem to produce identical results in most respects *but not all*. The newer theory has not fully supplanted the old, and having both theories side by side for the best part of two centuries has led to a number of areas of confusion. These, and the reasons behind them, are the main thrust of this article, as we shall now examine.

2 The Legacy

In this section, we examine the specific issues that have arisen from the historical legacy, including a mixture of nomenclature, conceptual problems and casual usage.

2.1 Referral to H Rather than B in Key Equations

Within a magnetic field, the equations for the torque $\mathbf{\Gamma}$ acting on a magnetic dipole $\mathbf{m}$ [22, p. 150] and the Lorentz force $\mathbf{F}$ acting on a point charge q [22, p. 191] moving with velocity $\mathbf{v}$ are given as

$$\mathbf{\Gamma} = \mathbf{m} \times \mathbf{B} \qquad (2)$$

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad (3)$$

where $\mathbf{B}$ is the so-called magnetic flux density or induction. Misunderstandings over which field should be present in these equations, $\mathbf{B}$, or the magnetic field intensity, $\mathbf{H}$, frequently arise in published works [23, pp. 262-263; 24, p4.11; 14, pp. 155 and 241; 25, p. 503; 26-30; 31, p. 9.3]. In these examples, one sees $\mathbf{H}$ used in place of $\mathbf{B}$ with no apparent explanation. Some of the reasons for the use of these apparently erroneous forms involving $\mathbf{H}$ rather than $\mathbf{B}$ are discussed below.

2.2 Expediency

One of the common reasons for interchanging $\mathbf{B}$ and $\mathbf{H}$ is simply that a particular equation of interest may involve $\mathbf{B}$, say, as a variable. To apply the equation, we need to find $\mathbf{B}$, but instead have an expression for $\mathbf{H}$ to hand. Rather than explicitly converting from $\mathbf{H}$ to $\mathbf{B}$ as an intermediate step, from the outset the author simply writes down the original equation in terms of $\mu\mathbf{H}$ rather than $\mathbf{B}$. And so we generate an equation such as $\mathbf{\Gamma} = \mathbf{m} \times \mathbf{H}$ that can only be correct within a given context (although this often goes without mention).

2.3 Semantics

Is Equation (3) effectively a definition of the field **B**, or is it an equality that is valid only in specific circumstances? Therefore, does **B** here really mean **B** under all circumstances, or is it just a convenient way of writing μ**H**, or even μ_0**H** *in vacuo* ? This is a common enough issue when dealing with formulas in the literature. The context, and the usage that was customary in a particular discipline *at the time of writing*, often need to be known in order to be certain as to the meaning. In the present case, we may take Equations (2) and (3) as defining equations, whereas all the forms involving **H** can only be correct in a given context.

This assertion *effectively defines* **B** *as the fundamental field giving rise to the magnetic force*, and not **H** as originally assumed by the earlier pole theorists. This, therefore, represents the watershed between the modern and historical views of electromagnetic theory. To some, the difference between **B** and μ**H** might seem to be hair-splitting, but not so. Remember that equating the two is only a special case that applies when a linear relationship may be assumed. It certainly does not apply in the presence of permanent magnetization, in which case **B** and μ**H** may even be more or less in *opposite directions*, as shown in Figures 8b and 8c, later in the article.

2.4 Casual Usage

Problems sometimes arise out of a familiarity with the subject which causes inconvenient detail to be glossed-over, or from taking an over-casual approach with concepts and nomenclature. Provided the treatment has been self consistent, the end results in such circumstances are often still factually correct, given that the author will have taken the trouble to confirm them before publication. Nevertheless, since understanding is likely to be impaired and intermediate steps may be technically wrong, or at least open to doubt, casual usage is not to be condoned.

2.5 Unhelpful Nomenclature

Notwithstanding the modern view that poles do not exist, **H** is still termed the 'magnetic field' or 'magnetic field intensity' while **B** is referred to as the 'magnetic flux density' or 'magnetic induction'. Even prior to 1890, Heaviside expressed doubts about the use of the term induction [10, Vol. 2, p27], while even now it may be read as describing induced rather than primary effects. In free space this choice of **B** or **H** makes little difference, effectively only a difference of units in the SI system, but nevertheless it gives rise to confusion because it retains the suggestion that **H**, rather than **B**, is the primary field associated with the magnetic force. Kraus [32], refers to both **B** and **H** alike as the magnetic field, perhaps in a genuine attempt to avoid the issue.

In spite of the concept of poles as such being obsolete, we still use the term in order to identify the sense of magnetization, *i.e.* north and south poles. Similarly, even though we know that an actual current may be due to a flow of negative electrons, we still retain the concept of a conventional positive current flow in the opposite direction. However, without referring to poles *per se*, we can still choose to define the north pole of a current dipole as being the face from which the direction of **B** emanates and the south pole as being the face on which it enters [33], with the right-hand screw rule relating the direction of **B** to the direction of circulation of the current. It is interesting to note that without the benefit of magnetism based on currents, we could only distinguish the north and south poles of a polar dipole by means of a reference dipole, and in fact the earth itself provided the original reference dipole as well as the terminology of north and south poles. The root 'pole' is now embodied in terms such as dipole, polar *etc*. It derives from *polus*, meaning the end of an axis, and in the context of a magnet the reference to the poles of a magnet is quite valid if we mean the where its axis of magnetization cuts the surface. It would make no sense, therefore, to try to eliminate or replace the word pole or the terms north and south pole as means of identifying the ends of either a magnet or an elementary dipole.

2.6 Confusion between the Roles of B and H

Putting nomenclature, semantics, customs and casual usage aside, a good deal of confusion about the respective roles and characters of **B** and **H** is evident. Many authors have in the past drawn analogies between **D** and **B** on the one hand and between **E** and **H** on the other [24, p. 4.11; 7 p. 496; 34]. The modern view, however, is that, if anything, **B** parallels **E** while **H** parallels **D** [22, p. 153; 9 pp. 12 and 242]. This assertion, however, requires careful qualification which we will address in Section 2.9 below.

2.7 Does H Apply to Magnets While B Applies to Currents?

Even a theoretically correct discussion on the equivalence of magnets and currents, as given by Kitaigorodsky [35], poses questions. Is it the case that **H** applies to magnets while **B** applies to currents? Indeed, we may suppose that such a view was held for some time, but, as we shall see later in Section 6.2, it cannot be valid – indeed if it were there would be some surprising results.

2.8. Different Systems of Units

2.8.1. Interchange of B and H in the Gaussian and emu systems

The different systems of units often take the blame for problems, if only for the fact that the most misleading examples tend to occur when the units of **B** and **H** are dimensionally identical (*i.e*. where μ_0 is dimensionless). It is of little help, and perhaps only more confusing, when the units are identical but have been given different names, such as the oersted and gauss. The numerical equality of **B** and **H** can make the casual practice of interchanging them all too easy. Doing this to simplify a calculation is one thing but, as previously discussed, replacing the one with the other in a defining formula is quite another. Not only can the occurrence of such 'context sensitive' formulas be a source of annoyance to those unfamiliar with the practice, it can certainly give rise to genuine misunderstanding.

2.8.2.Definitions Based on Poles

Undoubtedly, one of the main problem issues is the continuing use of the old terminology and definitions [13, pp. 38-41] that is associated with the emu. In spite of the fact the use of these is deprecated by the major technical bodies such as the IEC and IEEE, they still persist mainly through their use within the Gaussian formulation of the electromagnetic equations. For example, the oersted, the unit of magnetic field intensity, may still be found defined in terms on the force on a unit magnetic pole [13, pp. 38-41, 36].. It could be argued that such a unit pole can still be defined as a notional thing, but as discussed in Section 2.8.4 below, this type of unit pole cannot be consistent with a unit pole defined from forces between currents, as would now be required. There are similar problems with the definition of magnetic flux in maxwells, which derives directly from the definition of the magnetic field intensity, **H**, whereas in the SI system flux is inherently defined through the magnetic induction, **B**. These inconsistencies arise simply because the original pole concept is retained within these systems of units. While modernized definitions do exist, *e.g.* the oersted given in terms of the field of a current rather than that of a pole, these no longer have the authority of official recognition and have done little to supplant the older definitions.

The magnetic potential [13, pp. 38-41; 7 pp. 470-471], seems to be a concept which has survived poles. The electrical potential difference, $V_{ab}^{e} = \int_{a}^{b} \mathbf{E} \cdot \mathbf{d}\ell$ readily gives the work qV_{ab}^{e} done in moving a charge q from a to b through the electric field **E**. On the other hand, what does the analogous concept of magnetic potential difference $V_{ab}^{m} = \int_{a}^{b} H \cdot \mathbf{d}\ell$ give us without poles to be moved along the path? No doubt some interpretation may be found, as in the definition of magnetic reluctance where the need for poles is no longer apparent, but the point is that the definition itself tempts us to think of **H** like **E** in terms of a force field and leads some authors to refer to moving actual poles around a path [11, pp. 259 and 262].

2.8.3 Lack of Adherence to a Single System of Units

Notwithstanding the conceptual difficulties caused by retaining emu and other related CGS systems of units, there are the practical considerations. For years it has been necessary to do battle with (at least) two separate systems of electromagnetic units and, as a consequence, differing forms of equations. Many textbooks give two sets of equations or translation tables, some deliver lengthy recommendations of one system over the other, often advocating one for aesthetic reasons while saying the other is practical and for the convenience of engineers [37; 9, p. vii; 22, p. 621].

Such arguments are specious. While SI has been adopted as the ‘preferred’ system, the fact is that emu terms and units abound in the literature and can hardly be ignored. The critical issue, however, is not

the inconvenience of conversion factors, rather it is the old concept of magnetism which is still allowed to survive through the older definitions employed within emu. Unfortunately, as we have just discussed, this link with the old ideas often results in confusion about the basics, not just the units.

2.8.4 Inconsistency between emu and SI Poles

In isolation, the magnetic pole description provides a practical means of determining **H** in magnetostatic problems involving permanent magnets provided that we are able to specify the distribution of magnetic poles. The main difficulty lies in making this consistent with the modern view of magnetism produced by currents. In order to do so, the unit pole must be considered as induced [9 pp. 241-242; 11 pp. 5-6], rather than absolute. This is necessary simply because the pole distribution does not lead directly to **B**. One is first effectively determining the field **H** from the poles and then finding **B** as a result. The force between poles must depend on **B** in order to be consistent with the circulating current approach. Therefore, if we have $F = \overline{p}B$ for the magnetic force[2] in place of $F = pH$, then we must have $F = (\mu\,\overline{p})H$, so that $p = \mu\overline{p}$. The original pole is p while the so-called induced pole is $\overline{p}$, although we could argue that this should be stated the other way around.

In an attempt to bring over the pole concept into the then MKS system, a definition for poles and a Coulomb's magnetic law was included [13, pp. 40-41 and 43], but in contrast with the original emu definitions, these are based on the modern view, *i.e*. induced poles. The problem is that both the poles involved and the force laws are incompatible.

This anomaly can be examined by comparison of their definitions in terms of the inverse square law

$$F_{12} = \frac{p_1 p_2}{\mu_r r_{12}^2} \qquad \text{EMU}$$

$$F_{12} = \mu_r \mu_0 \frac{\overline{p}_1 \overline{p}_2}{4\pi r_{12}^2} \qquad \text{MKS} \tag{4}$$

One can see that μ_r, the dimensionless relative permeability of the intervening medium, appears in the *denominator* of the one form and in the *numerator* of the other. This is not just a mathematical difference – it was the borrowing of the original Coulomb force concept that was ill founded. As we shall explain in Section 3.2.4 below, the force between poles should *increase* as the material permeability increases, and not, as the early theorists held, the other way around. Had they but known this fact then they would no doubt have questioned the assumption of a direct parallel with the Coulomb field. With the same model at heart, the result cannot be one thing for the electric field and another for the magnetic field. As Stratton comments [9, p.239], "the properties of magnetic matter can be described more naturally…without fictitious 'magnetic charges' ". Indeed, it should now be said that the continuing existence of the older notions is quite undesirable from the standpoint of having a clear and consistent approach to the basics. We should no longer cling to them for purely historic reasons or as conveniences, but even the technically consistent representation of poles in the MKSA system only helps to prolong the concept. Poles no longer exist in SI, and those who still favor the notion of poles probably do so by their longstanding association with the emu system.

---

[2] In this context we can use poles with the field **B** provided we mean by a pole the positive half, mathematically speaking, of a real magnetic dipole. See Section 2.9.3.

2.9 Fundamental Field and Auxiliary Field

2.9.1 B is the Primary Field Associated with Magnetic Force

In a nutshell, the problem we are addressing stems from the original magnetostatic concept that the magnetic force field is defined in terms of **H** [38, p159], whereas it is *now* accepted that **B** is the fundamental field [22, p153; 9, pp. 12 and 242, 16, p. 157; 39, p. 17; 40]. and that it is the origin of the magnetic force, as stipulated in the correctly stated Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$, Equation (3). As we have already discussed, **H** was defined as a force field in a manner paralleling **E**, with the electric charges being replaced by magnetic ones, called poles. The description of magnetism always tends to be more complex than electricity, and in keeping with this we cannot simply rectify the situation by calling **B** the magnetic force field, or similar, because

- the force generated is not even parallel to **B**, and
- **B** is an axial vector and not a true vector [41, pp. 39-40] (see Section 2.9.2).

But neither should we fudge the issue and say that **H** is the *force* field. It may have an interpretation as an analogue of a force field, *e.g.* within in the expression $\int\mathbf{H}\cdot d\mathbf{B}$ for energy density [7, p. 494] and the dubiously defined magnetic potential $\int\mathbf{H}\cdot d\mathbf{l}$ discussed in 2.8.2 above, but not as the fundamental magnetic force field.

2.9.2 Ambiguous Real versus Axial Nature of the Magnetic Field Vectors

A true vector, also sometimes called a real vector or polar vector, is inverted if the co-ordinate system is inverted, *i.e.* $(x, y, z) \rightarrow (-x, -y, -z)$. Examples are force, position and velocity. On the other hand, a pseudo-vector, also referred to as an axial vector, is not inverted. Examples are torque and angular momentum. The cross product of two vectors of the same sort is an axial vector, while for different sorts it results in a true vector, and multiplication of any vector by a scalar does not affect its type. The field **E** is related to force by a simple scalar factor (charge) and so we can readily regard it both as a true vector and as a force field. There is objection, however, to referring to **B** as a force field since **B** is not a true vector and cannot therefore directly represent a force. But in fact an equal difficulty applies to **H**. Were **H** to be defined analogously to **E**, as in the original concept of magnetism, then it would indeed be a true vector. But because **H**, **B** and **M** (see Section 2.12.4 below) all require to be compatible, they must all be either true or axial vectors, otherwise linear combinations of **B**, **H** and **M**

would be neither axial nor polar in general. As a particular example, we could not write, as we so often take for granted, $\mathbf{B}=\mu\mathbf{H}$ with μ being a simple scalar. There is no longer any argument about **B** being an axial vector[3], and consequently all three magnetic field quantities must be axial vectors. This result

[3] To see this, consider the field **B** generated by a current in a circular loop of wire centred on the origin. A charge at position r on the loop has velocity **v**, say. On applying an inversion, $\mathbf{r}\rightarrow -\mathbf{r}$ and $\mathbf{v}\rightarrow -\mathbf{v}$. That is to say, the charge now has the opposite velocity but is *on the opposite side of the loop*. It is therefore continues to go round the loop in the same direction leaving **B** unaltered.

is completely consistent with Maxwell's equations, Equations (20), with **J** being taken as a true vector based on its equivalence to a flow of charge[4], $\rho\mathbf{v}$, where ρ is a scalar. Note that **J** is sometimes taken to be an axial vector by identifying its vector character with the normal to an infinitesimal element of surface through which a *positive* scalar current passes. In this case, applying an inversion leaves **J** unchanged so that **H** requires to be a true vector. The latter definition is therefore to be avoided in general, but see Nye [41, p. 54] for a full discussion.

Taken on their own, however, whether **M** and **H** are axial or polar simply depends on how they are defined. If they are defined based on poles, they must be polar. If they are defined based on circulating currents then, like **B**, they must be axial. Retaining a field **H** that is polar rather than axial may be consistent with being able to use it to represent a force field, but in a world without poles this is purely artificial. To do so simply provides an operational model which can be used to simplify analysis rather than a physical model which, although less convenient, more truly reflects nature. We cannot say that **B** is a force field in the proper sense, but neither can we say that **H** is, except in a restricted operational sense. Importantly, and as we shall subsequently reinforce, it is the field **B** rather than **H** that is, under all circumstances, directly associated with the Lorentz force - but this is nothing to do with the characters of **B** and **H** being axial or polar.

2.9.3 Definition of Magnetic Force Field without Poles

As explained in 2.8.2, the use of poles to define the magnetic force field is still extant in emu [13, p. 40]. But we must recognize that in reality there are no poles. Moreover, **B** rather than **H** is the origin of the magnetic force, even though we cannot truly refer to **B** itself as a force field in the usual sense (2.9.1 and 2.9.2 above). It is still a problem that the old pole concept of magnetism is easier to grasp at a basic level. It would be helpful if the true position could be restated in some simpler way without the notion of poles *per se*. Certainly, we can still visualize a magnetic 'force field' based on the alignment of an infinitesimal permanent magnetic dipole or current loop rather than the force exerted on a conceptual pole and whenever we refer to this meaning we shall use the quotation marks. As there never has been the possibility of using free poles in an experiment to measure magnetic field strength, the original method, due to Gauss [38, p. 160-161], was in any case based on the use of a test dipole. This is still valid for present day purposes as in the real world the behavior of any isolated test dipole is unaffected by the model we choose for it. As before, the orientation of the dipole gives the field direction, while the torque required to displace it from this orientation by a given small angle is a measure of the field strength. When we later refer to a magnetic 'force field', this is exactly what we shall mean. At this stage, this definition could apply either to **B** or **H**, and in a vacuum there could be no discernible difference. We shall, however, deal with the critical case that applies within a magnetic continuum in Section 3.2.4.

[4] The term convection current is sometimes used in this context. It would seem to indicate that the current flow involved is solenoidal, *e.g.* as in an eddy current, rather than flowing from a source to a sink.

2.9.4 Primary and Auxiliary Fields Defined

As a way round the joint problems of unhelpful nomenclature and identification of actual 'force fields', the terms primary and auxiliary fields are helpful. Both **B** and **H** appear in Maxwell's equations, which we will come to in due course. One of the two can be held to be the fundamental, or microscopic, quantity *which will always be required in free space*, while the other is introduced as an auxiliary field [39 p. 18; 42], *which is only genuinely required to describe the macroscopic effects arising in magnetic materials*. If **H** were the original force field that applied before the introduction of magnetic matter, then **B** would be the auxiliary field that allows the description of magnetic materials, or *vice versa*.

While the situation with the electric field is similar, there is no such confusion. **E** is, and always has been, the fundamental force field, or microscopic field, while **D** is purely an auxiliary field that allows for the macroscopic description of dielectric materials.

However it was that **B** came to be originally defined, nowadays it is recognized that it does represent the origin of magnetic force. Consistent with this standpoint, it is also appreciated that magnetic effects originate from circulating electrical charges, as originally conceived by Ampere [5, 6] and asserted by Maxwell [3, Vol. 2, p275], or by their quantum equivalent, rather than from the completely separate original concept of magnetic charges, or poles. We shall see in Section 4.2 that, more generally, Maxwell's equations can be arranged in terms of the fundamental fields **E** and **B** alone, to form the basic microscopic equations, with **D** and **H** being brought in through so-called constitutive relations in order to describe the effects of matter. In doing so, the important step is the identification of magnetic fields with circulating currents alone.

2.9.5 Summary of the Characteristics of the Field Vectors

Table 1 summarizes the main properties of the field vectors. **H** is a force field only if we consider a Coulomb theory of magnetism with poles. It then must also take the character of a true vector, rather than an axial one, which is no longer consistent with the character of **B**. Although it originates a force for moving charged particles, **B** is not a force field *per se* for reasons explained in 2.9.1 and 2.9.2 above.

Table 1: a summary of the key characteristics of the field vectors

Property	E	D	B	H
Fundamental field (origin of forces as in the Lorentz force)	●		●	
Auxiliary (required to account for macroscopic media on the basis of free charges and currents alone)		●		●
True Vector	●	●		
Axial Vector			●	●
Actual Force Field	●			

2.10 Special Relativity

2.10.1 Origins of Magnetism within the Theory of Special Relativity

It is now understood that magnetism arises naturally as a basic result of the theory of special relativity in which the magnetic field emerges out of the need to describe the observed electric force between moving charges. Just like the original Coulomb's law, it gives the force between particles directly, but now we must also take into account the particle velocities. The result agrees with the equation for **B** derived classically from Ampere's force law, Equation (11) below. Furthermore, the result for the transformation of the electric field, **E**, given for example by Jackson [22, pp. 380-382], agrees with the Lorentz force, Equation (3).

The satisfying outcome is that special relativity effectively yields the Lorentz force without any *a priori* need to prescribe the separate concept of a magnetic field. The Lorentz force therefore *defines* the magnetic field.

It is clearly quite difficult at the earlier stages of teaching electromagnetic theory to get this point across except perhaps in the mention, but on the other hand this does not prevent us from keeping the information imparted as consistent as possible with this cornerstone of the theory. By the stage that some of the basic results of the theory of special relativity have been appreciated, students would be in a position to make the connection between Coulomb's law and the magnetic field. In Section 5 below we examine how the basic results can be applied to some illustrative cases: transformation of the electric field, Faradays law of induction and displacement current, and finally, the universality of the speed of light *in vacuo.*

2.10.2 Significance of Magnetism in the Theory of Special Relativity

Text books on special relativity usually cite a familiar list of 'relativistic corrections' which have been duly confirmed experimentally, furnishing evidence in favor of the theory. In some of these books, magnetism is then shown to arise as a consequence of the theory, but very few, if any, ever mention magnetism as being the only basic evidence for the theory of special relativity that we can observe in everyday situations. All relativistic 'corrections', being typically of order $(v/c)^2$, are very small at ordinary velocities. Here v is the velocity of, say, a moving particle or other frame of reference, and by 'ordinary velocities', we mean velocities such that v/c is no greater than about 10^{-5}. The key difference that makes magnetism so readily observable at 'ordinary' velocities is that while the electric force between bodies filled with electrically neutral matter vanishes, the relativistic correction does not do so when they maintain a current. The magnetic force can therefore be observed quite readily when it is not masked by the presence of a net electric force which would be many orders of magnitude greater. While we can indeed encounter very large magnetic forces, this is due to the fact that the underlying electric forces, if unbalanced, would be quite enormous by comparison!

A much more subtle point is that Ampere's force law, Equation (11), does not obey Newtonian relativity as the forces exerted by one infinitesimal current element (or moving charge) upon another are not generally equal and opposite. This can be readily seen from Equation (11) for the case of two current elements, one parallel to the spatial vector $\mathbf{r}_{12}$ separating them while the other is perpendicular to it. The force acting on the perpendicular element is zero while the force acting on the parallel element is not, so that they evidently cannot be equal and opposite. To an observer at rest, therefore, there is a nonzero net force acting on the pair even in the absence of any external influence. How well forgotten is this inconvenient fact! Inconvenient, though, only because its explanation draws us into special relativity, something all but advanced textbooks on electromagnetic theory generally seek to avoid.

Surprising as this kind of behavior may seem even today, the whole supposition that the force between charges varies with their velocities flies in the face on Newtonian relativity, as we can make the forces come and go depending on the motion of the observer. The variable element, the magnetic force, is therefore at the very heart of special relativity and not just one of the consequences of it.

2.10.3 Magnetism as Evidence for the Theory of Special Relativity?

Magnetism is has been part of the everyday world for centuries. It is in fact so commonplace that we do not even recognize it as evidence for the theory of special relativity, like Monsieur Jourdain in Molière's play [43], who was greatly surprised, and impressed, to discover he had been speaking prose all his life. The main evidence generally cited in support of special relativity is the Michelson-Morley experiment together with the aberration of starlight pointing to the constancy of the speed of light and the absence of an ether, and Fizeau's experiment on the speed of light in moving liquids (see Section 5.4 below). In view of what we have just discussed in the preceding section, it is a pity, therefore, that we rarely see statements such as:

> Because we know that magnetic poles do not exist, magnetism must be explained in terms of existing forces. Since the theory of special relativity applied to the electric force would give rise to a force identical to the magnetic force, we must consider the observation of magnetism as *prima facie* evidence supporting the theory of special relativity.

27

Rather, we see it stated the other way round, with magnetism being treated as an application rather than being placed along with the other evidence. This, however, is simply part of the legacy, a result of how the subject developed.

2.10.4 Issues Arising from Advanced Relativistic Theories

We must, however, be quite careful when we move away from these basics. Special relativity and general relativity have been applied to a wide range of problems including electrodynamics. In relativistic formulations [9, pp. 71-72; 11, pp. 384-385], the use of tensors such as $\mathfrak{F}$, which combines **E** and **B**, and $\mathfrak{G}$, which combines **D** and **H**, to describe fields changes little about the characteristics of **E**, **B**, **D** and **H** themselves.

The main issue here is that there are variants on this formulation in which **E** and **H** are combined as a tensor, and **H** is used in place of **B** in the Lorentz force [44]. While within a particular representation this may stem from convenience rather than necessity, it hardly serves to promote clarity and understanding about the basic roles of the four electromagnetic quantities. In some ways it may seem justified to use **H** as a force field like **E**, but it is not consistent to do this within the Lorentz force, which involves moving charges rather than magnetic poles.

While this may present little problem for experts in the more advanced formulations of relativistic electrodynamics, the rest of us must exercise caution in interpreting the roles played by **B** and **H** in these special contexts.

We will return to special relativity in Section 5.

2.11 Are Magnetic Poles Completely Redundant?

An approach consistent with the modern view does not require magnetic poles, and so neither does it require any alternative description for the magnetic ‘force field’. Although it should now be regarded as largely historical, the pole description is intriguing in that it may still be applied to magnetic problems and, with due care, it does achieve the correct results. But we must be clear that it does not give the correct conceptual picture. It works on its own but not within electromagnetics as a whole, and so we should really avoid using it as a basic introduction to the subject. It is misguided to think that poles will help understanding when later we have to alter the description to conform with an entirely different picture. In a sense, the pole is the ‘flat earth’ of electromagnetic theory – we can readily work with poles as a means to an end, but we should not confuse them with physical reality.

Despite being physically invalid at the fundamental level, the pole description can be placed on a sound mathematical basis, as shown in Section 4.2. In fact its main value, if any, would appear to be as a mathematical simplification over the circulating current theory of dipoles. As we shall later see, Section 3.2.1, the interaction between current dipoles is considerably more complicated and far less intuitive than the interaction between individual poles. This kind of situation is not uncommon, but it is an intriguing fact that there are many respects in which mathematical pole-based description appears to be fundamentally at odds with the circulating current description, yet it still seems to give equivalent results. We shall return to this puzzle in Section 6.

2.12 Real Matter and the Constitutive Relations

2.12.1 Microscopic versus Macroscopic

Within what we may term a *microscopic* basis, interactions between *all* the particles present in any system under consideration must be accounted for explicitly. This is applicable to situations where we are dealing with given charge and current distributions *in vacuo*. On the other hand, by a *macroscopic* basis we mean that we have matter present, giving rise to *unknown* distributions of charge and current that require to be determined from knowledge of their interaction with the known ones. In this macroscopic picture, matter is most easily characterized as being a continuous homogeneous medium rather than having a detailed microscopic structure. The microscopic structure can be dealt with separately, for example, in order to relate the macroscopic properties of the medium to the properties of its basic ingredients on the atomic or molecular scale, as in the Lorentz-Lorenz model referred to in Section 1.5 above.

2.12.2 Real Media

Real media contain charges that are bound to molecules[5]. In the quantum mechanical description, charges may also have an associated spin that carries an equivalent magnetic dipole moment, and even in the semi-classical atomic description, charges can move in orbits giving rise to net magnetic dipole moments. The influence of externally applied electric and magnetic fields causes charge and current on the microscopic scale to be subjected to the Lorentz force, Equation (3) above. This in turn causes forces and torques on the molecules or ions to which they are bound, and the resulting motion is a problem in mechanics. Generally, a linear model for the displacements, or polarization, suffices. The displaced ions and molecules in turn give rise to their own contributions to the field affecting them, and so the whole problem of finding the actual field for a given applied field must be solved self-consistently, as in the Lorentz-Lorenz model. Within a real medium, therefore, the local field values that provide the forcing term driving the equations of motion are quite different to both the externally applied field and the average *macroscopic* field within the medium, that is, the field that pertains when we treat it as though it were conceptually a continuum.

2.12.3 Free and Bound Quantities

The terms free and bound may have valid connotations of distinguishing, say, charge that is *bound* within some matter from charge that is entirely separate and therefore *free* from any other matter. But the real significance of these terms is that free quantities represent the independent, or explicit, variables in a system, *the sources*, while the bound quantities are generally the dependent, or implicit, variables. *Bound* quantities are entirely associated with the macroscopic picture, whereas in the microscopic picture all quantities must be considered as being *free*.

2.12.4 The Electric and Magnetic Polarizations

On a macroscopic scale, the quantities **P** and **M**, which are the electric and magnetic polarizations per unit volume respectively, have been defined so as to describe the average instantaneous displacement, or polarization, of the bound charges and magnetic dipoles[6]. In the macroscopic context, **P** and **M** are fundamental in the sense that we can relate them directly to the individual molecular electric and magnetic dipole moments, **p** and **m**, which certainly are fundamental at the molecular, *i.e.* microscopic, level. Furthermore, since **P** and **M** prescribe the electric and magnetic dipole distributions, they are associated with electric and magnetic fields respectively, and we can determine

---

[5] As before, we include ions and atoms within ‘molecules’.

[6] Here displacement also includes any dipole rotation. Note also that while the term magnetic polarisation literally means the magnetisation, **M**, in some places, e.g.IEC60050 IEV 121-11-54, it is taken to be $\mu_0\mathbf{M}$ with the symbol **J** (also used for current density!).

these directly from the given distributions. While in principle the introduction of **P** and **M**, together with the historically defined free-space quantities **E** and **H**, are all that is required in order to take account of media other than free space, two additional macroscopic quantities, **D** and **B** were introduced and defined as

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$$
$$\mathbf{B} = \mu_0 \left(\mathbf{H} + \mathbf{M} \right) \tag{5}$$

The motivation for this was that **D** is analogous to **E** but is associated with free charge alone, as in $\nabla \cdot \mathbf{D} = \rho_{free}$, while $\nabla \cdot \mathbf{E} = \rho_{total}$. The equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot \mathbf{H} = \Pi$, where Π is the density of poles, was seen as a direct counterpart of the electric case, given that Π_{total} can be written simply as Π since there are no free poles. But, there are no poles whatsoever, and the introduction of equations for magnetic quantities simply on the basis of paralleling the electric ones is physically wrong, even if mathematically justifiable. We shall show a different rationale below that avoids poles altogether and is based on **B** being the primary magnetic field.

2.12.5 H as the Dependent Variable

The so-called constitutive relations started out in the form of Equations (5) above with **D** and **B** as the dependent variables and with the magnetic form being the complete parallel of the electric one, allowing of course the trivial difference that the counterpart of **P** is taken as $\mu_0\mathbf{M}$ rather than **M**, which is simply a matter of definition.

If we reverse this picture, as in the modern description where **H** depends on **B** rather than the other way around. Somewhat annoyingly, the associated magnetic constitutive relations as in Equations (6) below no longer parallel the electric ones. In terms of **H** being the dependent variable, the appropriate form of the constitutive relations is:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$$
$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{6}$$

Some authors, wishing to draw attention to the point, do introduce these forms of the equations [30, pp. 18 and 27; 22, p. 153; 45, p. 276] but they revert to the original forms which are still commonplace since they seem to be intuitively more acceptable. First of all, this may be because of the temptation to have the two sets of equations in a ‘neat’ parallel form. Secondly, it is hardly intuitively obvious why the auxiliary magnetic field, as we shall now refer to **H**, should appear to *reduce* with increasing magnetization of the medium, while the **D**, the electric displacement, or auxiliary electric field, *increases* along with the degree of polarization.

As we shall see in Section 4, the true comparison with $\nabla \cdot \mathbf{D} = \rho_{free}$ and $\nabla \cdot \mathbf{E} = \rho_{total}$ is not $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot \mathbf{H} = \Pi$, but rather, $\nabla \times \mathbf{B} = \mathbf{J}_{\mathrm{tot}}$ and $\nabla \times \mathbf{H} = \mathbf{J}_{\mathrm{free}}$, where **J** is current density. This requires no poles whatsoever, and it is clear that any comparison with the electric equations is on a basis that allows for the different vector character of the magnetic field quantities and for vector rather than scalar sources.

Unfortunately, Equations (5) rather than Equations (6) are generally still taken as the basic form, in spite of the outmoded description of the magnetic quantities that this reflects.

2.12.6 The Linear Description of Media Is Not Fundamental

With simple linear relationships between induced and applied fields, Equations (5) become

$$\begin{aligned}\mathbf{D} &= \boldsymbol{\varepsilon}\cdot\mathbf{E}\\ \mathbf{B} &= \boldsymbol{\mu}\cdot\mathbf{H}\end{aligned} \qquad (7)$$

or, as they may be written to better reflect the modern view as expressed by the form of Equations (6)

$$\begin{aligned}\mathbf{D} &= \boldsymbol{\varepsilon}\cdot\mathbf{E}\\ \mathbf{H} &= \boldsymbol{\mu}^{-1}\cdot\mathbf{B}\end{aligned} \qquad (8)$$

The proposition of linear relationships as in Equations (7) or (8) is not of itself fundamental and is only required in order to allow the straightforward solution of a great number of common physical problems, such as wave propagation in a given medium. When $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are simply scalar constants, with ε and μ both, in the main, being close to or greater than unity, the older forms in Equations (7) somehow seem more amenable [22, p153 (footnote)]. This is simply one more convention that makes us tend to see a parallel between **E** and **H**. Most of us are not quite at ease with routinely writing instead $\mathbf{H}=\mu^{-1}\mathbf{B}$, although there is no logical reason why we should not do so.

2.12.7 The Definitions of D and H are not Simply Arbitrary

It could be said that the definitions of the auxiliary fields **D** and **H** are arbitrary, but this is not the case. One could also ask why they are necessary at all, given that **P** and **M** are more fundamental quantities

which, for example, the Ewald-Oseen approach employs directly without any recourse to **D** and **H**. The answer to this question lies first of all in the fact that the use of **B**, **E**, **H** and **D** allows Maxwell's equations to be written in a form that depends only on *free* sources of charge and current, rather than all sources including the *bound* ones. A secondary but useful point is that these four fields all obey specific boundary conditions whereas **M** and **P** do not.

2.12.8 The Boundary Conditions

The combination of the four boundary conditions, as represented in Figure 1, allows straightforward solutions to some common problems, such as the transmission and reflection of electromagnetic waves at a plane interface. In general, they may be stated as

$$\begin{aligned}\mathbf{E}_1^{//}-\mathbf{E}_2^{//}&=\mathbf{0} \qquad & D_1^{\perp}-D_2^{\perp}&=\sigma_{free}\\ \mathbf{H}_1^{//}-\mathbf{H}_2^{//}&=\mathbf{K}_{free} \qquad & B_1^{\perp}-B_2^{\perp}&=0\end{aligned} \qquad (9)$$

Here the superscript $\perp$ refers to a component perpendicular to the interface, while // refers to the part of the vector that is parallel to it. The symbol σ_{free} represents the density of free surface charge, Cm^{-2}, while $\mathbf{K}_{free}$ is the density of surface current, Am^{-1}. It should be noted, however, that these four boundary equations are only special cases of each of Maxwell's four equations, but while they may add nothing new in themselves, they do underline the usefulness of the choice of vectors.

Somewhat intriguingly, the boundary conditions for **D** and **B** appear very similar, as do those for **E** and **H**. The former conditions both involve only the parallel components whereas the latter involve only the perpendicular parts. So once more, if we are not on our guard we may draw the wrong conclusions as to parallels between the vectors. The truth is that these components are associated with the particular field vectors because in Maxwell's equations $\nabla\times\mathbf{E}$ and $\nabla\times\mathbf{H}$ are involved in the one set and $\nabla\cdot\mathbf{D}$ and $\nabla\cdot\mathbf{B}$ in the other. Looking at it another way, though, the conditions for **E** and **B** are both homogeneous, while those for **D** and **H** involve free sources, and this is actually how the physical parallels should be drawn.

A comparison of the key boundary condition properties of the four field vectors is shown in Table 2.

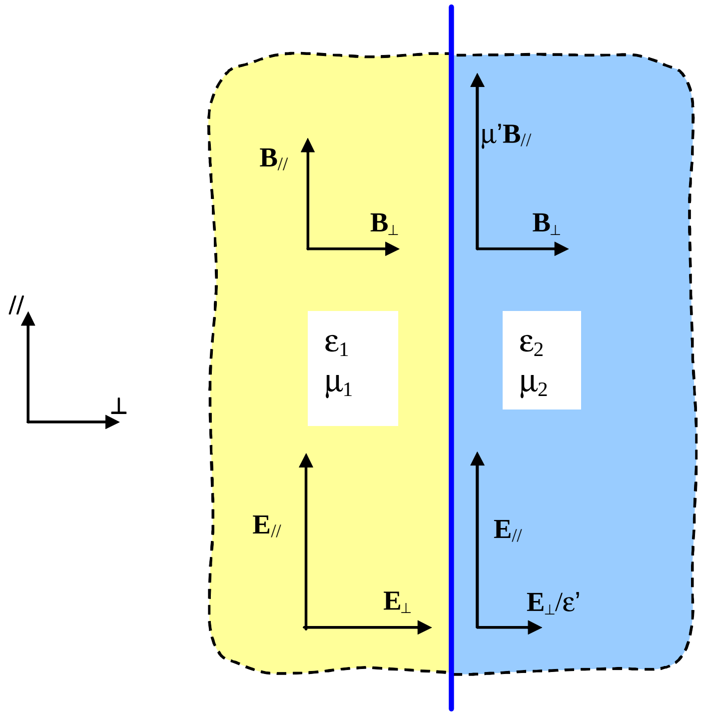


Figure 1 : The boundary conditions for E and B at a charge-free plane interface. On opposite sides of the interface we have media 1 and 2, where $\varepsilon' = \varepsilon_2/\varepsilon_1$ and $\mu' = \mu_2/\mu_1$. Both D and B obey a boundary condition in which the perpendicular component (⊥) is continuous, while E and H obey another in which the parallel component (//) is continuous. This can be misinterpreted as implying that B and D are analogous, both being auxiliary fields, while E and H are analogous, both being primary force fields. The different boundary conditions for E and B arise, however, from their origin as divergent and solenoidal fields respectively, in keeping with their separate origins in charges and currents respectively.

Table 2: A summary of the boundary condition properties of the field vectors and their counterparts in Maxwell's equations

Boundary Conditions	Maxwell's Equations	E	D	B	H
Homogeneous	Homogeneous	●		●	
Inhomogeneous	Inhomogeneous		●		●
Applies to Parallel Components	Involves Curl	●			●
Applies to Perpendicular Parts	Involves Divergence		●	●	

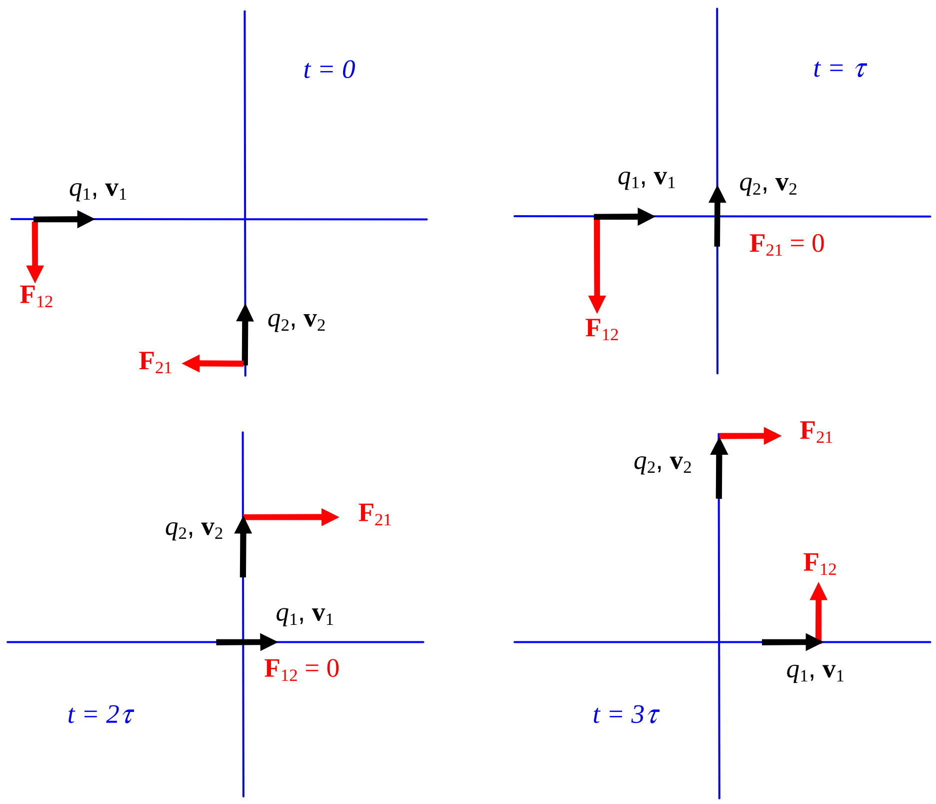


Figure 2 : How Newton's third law is broken. Two particles q_1 and q_2 are on orthogonal trajectories, as shown at instants t = 0, τ, 2τ and 3τ. For simplicity we take them both as being positively charged, but this in no way affects the generality of the results. As can be seen, the magnetic interaction as expressed in Equation (13) does not instantaneously obey Newton's Third law for at no instant do we have $\mathbf{F}_{12} = -\mathbf{F}_{21}$. It is easy to check these results conceptually by treating one of the moving charges as a current and deducing the direction of its field B at the location of the second charge by means of the right-hand screw rule, and also by noting that B must vanish along the axis of motion. The term v×B from the Lorentz force then gives the direction of the force on the second charge. The problem, therefore, is not with the equation, rather the equation exposes the physical problem.

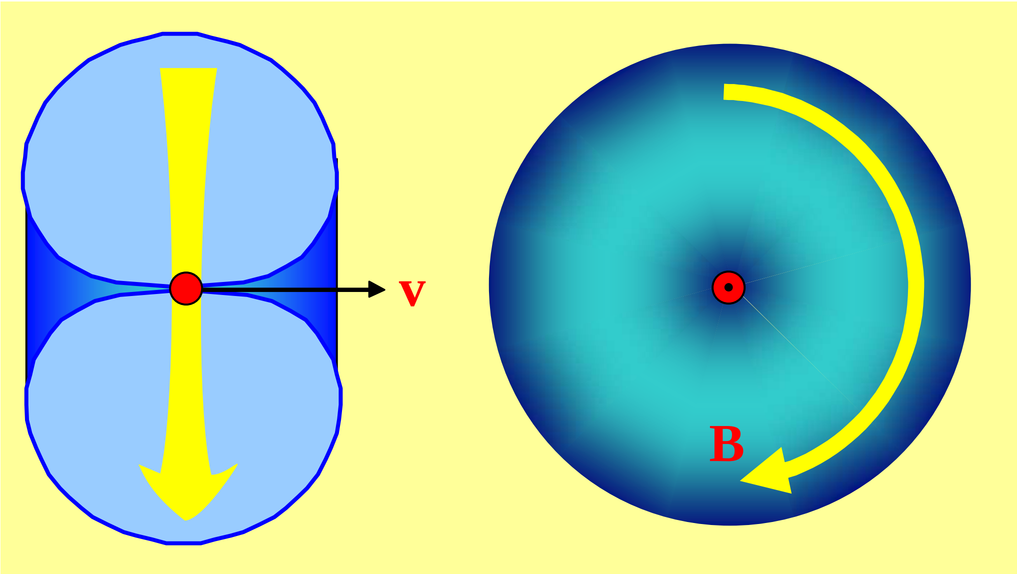


Figure 3 : The magnetic field due to a moving positive point charge. The field has constant magnitude over the toroid-like surface of rotation, while its direction lies in the surface and perpendicular to the motion as shown. The left-hand view is a plane section through the axis of motion, while the in right view the motion is out of the page.

2.12.9 Are D and H of any Fundamental Significance?

But the question still remains, are **H** and **D** of any more significance than that they are convenient? After all, the essence of Maxwell's contribution to electromagnetic theory is widely quoted as being the introduction of the displacement current $\frac{\partial \mathbf{D}}{\partial t}$ into Ampere's Circuital law, and is that not sufficiently significant? The point is debatable, as it is often possible to attach a physical significance to quantities that are arguably not in themselves fundamental. But we can say that as a minimum we require the fields **E** and **B** to represent the origin of the electric and magnetic forces, together with **P** and **M** to represent the state of any matter present. All four of these quantities have obvious physical definitions and are sufficient in themselves without any need for **H** and **D**. That is not to say that **H** and **D** are not without meaning or of significance, they are simply not so basic.

3 Fundamentals

3.1 Coulomb's Law and the Lorentz Force

Coulomb's law, Equation (1a), the fundamental equation of electrostatics, can be employed to define the electric field. For a point charge at the origin, the field **E** at the point **r** is given by

$$\mathbf{E}(\mathbf{r}) = \frac{q\hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2} \qquad (10)$$

There is no serious challenge to this description except that Coulomb's law itself is specific to stationary charges. It is necessary to combine the effects of Coulomb's law with a magnetic force in order to specify the forces acting on a point charge in motion. The combined electrostatic and magnetic force is known as the Lorentz force, referred to Equation (3). Rather than treat this as a separate effect, however, we can follow the work of Ampere on the magnetic forces between currents.

The force between two currents I_1 and I_2 flowing in the infinitesimal line elements $d\mathbf{l}_1$ and $d\mathbf{l}_2$ is given by

$$d\mathbf{F}_{12} = \frac{\mu_o}{4\pi} \frac{I_1 d\mathbf{l}_1 \times \left(I_2 d\mathbf{l}_2 \times \hat{\mathbf{r}}_{12}\right)}{r_{12}^2} \qquad (11)$$

Here $d\mathbf{F}_{12}$ represents the force acting on element 1 as a result of the current carried by element 2, while $\hat{\mathbf{r}}_{12}$ is a unit spatial vector directed from 2 to 1. While, as discussed in Section 2.10.2 above, $d\mathbf{F}_{12}$ and $d\mathbf{F}_{21}$ are not generally equal and opposite, the forces $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ integrated over actual closed current loops conveniently do balance [22, p. 136; 45, pp. 177-178].

Although he expressed it in a more basic form [5], we may nevertheless refer to Equation (11) as Ampere's force law. Ampere is given credit for its deduction [22, p. 135; 6, pp. 92-97; 24, p4.12] nearly a century before the Lorentz force. It deserves specific recognition in that it provides the earliest and most basic statement of the interaction between two currents in just the same way that Coulomb's law is the most basic statement of the interaction between two charges.

Because it expresses the basic magnetic interaction between infinitesimal current elements, Ampere's force law leads directly to the deduction of a force that arises between two charged particles only when they are both in motion (this not without some conceptual problems, as we shall soon discuss). All that is required is to express Equation (11) in microscopic form by replacing the vector current elements with moving point charges, q, and by making the equivalence between $Id\mathbf{l}$ and $q\mathbf{v}$, where $\mathbf{v}$ is the charge velocity, to obtain

$$\mathbf{F}_{12} = q_1 \mathbf{v}_1 \times \frac{\mu_o}{4\pi} \left(\frac{q_2 \mathbf{v}_2 \times \hat{\mathbf{r}}_{12}}{r_{12}^2} \right) \qquad (12)$$

By including the electric field, we must therefore have the basic form of the Coulomb force between moving particles

$$\mathbf{F}_{12} = \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}^2} \left(\hat{\mathbf{r}}_{12} + \frac{1}{c^2} \mathbf{v}_1 \times \left(\mathbf{v}_2 \times \hat{\mathbf{r}}_{12} \right) \right) \qquad (13)$$

The striking thing about Equation (13) is that it shows the magnetic interaction between two charged particles is simply a velocity dependent correction to the usual Coulomb's law for static particles. This correction applies only when *both* particles are in motion within the observers frame of reference and, being of order of $v_1 v_2/c^2$, appears to be entirely negligible at 'ordinary' velocities, *e.g.* up to Mach 10 the magnetic interaction is at most 10^{-10} of the electrostatic force. Surprisingly, however, there are commonplace situations in which the electrostatic part of Equation (13) completely vanishes, as when the charged particles in question are imbedded in electrically neutral conductors, and we are left only with the magnetic interaction. As everyday experience shows, however, the magnitude of the magnetic force on its own need not be negligible.

As already alluded to, however, there are difficulties with both Equations (11) and (13) concerning Newton's third law: "To every action there is an equal and opposite reaction". Taking first Equation (11), consider a situation in which $d\mathbf{l}_1$ and $d\mathbf{l}_2$ are mutually orthogonal, with $d\mathbf{l}_1$ parallel to $\mathbf{r}_{21}$ and $d\mathbf{l}_2$ perpendicular to it. We find $d\mathbf{F}_{12} = 0$, while by simply interchanging the roles of 1 and 2 we find $d\mathbf{F}_{21} \neq 0$. Although we know that this imbalance vanishes when the elemental forces are integrated over closed circuits and infinite conductors its existence on a microscopic scale is a little disconcerting. When we then come to address Equation (13), there is no such argument to fall back on as no circuits are involved, only two charged particles. The analogue of our example with two orthogonal current elements is now one in which we have charged particles q_1 and q_2 with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ such that they are on orthogonal trajectories, Figure 2. On the one hand it is futile to make attempts to argue away the fact that here Newton's Third law is being broken, but on the other we must recognize that the situation is entirely artificial. At low velocities, the interaction between the particles is dominated by the electrostatic force since the magnetic contribution, as we have seen, is negligible in such a case. Furthermore, at high velocities, Equation (13) cannot be used as a model for the interaction since a complete relativistic treatment is required [22, pp. 391 *et seq.*]

A simple observation that can be made to demonstrate the need for 'relativistic thinking', however, is that the forces can be made to balance if we simply move from our original reference frame, X, to a new frame, X'. Here we choose X' such that the average velocity of the two particles is observed to be zero. So when we, an observer at rest in X', move with a velocity $\bar{\mathbf{v}} = ½(\mathbf{v}_1 + \mathbf{v}_2)$ with respect to X, the particles now have velocities[7] $\mathbf{v}_1' = ½(\mathbf{v}_1 - \mathbf{v}_2)$ and $\mathbf{v}_2' = ½(\mathbf{v}_2 - \mathbf{v}_1)$, so that their relative motion is in opposite directions along *parallel* trajectories. Here we have no apparent conflict with Newton's Third law since we can simply exchange $\mathbf{v}_1'$ with $-\mathbf{v}_2'$ and *vice versa* in Equation (13) while leaving the result unchanged. On the other hand, replacing $\mathbf{r}_{12}$ with $\mathbf{r}_{21}$ alters the sign so that, in all, $\mathbf{F}_{21} = -\mathbf{F}_{12}$ [8]. This reinforces the notion that any conflict is to do with observational issues and as such as such the theory of special relativity must be involved. We will return to this issue in Section 5.

Page and Adams [46], however, offer an explanation by arguing that in spite of the force imbalance, the total momentum is conserved when it is taken to comprise both mechanical and electromagnetic momentum. The exchange of mechanical and electromagnetic momentum accounts for the apparent force imbalance. The momentum density of a traveling electromagnetic wave is generally taken to be given by $\varepsilon_0\mathbf{E}\times\mathbf{B}$, but the context here is quite different since no waves are involved. The interpretation can only be that the total energy in the electric field, $\int \frac{1}{2}\varepsilon_0 E^2 dV$, is equivalent to a mass $\frac{1}{c^2}\int \frac{1}{2}\varepsilon_0 E^2 dV$

[7] Newtonian relativity involving only vector addition of velocities suffices for this change of reference frame since the velocities involved need not be large. Applying special relativity would be more complicated and would make little difference to the end result.

[8] Indeed, this remains true even if we superimpose a further motion on X′ with any velocity proportional to $(\mathbf{v}_2 - \mathbf{v}_1)$, as the trajectories of the particles will still remain parallel. In short, the Third law holds good for any reference frame X′ having a velocity $\alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2$ with respect to the original rest frame, where α is any real number. This includes the situation where either particle is at rest.

that can carry momentum $\frac{\mathbf{v}}{c^2}\int \tfrac{1}{2}\varepsilon_0 E^2 dV$ when the charge is in motion. Since **E** and **B** are related through Equations (14a) and (14b) below, this last expression can be seen to be directly linked to $\int \varepsilon_0 \mathbf{E}\times\mathbf{B}dV$. Both Stratton [9, pp. 103-104] and Jackson [22, pp. 192-194] give detailed accounts of the conservation of momentum including the electromagnetic contribution

While Equation (13) is of limited practical use, it is nevertheless of fundamental interest because

- It shows that the Coulomb force of electrostatics must be modified when the charges are in motion
- The additional contribution that arises when the charges are in motion accounts for the origin of magnetic force
- It involves fundamental interactions between particles alone and puts magnetism and electricity together, whereas Ampere's force law, Equation (11), can be read as a effect between currents that is unconnected with Coulomb's law
- Ampere's law, the Biot and Savart law (see below) and the Lorentz force may all be derived from it
- The asymmetry $\mathbf{F}_{21} \neq -\mathbf{F}_{12}$ provokes a serious problem for Newtonian physics that must be recognized as evidence for special relativity.

From Equation (13) we can write fields **E** and **B** that are consistent with the Lorentz force as

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_0 r^2}\hat{\mathbf{r}} \tag{14a}$$

$$\begin{aligned}\mathbf{B}(\mathbf{r}) &= \frac{q}{4\pi\varepsilon_0 r^2}\cdot\frac{1}{c^2}\left(\mathbf{v}\times\hat{\mathbf{r}}\right)\\ &= \frac{\mu_0 q}{4\pi r^2}\left(\mathbf{v}\times\hat{\mathbf{r}}\right)\end{aligned} \tag{14b}$$

These apply to a particle of charge q situated at the origin and having a velocity **v**. Integration of **B** over a given charge distribution gives the result in the more familiar form based on current density rather than discrete charges in motion. As illustrated in Figure 3, the form of **B** in Equation (14b) contrasts strongly with the radially symmetric Coulomb field in Equation (14a).

The magnetic field due to an infinitesimal current element, effectively the magnetic field of a moving charge, was the contribution of Biot and Savart [22, pp. 133-134; 7, p. 174; 9, pp. 230-232; 45, pp. 178-179]. As an aside, it is to be noted that two of the references cited here have the law of Biot and Savart, as it is known, defining **B** while the others have it defining **H**. The difference lies in whether we take the current involved to be total current, or simply free current (Section 2.12.3). In the former case it defines **B**, and in the latter it defines **H**. This kind of ambiguity never seems to arise with the electrostatic field where Coulomb's law is always treated as defining **E** rather than **D**, even though to do so would enable the definition to directly carry over into dielectric media. There is no doubt as to the form we have in Equation (14b), since we are dealing with microscopic rather than macroscopic fields (Section 2.12.1) and we are being specific that it is the field of the charge q alone that we are interested in, and not in the fields due to any bound charges that might be affected by these fields. What applies to **E** in this case applies to **B**, and it would be inconsistent to quote **E** on the one hand and **H** on the other.

We may take Equations (14a) and (14b) as fundamentally defining the forces between charged particles and their associated fields. Putting the two together with Equation (13), we derive the Lorentz force as in Equation (3). With **B** originating from Equation (14b), the magnetic field involved in the Lorentz force *must be* **B**. *The forces on specified charged particles and currents always depend on* **E** *and* **B** *and never directly on* **D** *and* **H** *- there is no option about using either microscopic or macroscopic quantities in this context.*

3.1.1 Verification the Lorentz Force

While we postulate that **B** is the fundamental magnetic field on the basis that all magnetization arises from currents, what is the proof? Does it carry over, for example, into magnetic media? It has been asserted by one author [47, p. 118] that the experimental proof of the correctness of Equation (3) has been demonstrated on numerous occasions, but if so, such proof is now hard to track down as it has not been given the place in the literature that it truly deserves - text books in particular appear to avoid the issue.

Lorentz in fact introduced his 'law' only in the form of a postulate to the effect that the force acting on a unit charge is given by [23; 48; 49, p. 14;1, vol. 1 p. 422]

$$\mathbf{F} = 4\pi^2 c^2 \mathbf{D} + \mathbf{v}\times\mathbf{H} \quad \text{or} \qquad (15a)$$

$$\mathbf{f} = \mathbf{d} + \tfrac{1}{c}\mathbf{v}\times\mathbf{h} \qquad (15b)$$

Note that here we have dropped the original gothic script. This was not, as we have seen the first recognition that magnetic fields caused forces on infinitesimal current elements, as recorded by Oersted and Ampere. Even Maxwell included the force on moving charges among his original equations, first expressing it in his original paper in terms of $\mathbf{I}\times\mu\mathbf{H}$ [2, p. 489] and then later, in his treatise, he unambiguously states the force on a unit charge in motion to be $\mathbf{v}\times\mathbf{B}$ [3; vol. 2, p240]. Nowadays the notation used by both Maxwell and Lorentz is quite unfamiliar, and there may be more to Lorentz' choice of the variables **d** and **h** (*i.e*. **D** and **H**) than meets the eye. But because of their use, neither of the equivalent forms in Equation (15) is unambiguously recognizable as being identical to Equation (3) which is based on the fields **E** and **B.** How and when this later form came into regular use is now obscure, as it is simply stated without comment in most of the recent literature.

One would think that are some simple observations in magnetic materials that would provide an experimental basis for distinguishing between **B** and **H** in the Lorentz force, for example

- electron spin resonance, ESR, in magnetized ferrites,
- the Hall effect within a conducting magnetic sample, and
- the force on a magnetically soft conductor carrying current in the presence of a magnetic field.

Surprisingly perhaps, ESR shows no difference between **B** and **H**, the Hall Effect offers no clues as it is anomalous in magnetic conductors, while it turns out that the force on a magnetic conductor is essentially no different from a non-magnetic one. The test would instead require to be carried out within a magnetic fluid. Appendix 2 gives further background on these examples showing (a) the difficulties in finding a test capable of producing a conclusive result[9] and (b) in the case of ESR and the Hall effect, illustrating some of the difficulties with the available literature that have been discussed in Section 2.1 above.

3.1.2 Theoretical Grounds for the Lorentz Force in Magnetic Media

While the experimental evidence is hard to pin down, Wannier [50] has addressed the issue by means of calculations on beams of charged particles penetrating ferromagnetic materials. He uses **b** to denote the field actually responsible for the Lorentz force and concludes that, except under rather special circumstances, *e.g.* allowing for close range forces and quantum effects, **b = B**.

As discussed above, however, the conclusive theoretical arguments for both the origin and nature of the Lorentz force are more fundamentally based on the premise that all magnetic fields originate from moving charges or intrinsic circulating currents together with the theory of special relativity.

3.2 The Force Arising From Magnets

In spite of the relative simplicity of the description of the magnetic force as an interaction between directed current elements or moving charges, magnetic dipoles do exist and have to be dealt with. Their occurrence is commonplace in the fields of applied physics and electromagnetics, and only very rarely do we deal with the magnetic interactions between individual electric charges.

3.2.1 Dipole Interactions

A current I circulating around an infinitesimal loop of area dA has a magnetic moment[10] $\mathbf{m}$ given by

$$\mathbf{m} = I\hat{\mathbf{n}}dA \tag{16}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the plane of the loop and following the right-hand screw rule in relation to the direction of the positive current flow I.

While it is relatively straightforward to define the dipole moment thus, the interaction between two such infinitesimal dipoles of magnetic moment $\mathbf{m}_1$ and $\mathbf{m}_2$ is considerably more complicated than for the simple Coulomb interaction between individual charges or poles. The primary interaction is in the form of a torque $\mathbf{\Gamma}_{12}$ given by

$$\mathbf{\Gamma}_{12} = \frac{\mu_0}{r^3}\left[3(\mathbf{m}_1 \cdot \hat{\mathbf{r}})\mathbf{m}_2 \times \hat{\mathbf{r}} \quad - \quad \mathbf{m}_2 \times \mathbf{m}_1\right] \tag{17}$$

where $\hat{\mathbf{r}}$ is the unit vector from $\mathbf{m}_1$ to $\mathbf{m}_2$ and r is the distance in between. In addition, however, there is also a direct force $\mathbf{F}_{12}$ that is often overlooked. This is not a phenomenon that applies only to the

[9] The author would be grateful for any references to conclusive experimental evidence.

[10] IEC60050 most confusingly terms this the magnetic area moment (IEV 121-11-49) and defines the magnetic dipole moment as $\mu_0\mathbf{m}$ with symbol $\mathbf{j}$ (IEV 121-11-55).

macroscopic dipoles such as bar magnets that we are all familiar with, it applies equally to the infinitesimal dipole. It is often forgotten about because a single dipole experiences no net force in a uniform field, only a torque, but between two elementary dipoles there is a residual force because the dipole fields themselves are not uniform, and a dipole **m** will experience a net force **F** equal to $(\mathbf{m}\cdot\nabla)\mathbf{B}$ in a non-uniform field **B**.

Even neglecting the direct force, Equation (17) is not exactly amenable to further reduction or simple interpretation. Difficulties of this sort are no basis for arguing that the fundamental magnetic force is better expressed in the form of a Coulomb interaction between magnetic poles. In fact it is Equation (13), the microscopic equivalent of Ampere's force law, that provides the simplest basis for describing the elementary magnetic interaction. Ampere's force law itself, Equation (11), is equally satisfactory when dealing with currents rather than discrete charges.

The torque on a magnetic dipole in a uniform magnetic field **B** as given in Equation (2) follows directly from Equation (16) and the Lorentz force, Equation (3). This can easily be tested by considering a square loop of side a, with **B** in the plane of the loop and parallel to one of the sides. Two opposing sides of the loop will experience equal and opposite forces equal to BIa while the remaining two will experience no force at all, being parallel to **B**. The net torque is then $(BIa)\times a = Bm$, in agreement with Equation (2). This result, at least, is relatively simple.

3.2.2 Polar Versus Magnetic Dipoles

While the physical basis of the current dipole is simple and more rational than the polar model dipole, we cannot physically split a solenoidal dipole into two simpler equal and opposite parts. It is therefore often tempting to see if it is possible to make the polar dipole model work as a substitute for the current dipole model. The polar dipole moment $\mathbf{m}_p$ is given as per an electric dipole with $\mathbf{m}_p = pd\mathbf{l}$, where $\pm p$ are the pole strengths separated by $d\mathbf{l}$, and so we have to 'equate' this to $\mathbf{m}_I = I\hat{\mathbf{n}}dA$. If we did so, however, we would not be justified in writing Equation (2). This arises because the vector characters of $\mathbf{m}_I$ and $\mathbf{m}_p$ are different, the former being axial and the latter being polar. It would imply that the character of the torque, $\boldsymbol{\Gamma}$, was that of a real vector[11]. We could be tempted to write $\boldsymbol{\Gamma} = \mathbf{m}_p\times\mu_0\mathbf{H}$, but this is clearly not compatible with Equation (2) in general. Equation (17), on the other hand, is valid whatever the dipole character.

Figures 4a and 4b illustrate the difference between finite polar and current dipoles. The forms of the fields are given in several texts [11, pp. 253-255 and 271-272; 45 pp. 48-49, 205-210 and 267; 22, pp. 98-101 and 141-143]. The difference is in fact only discernable on a scale comparable to the size of the dipole itself, but it is a very significant difference in that the field directions in the centre of the dipole are reversed. It will be seen below that these inherent differences, including the vector character, are retained even when we consider an assembly of infinitesimal dipoles.

[11] Since poles are fictitious we could patch up this problem by assuming that a pole p would reverse its sign under inversion, but this would be simply adding to the artifice.

3.2.3 Properties of Dipole Fields

The properties of the 'force field' due to *infinitesimal* polar and current dipoles are indistinguishable in that by making the dipoles vanishingly small, we can only see the fields on a large scale as in Figure 4c.

Unfortunately, any discussion of the basic nature of these fields is somewhat limited without reference to two concepts from vector analysis, divergence and curl, with symbols $\nabla\cdot$ and $\nabla\times$ respectively. While they are not the easiest of concepts to grapple with, their names give a guide to their meaning. A divergent field emanates from a point, while a solenoidal field curls, or circulates, around an axis. There are analogies with fluid flow, with divergent flow being associated with a source (or sink) of fluid, while solenoidal behavior describes a vortex *i.e.* circulation of fluid around an axis. The picture of swirling bathwater disappearing down the plughole is therefore a combination of the two. Whirlpools at sea, however, are created by opposing currents, and do not have an associated source or divergence, and therefore the two concepts of divergence and curl appear quite independent.

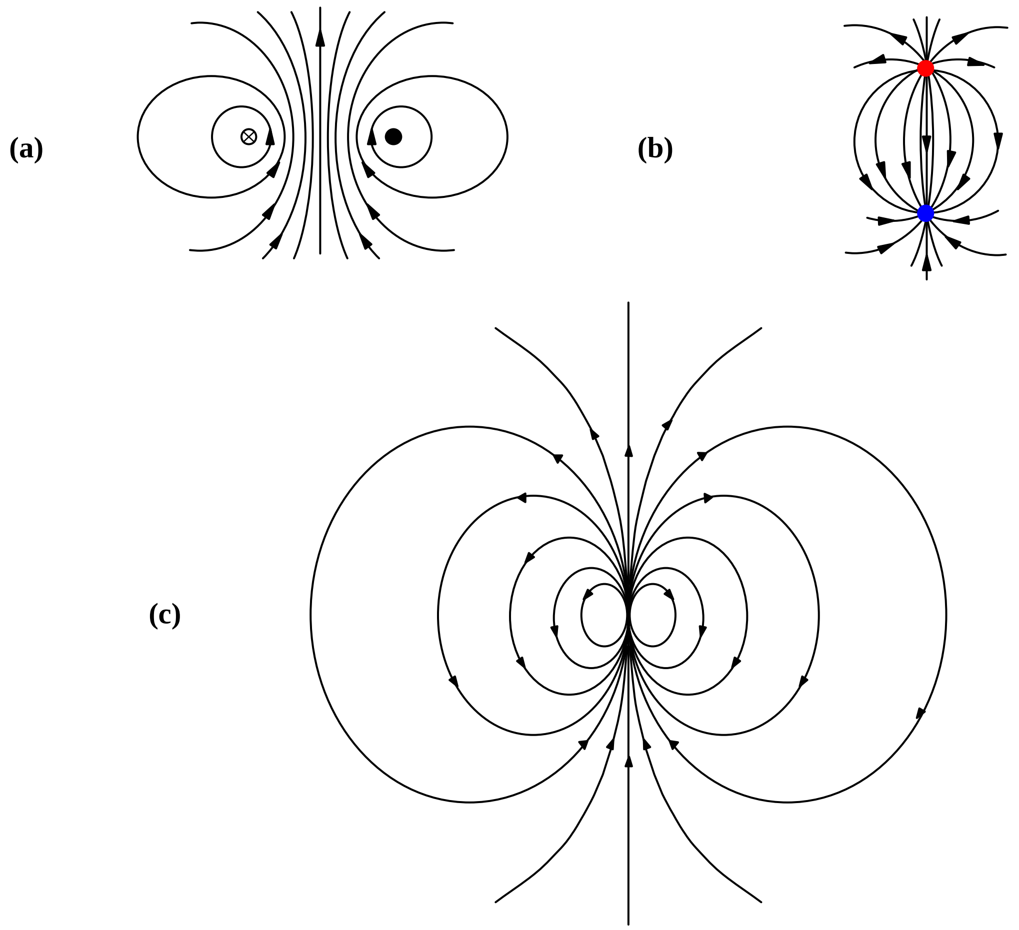


Figure 4 : A comparison of solenoidal and polar dipoles. (a) Local 'force field' for a current loop dipole. (b) Local 'force field' for a polar dipole. (c) Distant 'force field' for either current loop dipole or polar dipole. On the microscopic scale the characteristics of these two fields are are entirely different, while on the large scale the characteristics of both fields are are identical.

A field that is either divergent or solenoidal will often have both zero curl and zero divergence in a given region of space. In fact, the divergence or curl may be zero everywhere except at a single point. But the existence of that point determines the nature of the field around the point, as we will see in the examples below.

The field resulting from a current element is solenoidal, that is to say it has zero divergence everywhere but locally it has a non-zero curl everywhere, while the field due to a pole or charge is locally divergent and has zero curl. Solenoidal and divergent quantities are characteristically different and consequently there is no possibility of having two fields **b** and **h** which are *everywhere* equal while one is solenoidal and the other divergent. This is quite evident from Figure 5a where the elementary solenoidal field encircles its source, while in Figure 5b the divergent field emanates or converges directly from or to it, depending on the polarity of the source. The fields **b** and **h** are both mathematically and conceptually 'orthogonal'.

This discussion of curl and divergence leads to an apparent paradox. If we hold **B** to originate from currents and **H** from poles, then, how can they provide seemingly compatible alternative descriptions of the same phenomenon, given the completely opposite solenoidal and divergent characters of the fields that they generate? And on the other hand, if solenoidal dipoles are identical to polar dipoles on a large scale, why, fundamentally, can we not have a fully consistent physical description of magnetism in terms of poles which would be simpler to deal with mathematically?

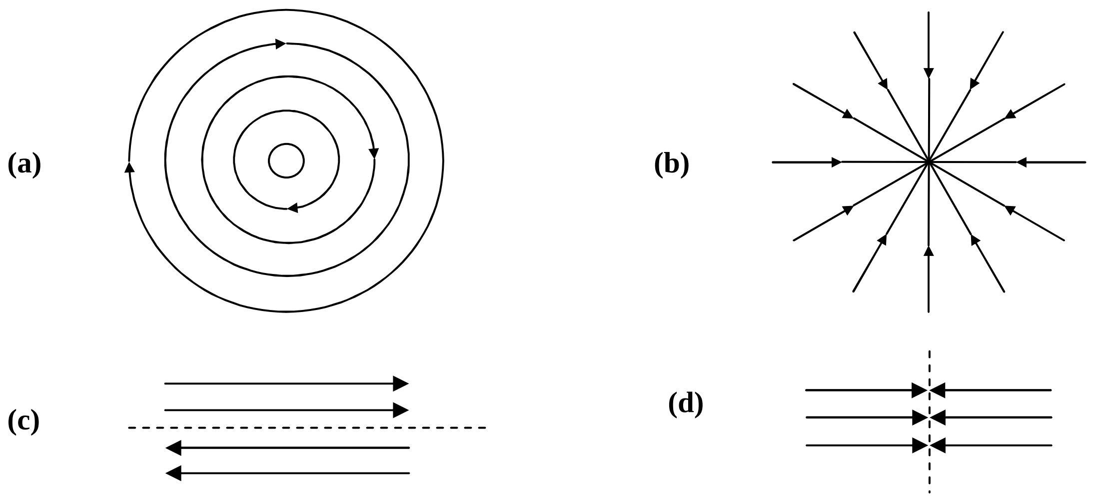
The answer is to these questions are relatively straightforward, if not immediately obvious. Given the large-scale dipole shown in Figure 4c, we cannot tell at all if it is solenoidal or polar in origin because both the curl and divergence of the field are zero everywhere except within the infinitesimal origin of the dipole itself. But as we shall now see, it is possible to tell the difference for *macroscopic* dipoles even on a large scale.

3.2.4 Incompatibility between Polar and Current-Loop Magnetic Dipoles

There are three types of magnetic dipole that could be involved in magnetic phenomena

- Polar
- Solenoidal
- Induced

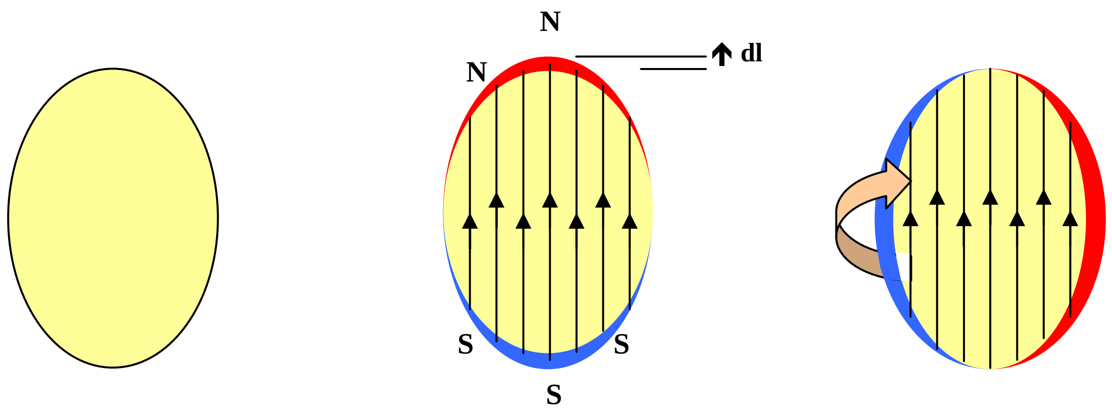
The first two types are permanent dipoles based on alternative descriptions, only one of which is conceptually correct. The third type is based on a secondary effect where there is ordinarily no net dipole moment at the atomic level until a magnetic field is applied. Our discussion here relates only to the first two types, and we only mention the third as it is the origin of so-called diamagnetism [14, pp. 134-135]. Induced magnetism will always tend to oppose the applied field, by Lenz's law, so that the associated magnetic susceptibility turns out negative. Nevertheless, such magnetic effects are still the equivalent of circulating currents so that for our discussion here, which has the objective of distinguishing between polar and solenoidal dipoles, we need only compare these two particular types.



		∇·	∇×
Solenoidal	(a), (c)	0	Nonzero somewhere
Divergent	(b), (d)	Nonzero somewhere	0

Figure 5 : A comparison of purely solenoidal and purely divergent fields.Diagram (a) is the solenoidal field for a line of current directed into the page, and diagram (b) is the divergent field for a negative charge or pole lying in the plane of the page. Diagrams (c) and (d) show the simplest of solenoidal and divergent fields having only a single nonzero component. Both such fields vanish within the plane out of the page indicated by the dotted line.

36



(a) unpolarized **(b) polar** **(c) solenoidal**

Figure 6 : A comparison of magnetization of a body due to polar and solenoidal sources. An ellipsoid, shown unpolarized in (a), has identical uniform magnetization shown by the arrows,, in (b) and (c), In (b) this magnetization is due to a surface pole distribution, while in (c) it is due to a surface current distribution. These two distributions are entirely different, even though the magnetization on the interior is the same.

37

We are all familiar with the result that two magnetic dipoles held at separate locations in free space will tend to line up antiparallel[12] to each other if they are free to rotate. Whether the dipoles are solenoidal or polar, the result is the same simply because the mathematical description we have for them makes no distinction between them except on an infinitesimal scale.

A single dipole placed within a continuum of polar dipoles, as in the historical description of a magnetized medium, will always align itself *antiparallel to the local magnetization*, *i.e.* antiparallel to the average dipole moment per unit volume where it is located.

Contrarily, a single dipole placed within a continuum of *solenoidal* dipoles, as in the modern description of a magnetized medium, will align itself *parallel to the magnetization*. The observation of this completely opposite behavior in practice would be, and is, of course, an effective qualitative demonstration that magnetization has its origin in circulating currents.

But, viewed this way, the paradox is why should there be any difference at all? If polar and current dipoles are essentially indistinguishable except in their infinitesimal detail, how can they bring about this fundamental and irreconcilable difference in the macroscopic behavior of a magnetized medium? To answer this we have to understand how the properties of an individual point dipole map onto a finite solid body.

While this discussion applies equally well to both poles and charges, for simplicity we will refer to poles alone. Taking first the infinitesimal polar dipole, it is formed by separating two equal but opposite poles, $+p$ and $-p$, by an infinitesimal distance $d\mathbf{l}$, with the dipole moment $\mathbf{m}$ being given effectively as in Section 3.2.2. Now, we can apply exactly the same concept to two equal and opposite pole *distributions* that are displaced by the infinitesimal distance $d\mathbf{l}$. In the straightforward case that the distributions are uniform everywhere inside a closed surface which defines the shape of the body in question and zero elsewhere, the net pole density within the body is zero prior to any such displacement, and in general this situation is unaffected by the displacement *except at any part of the surface that is not parallel to the displacement*. At any such point on the surface, the effect of the displacement will be to introduce a net surface pole density where the distribution has been displaced outside of the original surface of the body, for example, as in Figure 6b. The net surface pole density σ is given by

$$\begin{aligned}\sigma(\mathbf{r}) &= \Pi(\mathbf{r})\ \hat{\mathbf{n}}_s \cdot \hat{\mathbf{n}}_p dl \\ &= \mathbf{M} \cdot \hat{\mathbf{n}}_s\end{aligned} \tag{18}$$

where $\Pi(\mathbf{r})$ is the positive pole volume density within the body, the displacement is $d\mathbf{l} = \hat{\mathbf{n}}_p dl$, and ds is a given element of the surface whose unit normal directed out of the body is $\hat{\mathbf{n}}_s$. For simplicity we take $\Pi(\mathbf{r})$ to be constant within the interior of the body so that $\mathbf{M}$, the dipole moment per unit volume, is constant and equal to $\Pi d\mathbf{l}$.

---

[12] By antiparallel we shall mean geometrically parallel but directed in opposite orientations and, in the same context, we shall mean by parallel geometrically parallel and directed in the same orientation.

Because the net pole density is always zero within any region of the body, the dipole moment *within* the body has no discernible effect and the *surface* pole density created by the magnetization must account for *all* of its effects. Inside the body, the net 'force field' is therefore entirely determined by the surface pole distribution, so that we can leave the surface pole distributions in place, *and remove the body itself*. Everything in the interior will behave as if *in vacuo*, but under the influence of the induced surface pole (or charge) distributions. We cannot definitely say this, however, if the interior of the body has a non-uniform pole distribution, but such cases are more complex than we need consider here in elucidating the basic principles.

Given that the surface pole distribution induced by the magnetization determines everything we need to know for a body subjected to magnetization of this sort, it inevitably follows that the 'force field' in the *interior* of the body is directed in the opposite sense to the field emanating *from* the body. The reason for this is that when are near the positive pole distribution, say, the field will be directed away from it as shown in Figure 6a, irrespective of whether we are on the interior or exterior of the body. Similarly, if we are near the negative pole distribution, the field will be directed towards it.

Now let us turn to the solenoidal dipole which is represented by an infinitesimal current loop. As many textbooks show, a configuration of identical touching current loops lying in a plane behaves identically to one single loop of current passing around the perimeter of the configuration. This is an illustration of Stokes' theorem [22, p. 9; 11, p. 258-259; 9, pp. 237-238]. Therefore, if we take a uniformly polarized body comprising solenoidal dipoles and divide it into plane sections orthogonal to the direction of magnetization, the total magnetization of the body is represented by the collection of the currents traveling around the perimeters of all such sections. This in effect represents a net current lying entirely within the surface of the body and circulating around the axis of magnetization. The current is given in terms of a surface density **K** Am^{-1} traversing any line element that both lies within the surface and is coplanar with the axis of magnetization.

If we let each infinitesimal current loop in a plane have a magnetic moment $d\mathbf{m}$, then according to Equation (16) the current I circulating around a loop of elemental area $d\mathbf{A}$ is given from $d\mathbf{m}=Id\mathbf{A}$. As we argued earlier, all of the loop currents lying entirely within each plane section cancel by virtue of touching an identical adjacent loop, but around the edge of the plane section, *i.e.* at surface of the body, there is no touching loop to provide a canceling contribution, and so the net current traveling around the edge of the plane is simply I, the dipole current itself. If the plane we have taken as a section has thickness dz, then I/dz gives the linear current density, K, along the edge. However, we can also relate $d\mathbf{m}$ to the magnetization per unit volume, **M**, since the volume occupied by each dipole is $d\mathbf{A}dz$, and so

$$\begin{aligned} d\mathbf{m} &= Md\mathbf{A}dz \\ &= Id\mathbf{A} \\ \Rightarrow M &= \frac{I}{dz} = K \end{aligned}$$

We therefore have K equal to M or, rather more precisely,

$$\mathbf{K} = \mathbf{M} \times \hat{\mathbf{n}}_s \tag{19}$$

Similarly to the situation with polar dipoles, where we may remove the entire body and leave the polarization to be represented by the surface charge alone, we can remove the entire body and leave the surface current to represent the magnetization, Figure 6c.

While the surface current density provides a fully alternative description to a surface pole density, the two descriptions are not physically equivalent. Note in particular that while one, **K**, is essentially vector in character the other, σ, is scalar, and that their magnitudes have entirely different distributions over the surface, as we see from Figures 6b and Figure 6c.. The poles tend to concentrate where the axis of magnetization cuts the surface, whereas the current density concentrates around the equatorial plane. We might try to argue that these differences, though significant, tend to be mathematical in origin, but what we cannot escape from is that the 'force field' that exists inside a body magnetized with current dipoles must be in the same direction as the magnetization itself, *and that is completely opposed to the situation that holds within a body magnetized with polar dipoles*.

This is demonstrated in Figure 7. To convince ourselves of the point, we must follow it through from first principles and, for the moment, without thinking of *B* and *H*. We must think only about the actual 'force field', whatever it may turn out to be called, and we can deduce as much as we need to know about this field simply from the way that the dipoles tend to align. As for a test dipole, we choose a current loop simply for consistency with the modern view, but we already know that the choice will not affect the outcome because individual elemental dipoles have indistinguishable properties. They have identical external 'force fields' and we cannot detect the differences in their interior. However, in Figures 7b and 7c we have chosen different constituent dipoles for the magnetized body itself. The corresponding surface pole and surface current distributions would be as in Figures 6b and 6c.

In Figure 7b, the internal polar dipoles must be aligned 'up' in order to be consistent with the required magnetization and to have external field consistent with the orientation of the external test dipole. If the 'force field' $\mathfrak{F}$ emanating from the surface of the body is 'up', then on the other side, the interior, it must be 'down', as we have already discussed in relation to Figure 6b.

In Figure 7c, where we have a body of current dipoles, the effective surface current must be anti-clockwise (as viewed from the top) in order to be consistent with the required magnetization and to have an external field consistent with the orientation of the external test dipole. Remembering that parallel current elements attract whereas antiparallel ones repel, the current in the test dipole loop will tend to be parallel to the current around the perimeter of the plane cross-section in question. On the interior this must also be the case so that here the test dipole must now be 'up', and consequently the 'force field' in the interior of the body must also be 'up', just the opposite of the polar magnetized body.

This is the truly fundamental point. It tells us that, identical though the properties of the two different types of individual dipoles may seem, the 'force field' inside a magnetized medium - which is just an ensemble of such dipoles - is represented by Figure 8a, which describes a 'force field' due to circulating currents, and not as in Figure 8b which describes a 'force field' due to magnetic poles. Mathematically, there is a subtle difference between the two types of dipole such that when we

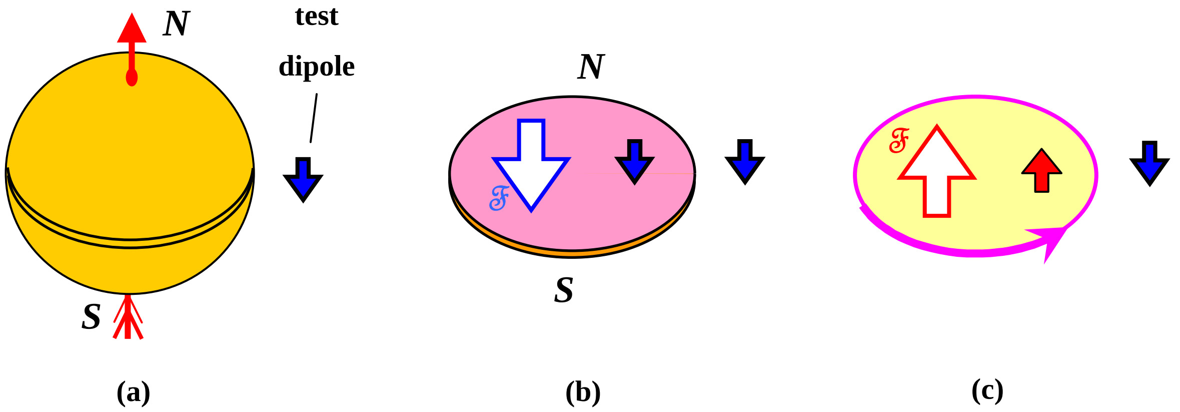


Figure 7 : The alignment of magnetic test dipoles inside and outside a magnetized body. In (a) we see the entire body, magnetised as shown. The external test dipole aligns antiparallel with the bulk magnetisation. Within a plane cross section of the body, as at (b), we assume polar magnetisation and the internal 'force field' $\mathcal{F}$ is 'down', with the test dipole seeking the South pole, while at (c) $\mathcal{F}$ is 'up' because the test dipole aligns itself so as to keep its current parallel to the effective magnetisation current circulating around the edge of the plane.

proceed from an individual dipole to a large-scale ensemble or continuum of dipoles, their original incompatible solenoidal and divergent characters, which are effectively suppressed by the process of making them infinitesimal, simply re-emerge. This arises because those distinct properties, which, although conveniently hidden by the smallness of scale, are still there, and they become transferred to the ensemble or continuum as though the result were simply a scaled-up version of the infinitesimal dipole, albeit possessing the actual geometrical detail of the ensemble. Finally, it is interesting to compare the three bulk pictures 8(a-c) with their microscopic counterparts 7(a-c).

It is an important point, however, that the behavior discussed above has been reasoned from the basic properties of the relevant types of magnetic dipoles in free space so that there can be no ambiguity resulting from the use of either **B** or **H**. We replaced the magnetized body by its equivalent pole or circulating current distribution, so we reasoned from fundamental free-space principles, not macroscopic ones. This therefore provides the crucial physical underpinning of the phenomenological description – the magnetic 'force field' is entirely solenoidal, never divergent. Consequently there are no real poles. It is not sufficient to state - as is almost invariably done – that this follows from $\nabla\cdot\mathbf{B} = 0$ alone. As this condition only negates the existence of free poles, it leaves open the possibility that there are intrinsic magnetic dipoles of the polar rather than solenoidal type. We must also state that the 'force field' is associated with **B** rather than **H**. If we did allow that the 'force field' could alternatively derive from **H**, then since we may have $\nabla\cdot\mathbf{H} \neq 0$, the converse of the standard argument for **B** would apply, so that poles *would* exist. **H** is only a 'force field' if poles exist, poles do not exist, ergo **H** is not a 'force field'.

Consider further the implications of the mode of interaction between dipoles. If an individual dipole tends to line up with the macroscopic magnetization, then this tends to *increase* the field rather than *decrease* it. This is the opposite behavior to an electric dipole in a polarized medium where the dipole alignment tends to reduce the field, as in the macroscopic form of Coulomb's law. This explains the essential difference between the SI and emu forms of Equation (4), since the force between magnetic poles tends to increase in the presence of a magnetic medium rather than be decreased by it. On these grounds alone, if we must introduce poles as an artifact, then only the MKSA form is valid. Moreover, to be consistent, it is actually the emu form that requires to be fixed up by an artificial concept of induced poles, which is no more than an apology for a problem which endures only for historical reasons.

In a 'polar' description, the equivalent distribution of bound charge is given by $\nabla\cdot\mathbf{P}$. Analogously, the equivalent distribution for poles is therefore given by $\nabla\cdot\mathbf{M}$. In the circulating current description, however, $\nabla\times\mathbf{M}$ is the equivalent circulating current responsible for magnetization. Note that we cannot describe the same magnetic field as originating from both $\nabla\times\mathbf{M}$ and $\nabla\cdot\mathbf{M}$ together; it has to be the one or the other. On the face of it, however, it is quite hard to see how such different quantities such as $\nabla\times\mathbf{M}$ and $\nabla\cdot\mathbf{M}$ can give the same field and we shall return to this problem later.

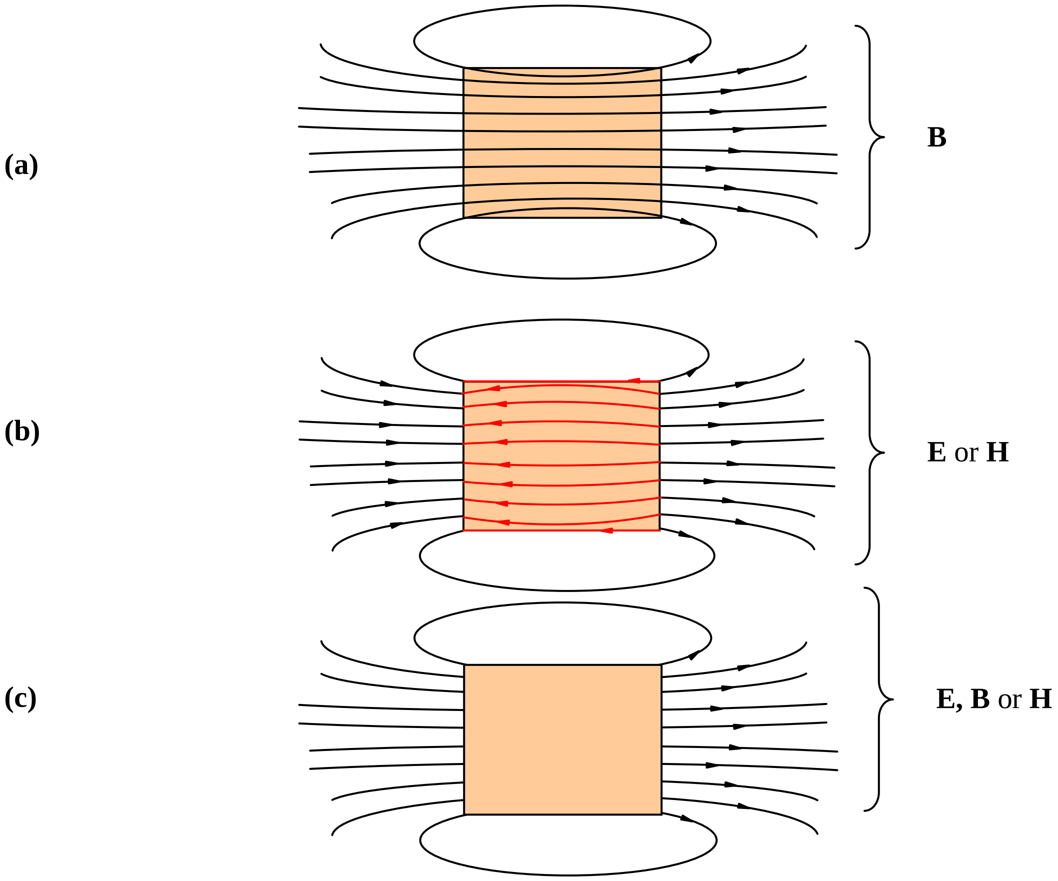


Figure 8. A comparison of the possible 'force fields' associated with a magnetized body. (a) The 'force field' inside and outside a sample polarized throughout by solenoidal dipoles. The field is definitely solenoidal in origin as the sample could be replaced by a current loop around its exterior. (b) The 'force field' inside and outside a sample polarized throughout by polar dipoles. The field is definitely divergent since the sample could be replaced by a pair of oppositely charged plates. (c) The 'force field' outside a sample of polarized or magnetized material. There is nothing to reveal whether the field is polar (divergent) or solenoidal in origin.

4 Maxwell's Equations

No discussion of electromagnetic theory can be complete without a discussion of Maxwell's equations. Maxwell's equations are so significant that they are sometimes seen as the starting point from which all else that is of interest follows. Even Stratton, to whom we so frequently refer, takes them as a postulate, introducing them on page 1 of his classic work [9, pp. 1-6]. But the standard form of the equations we see and use today is not quite as Maxwell first set them down [23, pp. 231-232]. As a result of refinements initiated by Heaviside [10, vol. 2, pp. 1-23], just four of his original eight sets of equations [2, pp. 480-486] are now taken as a basis, while his force equation equivalent to Equation (3) was later associated with Lorentz

$$\begin{aligned} &\nabla \cdot \mathbf{D} = \rho_{free} \\ &\nabla \cdot \mathbf{B} = 0 \\ &\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\ &\nabla \times \mathbf{H} = \mathbf{J}_{free} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned} \qquad (20)$$

These go together with the historical form of the constitutive relations, as given in Equations (5) above, and the Lorentz force, Equation (3), as a basis for describing a vast range of electromagnetic phenomena. But, in formulating the equations as a description of macroscopic phenomena by including **D** and **H** rather than **E** and **B** alone, the true set of microscopic or free-space equations is obscured. While substitution of $\varepsilon_0\mathbf{E}$ for **D** is an obvious step back towards the free-space equations, we 40

have an ambiguity with **B** and **H**, which we will now explore. If Maxwell's equations were entirely 42
fundamental, no such ambiguity should exist.

4.1 Free-Space versus Macroscopic Equations

By free-space we mean that the bulk polarization **P** and magnetization **M** are both taken as being zero. Any bound charge or magnetic current involved with matter has to be included explicitly as sources along with all the free charges and currents. Maxwell's equations can then be written in either of two forms, one dependent on **E** and **H** while the other is dependent on **E** and **B**. Conceptually, one and only one of the two forms is valid. First of all, in free space we must select only two fields, one electric and the other magnetic, as a third and fourth would be redundant. Provided that we accept that the proper choice for the electric field is **E**, against which there is little argument, this restricts our choice to just those two forms.

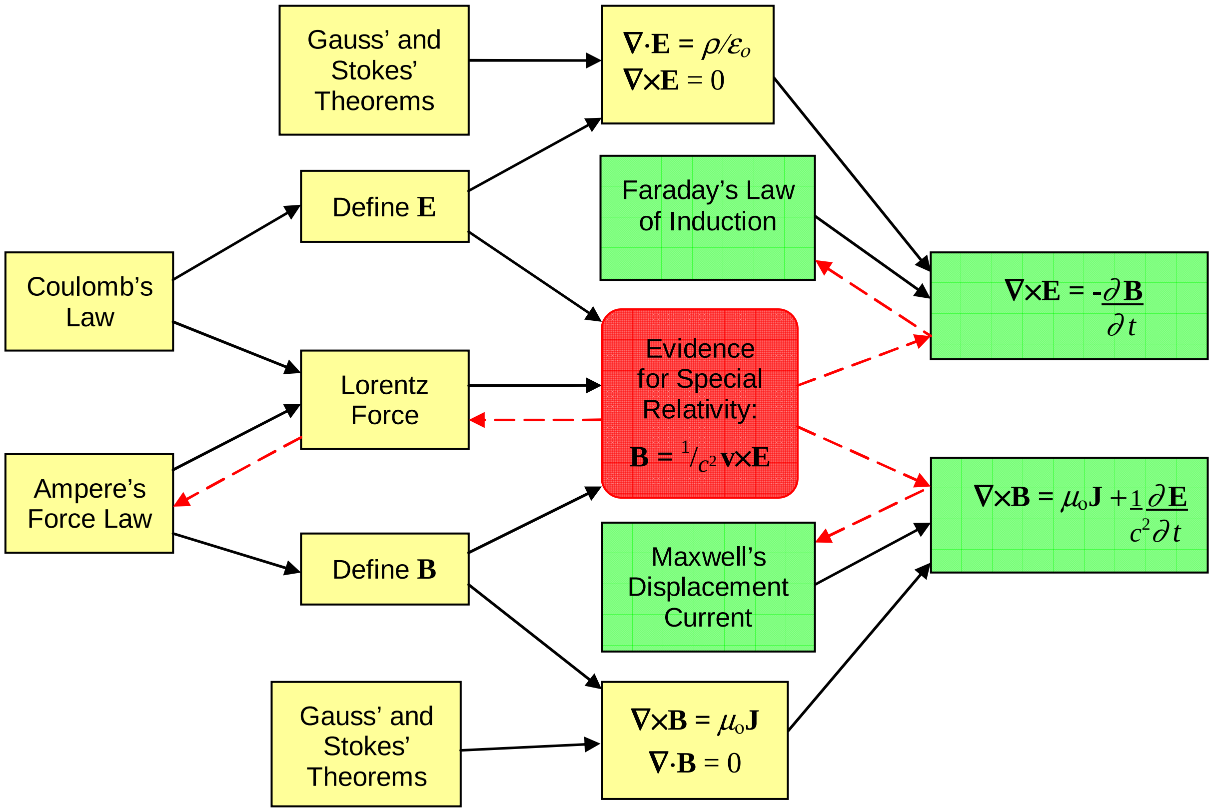


Figure 9a : The development of electromagnetic theory from basic principles. In the static cases, shown in the yellow, the starting points are Coulomb's and Ampere's laws. The Lorentz force can be argued as evidence for Einstein's theory of special relativity. Once that theory is independently established, the Lorentz force and the existence of the magnetic field itself can be derived. Crossing from statics to dynamics, shown in the green boxes, special relativity can be invoked to provide the time-dependent equations we know as Maxwell's equations, confirming both Faraday's and Maxwell's contributions. H and D are not essential and are required only in order to more readily describe fields in matter.

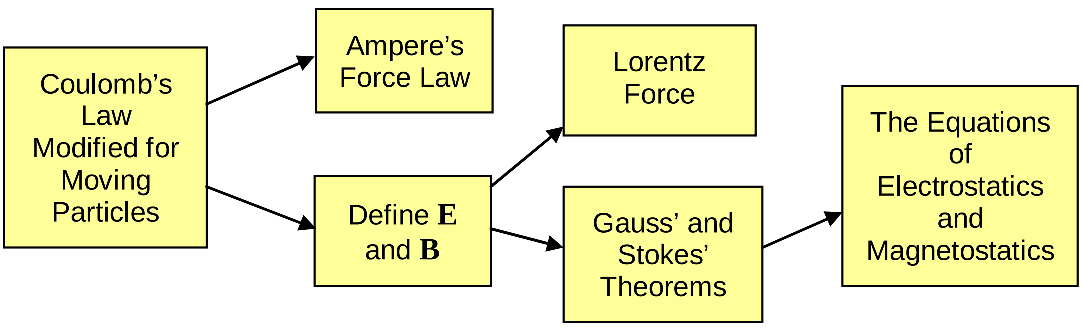


Figure 9b. A more elementary approach based on Coulomb's law modified as in Equation (13) allows the deduction of all the fundamentals of electrostatics and magnetostatics in free space. Ampere's force law is implicit, and, rather than arising separately, E and B are abstracted simultaneously leading to directly to the Lorentz force. As before H and D are not essential but can be added later, making their auxiliary roles clear.

4.1.1 Free-Space Maxwell's Equations in Terms of E and H

In this first arrangement we eliminate **D** by writing instead $\varepsilon_0\mathbf{E}$, since **P** = **0**. But for **B** we refer to the constitutive relations, Equation (5) or (6), and deliberately write $\mu_0(\mathbf{H}+\mathbf{M})$ in its place. The reason for this is that with Maxwell's equations as given in Equation (20), there could otherwise be no source for an intrinsic magnetic field since **J** is to be considered 'free' - a pure conduction current. With $\nabla\cdot\mathbf{B} = 0$, therefore, we must have $\nabla\cdot\mathbf{H} = -\nabla\cdot\mathbf{M}$. The required magnetic source is provided by the divergence of the macroscopic quantity -**M**. We can therefore define a *microscopic* quantity Π, a 'pole density', as the appropriate source. But we must apply constraints as to which functions are permissible for a pole density, for example, by requiring $\int \Pi dV = 0$ over of any region space that is taken sufficiently small but still finite. In addition to the pole density, Π, we must also have a pole current density, **Y**, as the pole distribution may move. In analogy with moving charges for which we have $\mathbf{J} = \rho\mathbf{v}$, we can write $\mathbf{Y} = \Pi\mathbf{v}$. Since $\Pi = -\nabla\cdot\mathbf{M}$, we must then have $\mathbf{Y} = \frac{\partial \mathbf{M}}{\partial t}$ from the continuity equation, assuming that this applies equally well for magnetic poles as it does for electric charges.

Consequently, we have for the free-space equations in this representation a formulation that appears to be consistent with the polar description of magnetization:

$$
\begin{aligned}
&\nabla\cdot\mathbf{E} = \frac{\rho}{\varepsilon_0}\\
&\nabla\cdot\mathbf{H} = \Pi\\
&\nabla\times\mathbf{E} + \frac{\partial \mu_0\mathbf{H}}{\partial t} = -\mu_0\mathbf{Y}\\
&\nabla\times\mathbf{H} - \frac{\partial \varepsilon_0\mathbf{E}}{\partial t} = \mathbf{J}
\end{aligned}
\qquad (21)
$$

4.1.2 Free-Space Maxwell's Equations in Terms of E and B

The other form we may take is

$$
\begin{aligned}
&\nabla\cdot\mathbf{E} = \frac{\rho}{\varepsilon_0}\\
&\nabla\cdot\mathbf{B} = 0\\
&\nabla\times\mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0\\
&\nabla\times\frac{\mathbf{B}}{\mu_0} - \frac{\partial \varepsilon_0\mathbf{E}}{\partial t} = \mathbf{J}_{free} + \mathbf{J}_{mag}
\end{aligned}
\qquad (22)
$$

Here we are only required to accept that all magnetism arises from currents alone. We must therefore include a term $\mathbf{J}_{mag}$ even in the free-space equations in order to represent any intrinsic or induced magnetic dipoles that arise from spin rather normal conduction currents. We have therefore written **J** as $\mathbf{J}_{free}+\mathbf{J}_{mag}$ in order to make the distinction between the conduction current and the intrinsic magnetic current. While $\mathbf{J}_{free}$ will be a function of both **E** and **B**, $\mathbf{J}_{mag}$ differs in that it will be entirely independent of **E**. Note that while it is true that an orbital electron can give rise to both polarisability and magnetization, in any basic phenomenological representation these effects would be seen as being separate.

To accept the first form of the free-space equations, Equations (21), is to adopt a model of magnetism that includes poles and, as we have seen in the previous section, effectively implies that **H** is the force field. In accepting the second form, Equations (22), we are recognizing that all magnetism, as Ampere was first to believe, arises from currents alone. It would save a lot of trouble if Equations (22), were given the recognition they deserve by referring to them as the free-space Maxwell's equations, and teaching them as such before launching in to the conventional macroscopic form as in Equations (20). With the free-space form as the starting point, a microscopic description of electrical and magnetic polarization can then be brought in to derive, as for example Scharf demonstrates [16, pp. 151-157], the full macroscopic form of Maxwell's equations by proceeding along the lines of the Lorentz-Lorenz treatment referred to in Section 1.2 above. The equations so derived are no other than the set of Maxwell's phenomenological equations that we are so familiar with, including the constitutive relations. The physical basis for the phenomenological theory of electrodynamics is therefore held within the free-space Maxwell's equations together with the Lorentz-Lorenz development of the macroscopic equations. Buchwald [51] presents a detailed account of the conceptual difficulties faced by the original Maxwellian macroscopic theorists in understanding the true nature of the interaction between electromagnetism and matter, and how the microscopic view eventually took hold as the electron theory progressed and the contentions were gradually resolved. Today, of course, with unavoidable hindsight, we think little of it. The macroscopic and microscopic pictures are seen to fit so obviously together that it is perhaps difficult to appreciate the quarter of a century of debate and counter debate that was involved in the transition.

4.2 Microscopic Form of Maxwell's Equations

There is another approach to Maxwell's macroscopic equations [38, pp. 427-428; 16, p. 151]. Here the phenomenological description of matter in terms of polarization **P** and magnetization **M** is dropped resulting in a description purely in terms of the fundamental fields **E** and **B** together with all source charges and currents, as in the free-space Maxwell's equations, Equations, (22). In doing so, however, we can still identify the source terms related to the presence of matter with appropriate functions of **P** and **M**. In principle, we can carry out a similar exercise using **E** and **H** as the basis fields, as in 4.1.1 above, in order to see how a pole based theory would look.

4.2.1 Maxwell's Equations in Terms of E and B Alone

In order to proceed we first of all note that

- Electrical charge comprises *all* charge, both bound and free
- The polarization **P** arises from the bound charge, ρ_{bound}, acting as source
- Motion of the bound charge constitutes a true current, $\mathbf{J}_{bound} = \dfrac{\partial \mathbf{P}}{\partial t}$

• The magnetization **M** arises from the implied circulating current $\mathbf{J}_{mag}$ acting as source [9, p. 242].

Turning then to Maxwell's original equations, Equations (20), only the first and fourth of need attention. In the first, if we wish to replace **D** with $\varepsilon_0\mathbf{E}+\mathbf{P}$ then we must simply replace ρ_{free} with *all* charges present, both bound and free, ρ_{total} . In the fourth equation, we need to do four separate things:

• replace **D** with $\varepsilon_0\mathbf{E}+\mathbf{P}$, as in the first equation,

• represent $\dfrac{\partial \mathbf{P}}{\partial t}$ by $\mathbf{J}_{bound}$,

• replace **H** with $\dfrac{\mathbf{B}}{\mu_0}-\mathbf{M}$,

• represent $\nabla\cdot\mathbf{M}$ by $\mathbf{J}_{mag}$.

Only the last of these four steps needs any further justification. If we have a magnetic field due to a real circulating current, $\mathbf{J}_{free}$, then $\nabla\times\mathbf{H} = \mathbf{J}_{free}$, and therefore, as in the free-space equations, Equations(22), we are simply ascribing the origin of any form of magnetization *that is not due to a conduction current*, to an equivalent current, $\mathbf{J}_{mag}$. In any case, $\mathbf{J}_{mag} = \nabla\times\mathbf{M}$ is simply the more general form of $\mathbf{K} = \mathbf{M}\times\hat{\mathbf{n}}_s$ which applies at the surface of a uniformly magnetized sample, as in Equation (19) above.

The resulting form of the four equations is, as we should expect, identical to Equations (20) except that the source terms ρ and **J** must now refer to *all* charges and currents, designated by ρ_{total} and $\mathbf{J}_{total}$. But in the process we have defined the association between certain of these quantities and the presence of matter which we more usually describe through the polarization **P** and magnetization **M**. We therefore have

$$\begin{aligned}
&\nabla\cdot(\varepsilon_0\mathbf{E}) = \rho_{total}\\
&\nabla\cdot\mathbf{B} = 0\\
&\nabla\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t} = 0\\
&\nabla\times\left(\frac{\mathbf{B}}{\mu_0}\right) - \frac{\partial(\varepsilon_0\mathbf{E})}{\partial t} = \mathbf{J}_{total}
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
\rho_{total} &= \rho_{free} + \rho_{cond} + \rho_{bound}\\
\rho_{bound} &= \nabla\cdot\mathbf{P}\\
\mathbf{J}_{total} &= \mathbf{J}_{free} + \mathbf{J}_{bound} + \mathbf{J}_{mag}\\
\mathbf{J}_{bound} &= \frac{\partial\mathbf{P}}{\partial t}\\
\mathbf{J}_{mag} &= \nabla\times\mathbf{M}
\end{aligned} \tag{24}$$

Scharf [16, p.151] refers to this particular set of equations as the 'Microscopic Equations', although he does not attempt to make the distinction between $\mathbf{J}_{free}$ and $\mathbf{J}_{mag}$, taking both together as simply being **J**. These equations are, as should be expected, identical with the free-space Maxwell's equations, Equations (22), except that the charge and current density quantities are now specifically labeled as 'total' to remind us to include these in all their forms, bound and free. We could therefore have recast the free-space equations into these forms merely by appropriately representing all the sources involved.

We may refer to Equations (22) and (23) alike as the free-space or microscopic equations. The free-space context can be recovered trivially from Equations (24) by taking **P** to be zero, although we must still allow $\mathbf{J}_{mag}$ to remain[13] but only for the purpose of including individual intrinsic magnetic dipoles such as electrons, atoms *etc*. Details of the terms in Equations (24) that are required to account for macroscopic effects are further explained in Table 3.

4.2.2 Maxwell's Equations in Terms of E and H Alone

Contrarily, if we were to choose **E** and **H** as the basis fields, we could equally have an entirely different form to the equations in Equations (23) and (24). Following a similar process we now find

$$\begin{aligned}
&\nabla\cdot\left(\varepsilon_0\mathbf{E}\right)=\rho_{total}\\
&\nabla\cdot(\mu_0\mathbf{H})=\Pi\\
&\nabla\times\mathbf{E}+\frac{\partial(\mu_0\mathbf{H})}{\partial t}=-\mu_0\mathbf{Y}\\
&\nabla\times\mathbf{H}-\frac{\partial\left(\varepsilon_0\mathbf{E}\right)}{\partial t}=\mathbf{J}_{free}+\mathbf{J}_{bound}
\end{aligned}\tag{25}$$

where

$$\begin{aligned}
&\Pi=-\nabla\cdot\mathbf{M}\\
&\mathbf{Y}=\frac{\partial\mathbf{M}}{\partial t}
\end{aligned}\tag{26}$$

As mathematical models, there is no argument in favor of either the set of Equations (23-24) or Equations (25-26). In the one set we have electrical charges and currents alone, with magnetism being related to a circulating current through $\nabla\times\mathbf{M}=\mathbf{J}_{mag}$, while in the other we have the presence of both magnetic poles, identified by $\Pi = -\nabla\cdot\mathbf{M}$, and magnetic currents due to their motion, $\mathbf{Y}=\frac{\partial\mathbf{M}}{\partial t}$ (note that magnetic poles so defined obey the same conservation law as does electric charge and since **Y** is intrinsically a bound quantity we need not identify it with a subscript).

Burris [31, p. 9.2] actually introduces Equations (25) together with pole density and pole current without further qualification, as though they were the basic entities. Within the definition of the Lorentz force given in this same relatively recent article, we have, again without further qualification, $\mu_0\mathbf{H}$ rather than **B**. The attempt to have everything in terms of **E** and **H** rather than **E** and **B** may suit the purposes of some authors, but this can only add to the general confusion surrounding **H** and **B**,

[13] In such a case, $\mathbf{M}dV$ would represent the dipole moment **m**, where dV is an arbitrary infinitesimal volume occupied by the dipole. $\mathbf{J}_{mag}$ then has to be expressed in terms of $\nabla\times\mathbf{M}$ by means of the calculus of generalised functions.

Table 3: A summary of electrical and magnetic source terms and their meanings

Source	*Meaning*	*Key Characteristics*
ρ, ρ_{total}	Electric charge density of all charges present	$\rho = \rho_{free} + \rho_{cond} + \rho_{bound}$
ρ_{free}	Free charge density	Charge that is isolated or that is surplus to an otherwise neutral body. Need not average to zero locally in space and may support direct current, DC, depending on whether free to move or fixed.
ρ_{cond}	Charge density giving rise to conduction	Charge that is available to flow in the presence of an electric field within a conducting medium.
ρ_{bound}	Bound charge density giving rise to polarization within a material body	Under the influence of an electric field, the relative displacement of the positive and negative components of a bound charge distribution gives rise to a polarization, **P**. Averages to zero locally in space and gives rise only to AC currents.
J, $\mathbf{J}_{total}$	Total current density	$\mathbf{J} = \rho\mathbf{v} + \mathbf{J}_{mag}$. At any point in space, the product of the net electrical charge density and its velocity. In general this will be a function of both **B** and **E**. We must also include $\mathbf{J}_{mag}$ (see below) in order to account for magnetism.
$\mathbf{J}_{free}$	Conduction current, due to motion of unbound charges ρ_{free} or ρ_{cond}	$\mathbf{J}_{free} = \rho_{free}\mathbf{v}_{free}$ or $\rho_{fcond}\mathbf{v}_{cond}$ as appropriate. The movement of bound charges is excluded. The summation is over all available free charge types. In general $\mathbf{J}_{free}$ will be a function of both **B** and **E**.
$\mathbf{J}_{bound}$	Current due to motion of ρ_{bound}	$\mathbf{J}_{bound} = \frac{\partial \mathbf{P}}{\partial t} = \rho_{bound}\mathbf{v}_{bound}$. Only bound charges are considered and the current is due to their relative motion alone. The summation is over all bound charge types. In general $\mathbf{J}_{bound}$ will be a function of both **B** and **E**.
$\mathbf{J}_{mag}$	Current giving rise to intrinsic magnetism	$\mathbf{J}_{mag} = \nabla \times \mathbf{M}$. $\mathbf{J}_{mag}$ is due to currents or spin at the atomic level. in general it is fixed or a function of **B**, but is independent of **E**.
Π	Pole density	$\Pi = -\nabla \cdot \mathbf{M}$, a fictitious source for an intrinsic magnetic field.
Y	Pole current	$\mathbf{Y} = \frac{\partial \mathbf{M}}{\partial t}$, a fictitious current of poles accounting for time varying magnetization

particularly in a reference work. At the very least authors who use **E** and **H** as the basis fields, with or without poles and pole currents, should give adequate cautions so as to prevent any such confusion.

As pointed out in Section 3.2.4 above, $\nabla \cdot \mathbf{B} = 0$ alone does not imply the non-existence of poles, rather it merely implies the non-existence of free poles. The non-existence of poles, therefore, does not arise out of Maxwell's equations, which could readily include poles as we have shown, but rather from the observed nature of the magnetic 'force field', which sides with the solenoidal field **B** as its origin rather than the divergent field **H**.

Magnetic poles do not exist, either as free monopoles or dipoles, and so the concepts of pole density Π and pole current **Y** are simply artificial. Furthermore, as they only stand for the macroscopic terms -$\nabla\cdot\mathbf{M}$ and $\dfrac{\partial \mathbf{M}}{\partial t}$ they are purely auxiliary. In addition, we have the identification of **E** and **B** as the two field quantities involved in the electromagnetic force, both microscopically and macroscopically. Finally, the free-space form of Maxwell's equations, Equations (22), identifies a specific relationship between **B** and **E** alone, while we cannot achieve the same for **E** and **H** without reverting to poles, Equations (21). All of these points argue towards Equations (23-24) rather than Equations (25-26), as being the physical underpinning of Maxwell's equations as we generally know them. 44

4.3 The Standard Form of Maxwell's Equations

Maxwell's equations as we generally know them in the form of Equations (20) lead us back to the free-space equations, Equations (28), provided that we:

- remember the basic quantities are **E** and **B**,
- convert **D** to $\varepsilon_0\mathbf{E} + \mathbf{P}$ and **H** to $\mathbf{B}/\mu_0 - \mathbf{M}$ using the constitutive relations, Equations (6),
- set to **M** to zero everywhere, except in the case of any individual intrinsic magnetic dipoles we want to keep as 'free' quantities[14], in which case we associate $\nabla\times\mathbf{M}$ with the magnetic current, $\mathbf{J}_{mag}$, describing such dipoles,
- set **P** to zero everywhere. Since a pair of free charges can account for any free point-dipole, it is not necessary to retain $\nabla\cdot\mathbf{P}$ as a local source of bound charge.

In turn, the free-space equations will lead us back to any form of the macroscopic equations that we choose, based either on a simple phenomenological approach or a detailed microscopic approach. What advantage, however, is their in choosing the standard form of Maxwell's equations? Often, when we have to solve them, the first part of the process is to find a way of eliminating two of the four variables in order to get, say, an equation in **E**, an equation in **H** and a final one linking both **E** and **H**. The quantities **D**, **B**, **P** and **M** will follow based on the original assumptions, *i.e.* the form taken by the constitutive relations in the given scenario. We may follow the same basic process starting instead with **E** and **B**, but the result will be the same. Even with Maxwell's equations in the form of Equations (23) or (25), where we have already reduced them to the unknowns **E** and **B** (or **E** and **H**) alone, we still have to resort to the supporting equations, Equations (24) or (26) in order to bring in the properties of the media involved.

[14] For example, on the atomic scale, when describing the interaction between the electromagnetic field and an individual particle possessing a net magnetic dipole moment. See also Section 4.2.1 and the footnote there.

Maxwell's equations in their standard form, however, do bring the supporting equations into play, but this is achieved by taking $\nabla\cdot\mathbf{P}$ and $\nabla\times\mathbf{M}$, the source terms for polarization and magnetization, together with $\nabla\cdot\mathbf{E}$ and $\nabla\times\mathbf{B}$ in the first and last equations to leave us with field terms only in $\nabla\cdot\mathbf{D}$ and $\nabla\times\mathbf{H}$. We can only interact explicitly with 'free' entities, be they charges, currents or magnetic dipoles. All 'bound' quantities are involved only implicitly through the characteristics of the media concerned. In that sense, the 'bound' quantities require to be either solved for or eliminated rather than specified, and so in dealing with macroscopic media it is more convenient to use **D** and **H** rather than **E** and **B** in the first and fourth equations. The standard form of Maxwell's equations, Equations (20), is therefore *essentially macroscopic* in character, whereas Equations (23) still retains the free-space form and requires to be taken together with Equations (24) in order to deal with macroscopic media. Equations (24) are therefore the bridge between the free-space and macroscopic forms. In the light of our discussion in the preceding sections concerning alternative views of **H** as arising from either poles or currents, it must be clear that the role of **H** here is strictly associated with currents alone and, in particular, only *free current*.

Finally, we may mention that when the standard form of Maxwell's equations is quoted in the form

$$\begin{aligned}\nabla\cdot\mathbf{B}&=0\\ \nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}&=0\\ \nabla\cdot\mathbf{D}&=\rho_{free}\\ \nabla\times\mathbf{H}-\frac{\partial\mathbf{D}}{\partial t}&=\mathbf{J}_{free}\end{aligned}\tag{27}$$

the result is a pair of homogeneous equations based on the microscopic quantities **E** and **B** alone, together with a pair of inhomogeneous equations based on **D** and **H** alone and with ρ_{free} and $\mathbf{J}_{free}$ as the source terms. The pair of homogeneous equations do not involve any charges, currents or magnetic sources, and therefore come over from the free-space form to the macroscopic unchanged. On the other hand, with the inhomogeneous equations **E** and **B** must be replaced with their macroscopic counterparts **D** and **H** so as to specifically to account for all charges, currents and magnetic sources. This form is by no means universal, but it is used for example in [52; 45, pp. 304; 9, p. 6; 15, p. 1; 53-55]. Any particular reason for preferring it goes unmentioned, but it is perhaps because this form carries over directly over into the relativistic formulation [15, footnote to p 1.] but it does also make it a little easier both to remember the rationale behind the uses of all four of the field quantities and to identify the free-space from the macroscopic quantities.

5 Special Relativity

While special relativity is at the very heart of modern electromagnetic theory, it is often considered too advanced a topic for inclusion in an introductory discussion of the fundamental theory. Indeed, in spite of its compact formalism, or rather because of it, the relativistic formulation of electromagnetic theory does not lend itself to easy understanding and it is generally only undertaken as part of an advanced mathematical physics course. But, as we have seen, the link to relativity is obvious through the existence of the magnetic field itself. First, it gives a full basis for classical magnetic phenomena without the need for artifices such as poles. Secondly, even if we were able to argue for some alternative non-relativistic theory, we would be left with the difficulty of explaining how the forces between electrical charges in uniform motion could contrive to break Newton’s third law, which they may do by Equations (11) and (13). It seems a pity to leave any discussion of the deep impact of this fundamental link simply to a mention in the passing. To demonstrate how deep the link is, let us review the basic structure of electromagnetic theory based on what we have established so far.

It has already been mentioned that Coulomb’s law is incomplete in as much as it provides a purely static description of the interaction of two charges. Coulomb’s law as applied to moving charges as seen through the theory of special relativity results in not just in a single modification, such as Equation (13), but in the *complete description of electrodynamics including the Lorentz force and all* *of Maxwell’s equations*. This is illustrated in Figure 9a. In the diagram, the starting points of the classical microscopic theory are Coulomb’s and Ampere’s force laws, Equations (1a) and (11) above, which lead to the definition of the electric and magnetic fields **E** and **B**, thence to the basic equations of electro- and magneto- statics, $\nabla\times\mathbf{E} = \mathbf{0}$, $\nabla\cdot\mathbf{E} = \rho/\varepsilon_0$, $\nabla\times\mathbf{B} = \mu_0\mathbf{J}$ and $\nabla\cdot\mathbf{B} = 0$, and to the Lorentz force, $\mathbf{F}=q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$. The Lorentz force, as we have seen, can be argued as evidence for Einstein’s Theory of special relativity. That theory having been independently established, however, the Lorentz force and the existence of the magnetic field itself can then be derived given only a purely electric field. Crossing from statics to dynamics, shown in the green boxes, special relativity can again be invoked to show that the time independent equations $\nabla\times\mathbf{E} = \mathbf{0}$ and $\nabla\times\mathbf{B} = \mu_0\mathbf{J}$, which we may suppose hold in a given rest frame, must transform over to the established time dependent form as per Maxwell’s equations within a moving frame. This provides a theoretical basis confirming both Faraday’s and Maxwell’s contributions to the theory. As before, **H** and **D** are required only in order to describe fields in matter.

A rigorous treatment of electromagnetics within the theory of special relativity is covered in detail by several textbooks [22, pp. 380-388; 9, pp. 78-80; 47, pp. 486-495; 56]. Here we only attempt to use the simplest results to justify the assertions made above. To this end we show that

- a force on a test charge placed between electrically charged plates will be seen to vary depending on the velocity of the observer. This paradox shows that the laws of physics based on Newtonian relativity are incomplete,
- this apparent paradox is resolved by applying special relativity to the observed electric force alone,
- consequently the magnetic force provides evidence for special relativity while conversely special relativity provides the exact form of the magnetic force,
- the time dependent contributions to Maxwell’s equations can be deduced from special relativity (Faraday's law of induction and Maxwell’s displacement current).

In addition, we examine how the theory of special relativity imposes the condition on the free-space electromagnetic wave equation that the wave velocity must be independent of the frame of reference of the observer, given that these wave equations are just like any other.

5.1 Transformation of the Electric Field

Figure 10 shows an infinite parallel plate capacitor having uniform charge distributions $+\sigma$ and $-\sigma$ (Cm^{-1}) on each plate. We take our frame of reference as being the 'rest frame', in which we, the observers, and the charge distributions are stationary. There is no current, and Coulomb's law alone applies so that the electric field is given by $\mathbf{E} = \sigma/\varepsilon_0 \hat{\mathbf{z}}$ and a test charge q lying between the plates will undergo a force $\mathbf{F} = q\dfrac{\sigma}{\varepsilon_0}\hat{\mathbf{z}}$.

5.1.1 Demonstration that the Force Depends on the Observer's Velocity

Within a frame of reference moving with a velocity $\mathbf{v} = v\hat{\mathbf{x}}$ with respect to the rest frame, however, the test charge and both of the surface charge distributions now have a velocity $-\mathbf{v}$, and therefore surface current densities $\mathbf{K} = \pm\sigma\mathbf{v}$ (Am^{-1}) are observed, the negative sign being for the positively charged plate. This gives rise to a magnetic field $\mathbf{B} = \mu_0 K\hat{\mathbf{y}}$ between the plates in this frame of reference, as the currents on each plate are equal in magnitude but oppositely directed. The fact that we can make **B** come and go with the frame of reference should be surprising, but of course we are already aware that **B** must be treated as an 'observational' effect dependent on **E** itself rather than on an entirely separate magnetic source.

The current seen in the moving frame results in a net force **F** on the test charge q given by the Lorentz force involving both **E** and **B** as found above: 46

48

$$\begin{aligned}\mathbf{F} &= q(E\hat{\mathbf{z}} + (-v\hat{\mathbf{x}})\times B\hat{\mathbf{y}})\\ &= q\left(\frac{\sigma}{\varepsilon_0}\hat{\mathbf{z}} - \mu_0 K v\hat{\mathbf{z}}\right)\\ &= q\frac{\sigma}{\varepsilon_0}\hat{\mathbf{z}}\left(1 - \varepsilon_0\mu_0 v\frac{K}{\sigma}\right)\\ &= q\mathbf{E}\left(1 - \frac{v^2}{c^2}\right)\end{aligned} \tag{28}$$

The net force acting on the test charge therefore appears less than it was in the rest frame due to the factor of $1-v^2/c^2$. This paradox of classical electromagnetism can only be resolved by accepting the theory of special relativity. While the theory of special relativity has many conflicts with the classical description and offers the prospect of numerous paradoxes within its own framework, paradoxes are rare within the closed confines a purely classical description, particularly when velocities nowhere near the speed of light are involved. This is one of those rare exceptions, and as we have mentioned earlier, it is so obvious as to go almost unnoticed. It is certainly mentioned by few, if any, authors as compelling basic evidence for the theory of special relativity, even if they do at some point fully acknowledge that magnetism arises out of special relativity!

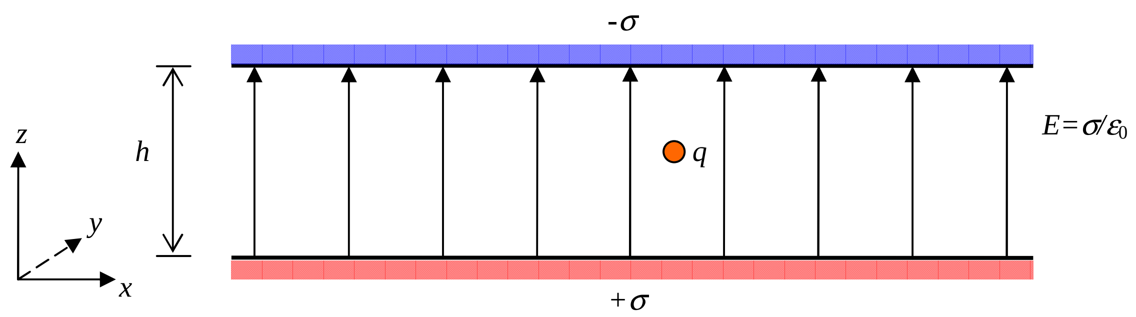


Figure 10a. The field within a moving charged parallel plate capacitor. In the rest frame, there is a static charge distribution σ on the bottom plate of the capacitor, and an equal and opposite static charge distribution on the top plate. An electric field E between the plates results, and there is no magnetic field B.

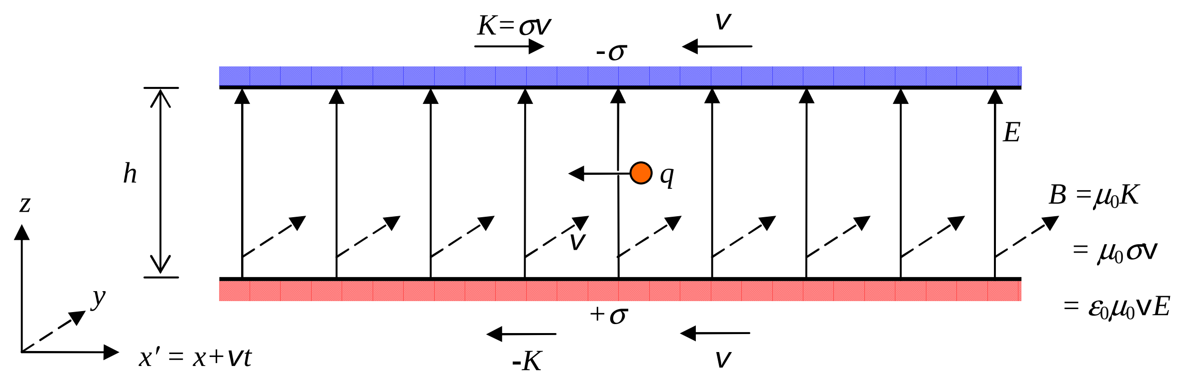


Figure 10b. The field within a moving charged parallel plate capacitor. Viewed from a frame of reference moving to the right with velocity v, the charge distributions appear to move with velocity $-v$, *i.e.* to the left. Therefore in this frame of reference we must have a magnetic field B due to the linear current densities K=±σv seen on each plate. This proves that the magnetic field is purely an observational effect, an implication of relativity. B is orthogonal to E and to the direction of motion, and its magnitude is simply $(v^2/c^2)E$. The effect of the motion on the attractive force between the capacitor plates is to reduce it, since opposite currents repel, and again, this must be an observational effect as all we have done is to change to a moving frame of reference, the force in the original frame of reference is unchanged. These conclusions can be reached with no *a priori* knowledge of special relativity.

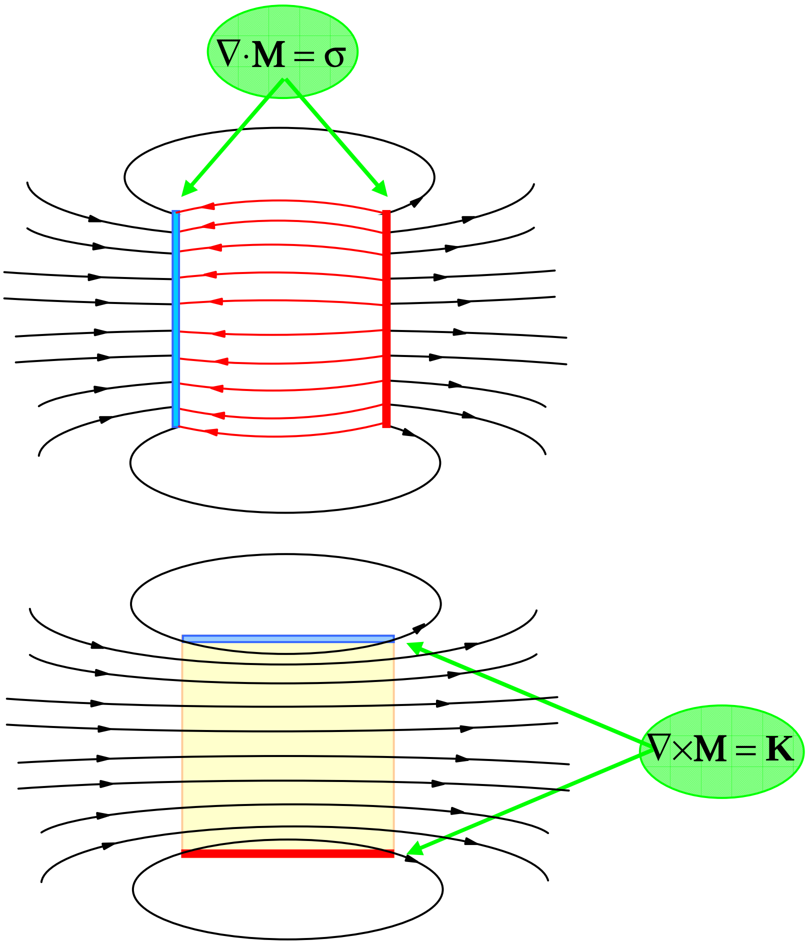


Figure 11 : Current and pole distributions producing identical external fields. A uniformly magnetized cylinder is taken to be magnetized, in one case with a surface current distribution K, and in the other with a pole distribution *σ*. The 'force field' for a polar description is shown in (a) based on two parallel discs (the pole faces) bearing equal and opposite uniform pole distributions. In (b), however, we see the 'force field' resulting from the surface cylinder carrying a uniform sheet of current such that a linear current density K circulates around the polar axis. The external fields are identical, implying that if we specify the one distribution then we are automatically specifying the other. But as they are defined on completely independent faces of the cylinder, this seems counter-intuitive. While the internal fields appear quite different they are trivially related through the curl and divergence of M.

5.1.2 Demonstration that Special Relativity Accounts for Magnetic Force

One of the main results in the field of special relativity is the Lorentz Transform which shows how the spatial and time co-ordinates (*x*, *y*, *x*, *t*) in one frame of reference map over into another frame of reference moving with a relative velocity **v** with respect to it. Almost every elementary text on special relativity, for example [57, pp.59-64; 58, p. 34; 11, pp. 375-378; 7, pp. 140-141], deals with the Lorentz transform and many show how it can be derived from three basic principles: the speed of light is a constant independent of the reference frame, invariance of measurement under pure translations, and the principle of relativity itself, that is to say, if we transform from frame A to A′ using velocity **v** as a parameter, then the very same transform with the parameter –**v** will take us from A′ back to A. We can express this as

$\boldsymbol{L_v}$ $(x, y, x, t) \rightarrow (x', y', x', t')$,

$\boldsymbol{L_{-v}}$ $(x', y', x', t') \rightarrow (x, y, x, t)$

The Lorentz transform $\boldsymbol{L_v}$ and its inverse $\boldsymbol{L_{-v}}$ are then given by

$\boldsymbol{L_v}$	$\boldsymbol{L_{-v}}$
$x' = \gamma(x - vt)$	$x = \gamma(x' + vt')$
$y' = y$	$y = y'$
$z' = z$	$z = z'$
$t' = \gamma(t - \frac{v}{c^2} x)$	$t = \gamma(t' + \frac{v}{c^2} x')$

where $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$ (29)

For simplicity we have assumed that relative motion between the primed and unprimed frames takes place along the *x*-direction with velocity $\mathbf{v} = v\hat{\mathbf{x}}$ as in our example above.

One immediate conclusion of the Lorentz Transform is that the charge density seen in the primed frame is $\gamma\sigma$ (not σ/γ!) because in the unprimed frame a charge q distributed over an area $dx \cdot dy$ will be seen in the primed frame as being distributed over an area $dx' \cdot dy' = (dx/\gamma) \cdot dy$. This change of scaling, known as the Lorentz contraction, comes about because in evaluating the differential dx', both x *and* $x+dx$ must be taken at the same instant t' within the moving frame, rather than the same instant in the rest frame. The quantity of charge itself cannot be affected and consequently we must have $\sigma' = \gamma\sigma$ and $E_z' = \gamma E_z$.

Besides length, forces are also observed differently in the primed frame. The definition of force as rate of change of momentum still holds valid in relativistic mechanics, but the transformation between the forces observed in different reference frames is rather complex because it involves the transformation of the observed momentum **p** as well as that of the observed time t [57, pp. 178-180]. But for a force, $F_z\hat{\mathbf{z}}$ say, that is acting perpendicular to the motion, $v\hat{\mathbf{x}}$, of the moving reference frame, the component of momentum involved is unaltered, $p'_z = p_z$ [57, pp. 178-180]. The transformation of $\frac{d\mathbf{p}}{dt}$ associated with the force $F_z\hat{\mathbf{z}}$ therefore depends only on the transformation of time. The net result for the specific situation in which the accelerated body is stationary in the rest frame is $F'_z = F_z/\gamma$, so that for a given force acting on the test charge in the rest frame, the force observed in the primed frame is reduced.

Taking the Coulomb force acting on the stationary test charge in the rest frame on its own, the force as seen in the primed frame would therefore be evaluated as

$$
\begin{aligned}
\mathbf{F}'_c &= \frac{1}{\gamma}(qE)\hat{\mathbf{z}} \\
&= \frac{1}{\gamma}\left(q\frac{E'}{\gamma}\right)\hat{\mathbf{z}} \\
&= \frac{1}{\gamma^2}(qE')\hat{\mathbf{z}} \\
&= \left(1-\frac{v^2}{c^2}\right)(qE')\hat{\mathbf{z}}
\end{aligned} \tag{30}
$$

This result is entirely in agreement with the earlier result of Equation (28) that was obtained by evaluating the effects of the magnetic field and Lorentz force, thereby resolving the apparent paradox. The undeniable conclusion is that transforming the pure Coulomb force seen in the rest frame to a moving frame results in a force which *agrees with the full Lorentz force acting on the test charge as seen from within that frame*. The classical result obtained by bringing in the magnetic field as an observational artifact together with the Lorentz force completely agrees with the Coulomb force, taken on its own, as long as the observational transformations of the theory of special relativity are applied. This simple 1-dimensional analysis carries over into a complete 3-d treatment. The only caution is regarding the literature on special relativity itself, in that it is quite common to see **H** used in place of **B** within the expression for the Lorentz force [44; 58, p. 42], but this is perhaps a syntactic error rather than a semantic error on the part of the authors, as is no doubt quite clear by now.

5.2 Faraday's Law of Induction and Maxwell's Displacement Current

The preceding discussion takes us from Coulomb's law via special relativity to the magnetic field and the Lorentz force, but it does not take us beyond electrostatics and magnetostatics. For a description of electrodynamics we require to bring in the explicitly time dependent terms in Maxwell's equations through Faraday's law of induction and Maxwell's displacement current. Are these phenomena original physical principles or are they also simply artifacts of special relativity?

In the introduction to Section 5, we supposed a scenario where the time dependent contributions to Maxwell's equations are unknown. Apart from simply introducing time dependence into the equations, these contributions are responsible for the coupling between the electric and magnetic fields *in a way that reaches back even to static fields.* That is to say, without them, electricity and magnetism are two entirely separate effects, whereas with them they become completely interdependent. As to the assertion that this also applies to statics, there is a semantic issue here as to what we mean by static. The term 'static' must be taken to include uniform motion since what is completely at rest in one reference frame may be in uniform motion when viewed in another equivalent reference frame. Consequently, a charge that is static within one reference frame may give rise to a current element in another. Since we can describe this as $\rho \rightarrow \mathbf{J}$, we can likewise write $\nabla\cdot\mathbf{E} \rightarrow \nabla\times\mathbf{B}$, that is to say the electric field in one frame of reference can give rise to a magnetic field in another, and *vice versa*.

In Sections 5.1.1 and 5.1.2 we saw that the connection between static electric and magnetic fields did indeed come about through special relativity. We now show how two of the equations of electrostatics and magnetostatics, $\nabla\times\mathbf{E} = \mathbf{0}$ and $\nabla\times\mathbf{B} = \mu_0\mathbf{J}$, need to be modified, so that in doing so that they become what we recognize as the two of Maxwell's four equations that include the time dependent contributions, $-\dfrac{\partial\mathbf{B}}{\partial t}$ and $\dfrac{\partial(\varepsilon_0\mathbf{E})}{\partial t}$

Our assumed initial scenario is that we are oblivious to any coupling between the electric and magnetic fields or any time dependent terms in the equations, and that it is sufficient to write $\mathbf{E}(\mathbf{r},t) \to \mathbf{E}(\mathbf{r}',t')$ and $\mathbf{B}(\mathbf{r},t) \to \mathbf{B}(\mathbf{r}',t')$ under any transformation of reference frame $\boldsymbol{L}_v(\mathbf{r},t) \to (\mathbf{r}',t')$. We are simply assuming this to be true so that we may consider the consequences.

If we start with $\nabla\times\mathbf{E}$ in the rest frame and apply the Lorentz Transform, we find $\nabla'\times\mathbf{E}$ in the moving frame by application of the general transformation for differentials, which follows directly from differentiating Equation (29):

$$\begin{aligned}
\frac{\partial}{\partial x'} &= \gamma\frac{\partial}{\partial x} + \frac{\gamma v}{c^2}\frac{\partial}{\partial t} & \qquad \frac{\partial}{\partial x} &= \gamma\frac{\partial}{\partial x'} - \frac{\gamma v}{c^2}\frac{\partial}{\partial t'} \\
\frac{\partial}{\partial y'} &= \frac{\partial}{\partial y} & \frac{\partial}{\partial y} &= \frac{\partial}{\partial y'} \\
\frac{\partial}{\partial z'} &= \frac{\partial}{\partial z} & \frac{\partial}{\partial z} &= \frac{\partial}{\partial z'} \\
\frac{\partial}{\partial t'} &= \gamma v\frac{\partial}{\partial x} + \gamma\frac{\partial}{\partial t} & \frac{\partial}{\partial t} &= -\gamma v\frac{\partial}{\partial x'} + \gamma\frac{\partial}{\partial t'}
\end{aligned} \tag{31}$$

We then find

$$\begin{aligned}
\nabla'\times\mathbf{E} &= \nabla\times\mathbf{E} + (\gamma-1)\frac{\partial}{\partial x}\left[E_y\hat{\mathbf{z}} - E_z\hat{\mathbf{y}}\right] + \frac{\gamma v}{c^2}\frac{\partial}{\partial t}\left[E_y\hat{\mathbf{z}} - E_z\hat{\mathbf{y}}\right] \\
&= \nabla\times\mathbf{E} + \frac{\partial}{\partial t}\left(\frac{\mathbf{v}}{c^2}\times\mathbf{E}\right) + O\left(\frac{v^2}{c^2}\right) \\
&= \nabla\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t}
\end{aligned} \tag{32}$$

It is also to be understood that on the left side of the Equation (32) $\mathbf{E}$ is a function of (x', y', x', t') while on the right both $\mathbf{E}$ and $\mathbf{B}$ are functions of (x, y, x, t). We again identify $\dfrac{\mathbf{v}}{c^2}\times\mathbf{E}$ with $\mathbf{B}$ from Equation 14b, and we neglect the terms of order v^2/c^2 that are effectively unmeasurable at ordinary velocities. We must therefore conclude that if $\nabla'\times\mathbf{E} = \mathbf{0}$ pertains in a given reference frame then $\nabla\times\mathbf{E} + \dfrac{\partial\mathbf{B}}{\partial t} = \mathbf{0}$ pertains in another. This then is the origin of electromagnetic induction - it is indeed a consequence of special relativity.

It would appear that Maxwell's displacement current would follow from applying the same procedure to $\nabla'\times\mathbf{B}$, but this leads only to higher order terms. However it must be borne in mind that $\mathbf{B}$ itself should be regarded as the lowest order relativistic correction to $\mathbf{E}$, and it is again to $\mathbf{E}$ that we should look as the starting point. This time we must ask how $\dfrac{\partial\mathbf{E}}{\partial t}$ transforms:

$$
\begin{aligned}
\mu_0 \frac{\partial \varepsilon_0 \mathbf{E}}{\partial t'} &= \frac{1}{c^2}\left(\gamma v \frac{\partial \mathbf{E}}{\partial x} + \gamma \frac{\partial \mathbf{E}}{\partial t} \right) \\
&= \gamma\left(\frac{1}{c^2}(\mathbf{v}\nabla\cdot\mathbf{E}) - \nabla\times(\frac{\mathbf{v}}{c^2}\times\mathbf{E}) \right) + \frac{\gamma}{c^2}\frac{\partial \mathbf{E}}{\partial t} \\
&= (\mu_0\mathbf{J} - \nabla\times\mathbf{B}) + \mu_0 \frac{\partial \varepsilon_0 \mathbf{E}}{\partial t} + O\left(\frac{v^2}{c^2}\right)
\end{aligned} \tag{33}
$$

Here, as previously, it is to be understood that on the left side of Equation (33), **E** is a function of (x', y', z', t') while on the right both **E** and **B** are functions of (x, y, z, t). Given v_y=v_z=0, it is possible 49

to write $v\frac{\partial \mathbf{E}}{\partial x}$ in the form $(\mathbf{v}\cdot\nabla)\mathbf{E}$, which by a standard identity is equal to $\mathbf{v}(\nabla\cdot\mathbf{E}) - \nabla\times(\mathbf{v}\times\mathbf{E})$.

Finally, we again identify $\frac{\mathbf{v}}{c^2}\times\mathbf{E}$ with the magnetic field **B.**

Now if all charge ρ is stationary in the primed reference frame then there is no change in the electric field with time so that $\mu_0 \frac{\partial \varepsilon_0 \mathbf{E}}{\partial t'} = \mathbf{0}$, while in the unprimed rest frame the charge distribution must appear to have a uniform velocity $+v$ in the x-direction, giving rise to the current density $\mathbf{J} = \rho\mathbf{v} = \varepsilon_0 \mathbf{v}\nabla\cdot\mathbf{E}$. We associate the current density **J** with a magnetic field **B** originating from the now familiar relativistic correction to the electric field, $\frac{\mathbf{v}}{c^2}\times\mathbf{E}$. In addition we have the term $\mu_0 \frac{\partial \varepsilon_0 \mathbf{E}}{\partial t}$, which will not in general be zero because at a given point the electric field will be changing as a result of the moving charge distribution, as in the case of a moving point charge. In the primed frame we have a purely electrostatic scenario, whereas in the rest frame we have $\nabla\times\mathbf{B} = \mu_0\left(\mathbf{J} + \frac{\partial \varepsilon_0 \mathbf{E}}{\partial t}\right)$, again neglecting terms of order v^2/c^2 . This, then, is the origin of the displacement current[15].

While we have achieved these results through a scenario that is only valid for $v \ll c$, they turn out to be valid in general and fully consistent with special relativity. The time dependent forms of Maxwell's equations that are revealed, however, may in turn be used to show the exact form of the coupling between the electric and magnetic fields under a change of reference frame [22, pp. 380-382; 9, 78-80]:

$$
\begin{aligned}
\mathbf{E}'_{//}(\mathbf{r}', t') &= \mathbf{E}_{//}(\mathbf{r}, t) & B'_{//}(\mathbf{r}', t') &= B_{//}(\mathbf{r}, t) \\
\mathbf{E}'_{\perp}(\mathbf{r}', t') &= \gamma(\mathbf{E}_{\perp}(\mathbf{r}, t) + \mathbf{v}\times\mathbf{B}(\mathbf{r}, t)) & \mathbf{B}'_{\perp}(\mathbf{r}', t') &= \gamma\left(\mathbf{B}_{\perp}(\mathbf{r}, t) - \frac{1}{c^2}\mathbf{v}\times\mathbf{E}(\mathbf{r}, t)\right)
\end{aligned} \tag{34}
$$

---

[15] Since the displacement current may also be inferred simply by taking the divergence of Maxwell's fourth equation and requiring the conservation of charge, $\frac{\partial \rho}{\partial t} = -\nabla\cdot\mathbf{J}$ one may ask why appeal to special relativity? But, going back a step, relativity itself is based on simple and fundamental tenets such as translational invariance and conservation laws. While both methods of establishing the existence of the displacement current are equally valid, we have chosen this one specifically to illustrate the same connexion with special relativity as with Faraday's law of induction.

where the subscripts // and ⊥ denote respectively parallel to **v** and perpendicular to **v**, and the factor γ can be ignored except at extreme velocities. Note that only the perpendicular parts of **E** and **B** are coupled by the change of reference frame. Also, the transformed electric field has a strong similarity to the field part of the Lorentz force (see 5.3 below).

Returning to Figure 9a, we see that special relativity is indeed the bridge between electrostatics and electrodynamics. As with Ampere's force law, there is no need to adopt Faraday's law of induction and Maxwell's displacement current as separate postulates. Coulomb's law and the theory of special relativity are sufficient.

5.3 Lorentz Force

In Section 5.2 we saw that under a change of reference frame the transformation of an electric field, $\mathbf{E}'$, bears a strong resemblance to the field part of the Lorentz force, $\mathbf{E} + \mathbf{v}\times\mathbf{B}$. Conceptually, we can hardly escape from the fact that the Lorentz force must be identical with the Coulomb force, the only difference being the velocity of the charge under observation. In the rest frame of this charge, only the Coulomb force is ever experienced. Einstein drew attention to this in his seminal 1905 paper [59], the relevant part being:

1. If a unit electric point charge is in motion in an electromagnetic field, there acts 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]…
2. 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 co-ordinates at rest relatively to the electrical charge. [new manner of expression].

Accordingly, the equation $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$ is none other than $\mathbf{F}' = q\mathbf{E}'$ within the appropriate frame of reference. The term $(\mathbf{E} + \mathbf{v}\times\mathbf{B})$ is simply *the electric field* $\mathbf{E}'$ *perceived by the charge q within its own rest frame*.

The form of the Lorentz force is correct for any particle velocity and no approximation is involved. According to Jackson [22, p. 191], it has been experimentally verified to very large velocities. The question that hangs, however, is what happens inside a magnetic medium? As mentioned in Section 3.1.2 above, the Lorentz force is still expected to hold good except in special cases where there may be close range forces or quantum effects to consider.

5.4 The Speed of Light as a Universal Constant

That the speed of light *in vacuo* is a constant independent of any reference frame was a revolutionary concept in its time and is held to be the primary evidence for special relativity. But the significance of this is more than purely conceptual and its effects are not simply confined to the peculiar domain of relativistic physics, for it has an impact on the solution of Maxwell's equations for propagating waves.

The most general and simplest form of wave equation is

$$\nabla^2 U = \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} \qquad (35)$$

This equation, applicable to all components of **E** and **B** simultaneously, also applies to other linear phenomena such as the propagation of sound waves. But we know that electromagnetic waves and sound waves are fundamentally different in that the apparent velocity of sound waves is fixed with respect to the reference frame in which the underlying medium appears to be stationary, and consequently the observed wave velocity depends on the motion of the observer, while on the other hand the velocity of electromagnetic waves appears the same in any reference frame. How can both systems obey the same equation?

The answer lies once more in the Lorentz transform. Using Equation (31) to find the transformation of the double partial derivatives with respect to *x* and *t*, it can be readily verified that the wave equation, Equation (35), is obeyed in all reference frames undergoing uniform motion with respect to the original rest frame [22, pp. 352-353 and 378]. We can therefore apply the Lorentz transform to any solution of Equation (35) and be confident that it will still be a valid solution within the new reference frame. A plane wave solution to Equation (35) propagating in the *x*-direction with frequency ω and wavenumber k has a phase at any point along the *x*-axis given by $\phi(x,t) = \omega t - kx$, where $\omega/k = c$. Applying the Lorentz transform to *x* and *t*, the phase observed in a frame moving along the x-axis with velocity v with respect to the rest frame will be given by:

$$\begin{aligned}
\phi(x',t') &= \gamma(\omega - k\mathrm{v})t' - \gamma\left(k - \frac{\omega\mathrm{v}}{c^2}\right)x' \\
&= \omega' t' - k' x' \\
\textit{where} & \\
\omega' &= \gamma(\omega\text{-}k\mathrm{v}) \\
k' &= \gamma(k\text{-}\omega\mathrm{v}/c^2) \\
\omega'/k' &= \frac{\gamma(\omega\text{-}k\mathrm{v})}{\gamma(k\text{-}\omega\mathrm{v}/c^2)} = c
\end{aligned} \qquad (36)$$

Both the frequency ω' and wavenumber k' are different in the primed frame, while the phase velocity, given by ω'/k', is unchanged. On the one hand the frequency is Doppler-shifted down by a factor of $\sqrt{\frac{1-\mathrm{v}/c}{1+\mathrm{v}/c}}$, whereas on the other the wavelength is observed to increase by exactly the same factor so that there is no change in velocity.

For a sound wave, however, the ratio ω/k is not the speed of light, but speed of sound, say c_s, which is considerably smaller. Accordingly, to first order in v^2/c^2 we have $\omega' = (1 - \mathrm{v}/c_s)\omega$ and $k' = k$, so that we have $\omega'/k' = c_s\text{-}\mathrm{v}$, which indicates both a Doppler shift and a change in apparent sound velocity. Conversely, no change in wavenumber (or wavelength) is observed. This all conforms to what we would expect at low velocities. Because the wavelength is unchanged, any Doppler shift must be accompanied by a change in velocity, whereas for electromagnetic waves in free space the frequency and the wavenumber change by exactly the same factor such that the speed of light remains invariant. Although the results of a change of reference frame are characteristically different for sound waves and electromagnetic waves in free space, they nevertheless obey the same wave equation without contradiction. The transformation properties of the wave are built into the transformation of coordinates rather than the equation itself, and only when the velocity of the wave is identical to the speed of light in vacuo is it invariant under a change of reference frame. For electromagnetic waves in matter, however, when the wave velocity, c_m, is significantly different from c, then $\omega'/k' \neq \omega/k$. To first order in v, c_m would be shifted by an amount approximately equal to $-v \cdot \left(1 - \frac{c_m^2}{c^2}\right)$. While in general, the shift amounts to at most hundreds of meters per second in normal terrestrial situations, in extremely high velocity ionized gases, say, it could be quite significant. Nevertheless, the result forms the basis of Fizeau's experiment [60; 57, pp. 68-70], which, by means of an interferometer, produced a measurable result for the shift in water flowing at a modest speed. The significance of this at the time was that it produced less than half the value that would be expected in an ether based theory where the propagating wave would be borne along by the traveling medium.

6 Discussion

6.1 Is A Magnetic Field Really Necessary?

Given that the magnetic field arises from special relativity, as discussed in Section 5, and in particular that fact the whole of electromagnetic theory simply stems from treating Coulomb's law according with special relativity, can we not dispense with the notion of a magnetic field altogether and simply consider everything in terms of a proper relativistic treatment of the electric field? There are two main problems with this notion, however. One is that we prefer to avoid applying relativity theory to everyday phenomena, and the other is that we would have to consider separate reference frames for each different velocity involved, which is clearly impractical in all but the most trivial of circumstances. In particular, how would circulating currents be dealt with? The concept of a magnetic field, therefore, simply provides a straightforward method of dealing with this complex situation so that we can evaluate the forces due to any arbitrary collection of moving charges that is effectively inclusive of the relativistic corrections to the observed electric forces.

From a more elementary educational standpoint however, there is nothing overwhelmingly difficult about the modified Coulomb's law as given by Equation (13). The problem with forces that do not balance may be used as a means of touching on the link with special relativity, and of course it can be pointed out that on a practical footing all is well for closed circuits. It embodies the basis of magnetism and does not necessarily require explanation in terms of special relativity. As shown in Figure 9b, it can be taken as the starting point from which everything else follows in a consistent manner. Split into its two component parts, electrostatics and magnetostatics can be developed from a common footing. The connection between electricity and magnetism will therefore be recognized from the outset.

6.2 Equivalence of Solutions Based on Magnetic Poles and Currents

That two apparently distinct models of intrinsic magnetism - one based on poles and the other on circulating currents - should give identical solutions is more than a little surprising. It seems to imply that whenever we specify $\nabla\cdot\mathbf{M}$ throughout the volume of a particular medium, then $\nabla\times\mathbf{M}$ is automatically determined, and *vice versa*. Indeed, $\nabla\cdot\mathbf{M}$ and $\nabla\times\mathbf{M}$ must coexist in a form of duality. We know that by using Maxwell's equations in the form of first Equation (23) and then Equation (25), both approaches must provide fully equivalent field quantities, as illustrated for example in Figure 11, but due to the irreconcilable nature of solenoidal and divergent fields, the characters of $\nabla\cdot\mathbf{M}$ and $\nabla\times\mathbf{M}$ are very different. Even their spatial distributions are quite different, which all adds to the difficulty of seeing how this equivalence is possible.

Surprising or not, a full mathematical analysis does demonstrate the complete equivalence of these two descriptions [61]. While trying to understand this is a fascinating mathematical conundrum in itself, perhaps somewhat disappointingly it seems to have no great relevance from the point of view of electromagnetic theory. It suggests that the all-too-neat equivalence of the two different descriptions is more on a mathematical basis rather than a physical one. Nevertheless, it can be very confusing as, rather than having a single view, we tend to be split between two views, a simpler non-physical view on the one hand and a somewhat more complex physical view on the other. If, however, we put the pole description out of our minds, the problem no longer exists. All the same, the natural philosophers among us may wish to ponder as to why this tantalizingly close correspondence exists at all. Pole theory is so simple, it is correct in its results for the fields, and yet it gives the wrong behavior for the 'force field' inside a magnetic medium in that

- the force on poles follows **H** while the actual force detected by any infinitesimal dipole depends on **B** rather than **H**;
- within a medium consisting of current based dipoles, a test dipole tends to align parallel to the dipoles of the medium, while in a medium based on polar dipoles it would tend to do the opposite.

Even now, on the surface it could be possible to dismiss these issues on the grounds that it is, practically speaking, a simpler alternative to the somewhat more cumbersome circulating current theory. But, in the end we cannot accept as the basic theory something that is conceptually wrong, particularly when we already have a perfectly sound theory, albeit a little more mathematically difficult. In fact, the only real difficulty lies in avoiding the misleading notions and nomenclature left over from the pole theory, as discussed in Section 2 above. It would make matters considerably easier if we had never ventured down that route in the first place.

In the meantime, it is interesting to note that if magnetization were accounted for by a combination of both current and polar dipoles, as considered possible even in Maxwell's time, things would be much more complex. In order to describe this situation we would have to divide up the magnetization **M** into $\mathbf{M}_p$ for polar magnetization and $\mathbf{M}_I$ for the circulating current based kind. The 'force field' for our cylindrical sample would then be represented by a combination of Figure 11a for $\mathbf{M}_p$ and Figure 11b for $\mathbf{M}_I$. One could arrive at the seemingly paradoxical situation where, with balancing proportions of each type, the resulting 'force field' would be all but zero on the interior of the sample! Without trying to analyze this situation too deeply, it does appear to be quite implausible. It could make an interesting academic exercise, but as far as the known state of the universe is concerned we can say that the 'force field' in the interior conforms to Figure 11b alone. This is, therefore, just the same thing as saying that magnetic poles do not exist even as pairs, or that $\nabla\cdot\mathbf{B} = 0$ holds not only on a macroscopic scale but on the microscopic scale too. As far as we now know, there is no element of space small enough to be able to isolate a single pole from a dipole pair, or even to detect an imbalance in pole density.

6.3 Symmetry of Maxwell's Equations and Choice of Units

A number of writers have commented on the form or symmetry of Maxwell's equations under different systems of units, particularly regarding the appearance or non-appearance of particular constants, primarily 4π and c [22, p. 618; 62]. The fact that both ε_0 and μ_0 appear in the free-space or microscopic equations, Equations (22) or (23), and not in the macroscopic equations, Equations (20), however, should not be regarded as signifying any particular advantage or disadvantage either way. However things may be arranged, two independent constants are inevitably required so that we may relate both the electrical quantities to the inertial and the magnetic ones to the electric. Here the situation, in SI at least, is that with the microscopic forms the constants appear in Maxwell's equations while with the macroscopic form they appear in the constitutive relations, as in Equation (6) above. In general, however, we may choose where the constants occur and set the scale of measurement of **E** and **B** by their magnitude, but no more.

Today, Gaussian units and SI are the main contenders across the scientific and engineering world. The preferred SI system [63; 64], in which has μ_0 is given as $4\pi\times10^{-7}\,\mathrm{NA}^{-2}$ and $\varepsilon_0\mu_0 = 1/c^2$, is at least consistent over all physical quantities, not just electromagnetic, and is defined consistently with the modern view of magnetism. The derived electromagnetic units in SI are laid down by the International Electrotechnical Commission, IEC [33], and the poles embodied in the earlier MKS system have been quietly dropped.

As to the Gaussian system, this is still to be found in ISO 31-5:1992 [65]. However, its definitions vary from those given in IEC60050 and in particular while the definition of **H** through its curl does avoid poles, it seems of little practical use. As to the convenience of using these units, it is not always helpful to have the terms **B** and **H** being dimensionally the same (irrespective of units), for the temptation to interchange them casually is too easy to fall into. It is also a definite inconvenience to have to mentally adapt equations and units when going from one published work to another.

Even though the use of systems of units other than SI is now deprecated, it would still be valuable to retain formal definitions of Gaussian units that are consistent with SI. In this way it could be made clear that all the definitions have been unified. Explicit mention of magnetic poles being no longer valid would be also helpful in the short term. Giving a conversion factor, as for example in IEEE Std270-2006 [66] where the gauss is given as 10^{-4}T and the oersted as $250/\pi\,\mathrm{Am}^{-1}$, is helpful, but it would be even more so if we were reminded that these are to be taken as measures relating to the *modern* definitions of the magnetic field quantities, as in the SI system, as opposed to the old forms. The stance of the IEC appears to be somewhat different. The terms oersted and gauss are not to be found in IEC 60050 and so it can offer no guidance as to their usage and meaning. This is understandable as far as preserving the integrity of the SI system is concerned, but it leaves no way of finding a valid definition of those other terms and units that we still frequently have to deal with. There is, however, no shortage of definitions to be found and the old versions seem to turn up just as often as the new.

6.4 Choice of Field Variables

Maxwell's equations in the form of the free-space and microscopic equations, Equations (22-24), are stripped to the bare essentials, leaving no ambiguities as to the roles of the field quantities, and making it absolutely clear that the fields themselves originate entirely from the static and dynamic effects of spatial charge distributions. While we accept that this must be taken to include intrinsic magnetic effects, *i.e*. those arising from electronic orbital angular momentum and particle spin as opposed to observable current, their notional representation in terms of a microscopic current is not affected.

In terms of practical application to situations including real matter, the auxiliary fields **D** and **H**, preferably in the form of Equations (6) above, play a significant role in facilitating a model. Moreover, they obey complementary boundary conditions to **E** and **B**, whereas **P** and **M** obey none in particular. It is strange to consider in retrospect that, had the theory originally developed from the proper free-space form of Maxwell's equations, Equations (22), as a starting point, and then **D** and **H** had been introduced for just the purpose of describing matter, the resulting field **H** would still turn out to be the same as the field that was originally introduced as the force field based on a magnetic pole description. On the other hand it was **B** that was originally introduced as an auxiliary field in order to facilitate the description of induced magnetization! Serendipitous this may be, but nevertheless a historical accident that has led to a longstanding source of confusion. While the major textbooks are correct in their reference to the respective roles of **B** and **H**, few have done much to improve the basic understanding, either taking it as axiomatic or convention or leaving any explanations to a bare comment or footnote.

6.5 The Concept of Free and Bound Quantities

There is more significance to terms free and bound than at first meets the eye. At the simplest level, the term 'free' applies to isolated charges that we can cause to flow under the influence of an imposed electric field , if free to move, while the term 'bound' simply means all other charges which, being within a medium are not free in this sense but which may nevertheless be displaced to some degree. In the same sense, magnetic dipoles are almost always bound, as there is no magnetic equivalent of conduction.

Obvious as this may seem, the true significance of 'free' and 'bound' to electromagnetic theory is that free charges and currents are sources that we can manipulate *directly*, while the bound charges and magnetic dipoles are affected only *indirectly* as a result of this manipulation. For example, if we loop several elastic bands end to end so as to form a chain, we can exert a stretching force on the whole chain by gripping each of the two end bands and pulling. All the intermediate bands stretch or relax in response to the movement of the outer two. In this sense, the two end bands are free (to be manipulated) while all the others are bound (to each other). Given the force constants of all of the bands, we can work out the elongation of each band from the stretch applied to the two end-points because we know that the same force must apply to each band along the length of the chain. We can therefore eliminate the intermediate bands simply by evaluating an average force constant for the whole chain from the sum of all the individual displacements divided by the applied force.

The free quantities therefore move directly under the influence of applied forces, whereas the bound quantities move only so as to remain in equilibrium. The part played by the bound quantities is that they affect the net force seen by the free ones, and in order to solve specific problems, we generally need to eliminate the bound quantities from the equations through supporting equations such as the constitutive relations and $\mathbf{J} = \sigma\mathbf{E}$ in the case of conduction. Note, therefore, that even conduction charges, even though free to move, can be considered as 'bound' in this sense. We have therefore labeled them separately as ρ_{cond} so as to avoid any confusion. This distinction between free and bound quantities is therefore the main motivation for Maxwell's equations being written in terms of **B**, **E**, **H, D,** ρ_{free} and $\mathbf{J}_{free}$ as in either Equations (20) or (27) above.

Care must be taken, however, because this does not mean to say that problems can be readily solved with reference to the free quantities alone. For example, if a point charge q_{free} in free space is placed at a given distance from the plane surface of a semi-infinite dielectric body, what is the resulting electric field? This is determined not only by the single free charge, q_{free}, but by the bound surface charge density σ_{bound} that is induced on the dielectric's surface. We cannot simply use the equation $\nabla\cdot\mathbf{D} = \rho_{free}$ on its own, which clearly would have trivial results. Solutions must still be found be found for both **D** and **E**, and this is achieved by applying Maxwell's equations in each region separately, while uniting these by bringing in the boundary conditions given in Equations (9) above. The issue here is that the boundary conditions form an essential part of the supporting equations. Jackson [22, pp. 111-113] solves this problem by the method of images, without the aid of which the task would be far from trivial. This is a particular example of technique being an expedient to solving the problem rather than an aid to understanding the true nature of the problem itself. That is not to say that we do not learn from solving problems, but it is a bad thing to possess more by the way of technique rather than understanding.

6.6 Terminology and Definitions

Following the discussion of Section 2.5 above, renaming the fields more appropriately would do much to clarify their respective roles [9, pp. 241-242]. The problem is that while the original names appear in various forms and some are misleading to various degrees, they have been in use for a very long time.. In fact it is now quite common to refer to the more or less universal symbols **B**, **D**, **E**, **H**, **M** and **P** or terms such as B-field or H-field in place of the formal field names simply as a means of avoiding such problems, and this article is no exception.

With polarization and magnetization there are no real issues, the main problem lies with the magnetic field. Magnetic flux density is the SI preferred term for **B**, but this seems a little distant from its direct role in magnetic force. The term magnetic induction is widespread but has the same drawback. As previously discussed, the problem with a name like magnetic force field is that in this case the force is of a different vector character to the field and it does not even point in the same direction. It could be renamed simply as *the* magnetic field, which is now fairly common usage for either **B** or **H**, but it is **B** alone that truly deserves the name. Even the IEC definition [33, IEV 121-11-19] acknowledges the common use of the name magnetic field. Unfortunately the IEC definition of magnetic field [33, IEV 121-11-69] includes "**H** *together with*…**B**". Although this seems logical enough as things stand, this arises only because we are presently stuck with two fields appearing on a par with each other. It means that we cannot associate the term magnetic field with just one variable, as would be not only preferable but logical. Now when we wish to do so we must specify one or other of the IEC preferred terms, magnetic field strength [33, IEV 121-11-56] or magnetic flux density[16] [33, IEV 121-11-19], and it is this very choice that brings about the difficulties of which term should be used. It is therefore also doubly unfortunate in that **B** is given *second place* to **H** in the definition of magnetic field, nor is it the one that bears 'field' within its name.

Since the role of **H** can be considered auxiliary, a description already used by some authors [39, p. 18; 42] it could be renamed as auxiliary magnetic field, as suggested through Section 2.9. The advantage with **B** named the magnetic field and **H** as the auxiliary magnetic field is that there can be no doubt as to the roles and no need for yet another separate term to describe them both together as the magnetic field. As to other possible names for **H**, the problem with the term macroscopic field is that it has often been used to refer to any field within matter simply to distinguish it from the same field in free space, and it would also be useful to be able to refer to the electric field quantities in like terms.

In keeping with the names for the magnetic quantities, **E** would simply be referred to as the electric field, as is now common, while **D** would be referred to as the auxiliary electric field. There is some ambiguity associated with the term displacement, for example Beneson, Harris *et al* refer to the separate concept of displacement flux [7, pp. 451-452]. In the IEC nomenclature [33, IEV 121-11-40], displacement is secondary to the preferred term electric flux density, but the latter seems to have less currency.

As to the definitions themselves, as opposed to the names, the IEC International Electrotechnical Vocabulary, IEV, [33] appears to offer the best source in terms of completeness and consistency. There is no difficulty with the definition of either **B** or **M**, so that **H** is best defined through $\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$ [33, IEV 121-11-26]. In this way the concept of poles are avoided altogether, as is the notion of "ability to produce magnetic induction". As to **D**, the physical definition for displacement is now somewhat obscure (although it is in fact given by Beneson, Harris *et al* [7, pp. 453]) and displacement is now readily defined simply by $\mathbf{D} = \varepsilon_0\mathbf{E}+\mathbf{P}$ [33, IEV 121-11-40] (although not by either

[16] Magnetic induction does not appear in the index and is given only within of the definition of magnetic flux density as an alternative name.

ISO 31-5 or IEEE Std270 which both use $\nabla \cdot \mathbf{D} = \rho$). While IEEE Std270 and ISO 31-5 are generally consistent with IEC 60050, there are some other differences, *e.g*. in the definition of **H**. While IEEE Std270 does use $\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$, no separate definition for Magnetization is provided. On the other hand ISO 31-5 uses a different definition based on $\nabla \times \mathbf{H}$. There is therefore a lack of harmony and consistency between the various versions of the electromagnetic standards in some of the most crucial areas.

The existing names for the field quantities are summarized in Table 4 below together with suggested preferable forms. If these or similar names were adopted, both **B** and **E** would be clearly recognizable as the primary fields with **D** and **H** being clearly secondary in nature, and so over all four such electromagnetic field quantities it is only a matter distinguishing which is electric and magnetic, and which is auxiliary rather than primary. Unlike changes in the definition of units as suggested in Section 6.3 above, however, such changes would not be easy to bring about. It would inevitably be the subject of much debate – which would in turn lead to the proposal of more names or even combinations of old and new. But the problem does exist and so it would be better tackled sooner rather than later. At the moment it seems the only names that can be reliably agreed on are the actually the bare symbols **B**, **D**, **E**, **H**, **M** and **P** and so a complete break with past connotations is still something only to be hoped for!

6.7 Lorentz Force

If matter did include an intrinsic magnetic dipole that was not ultimately equivalent to a circulating current, the only description that would be available for this would be that of an inseparable pair of magnetic monopoles. If this were indeed the case, then the picture of a permanent magnet's 'force field' would need to be as in Figure 11a. Since Figure 11b would still apply to an electromagnet, the Lorentz force might then take a form such as

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times (\mu_0 \mathbf{H}_{mag} + \mathbf{B}_{cur})) \qquad (37)$$

This echoes Kitaigordsky's approach, Section 2.7, which appears to be in line with the historical development of the subject according to which **B** is for currents while **H** is for magnets. As shown in the 11a, the 'force field' within a magnet would tend to be antiparallel to that which emanates from the 54

pole faces. As we have stressed, this is entirely different from the situation in which the elementary dipoles are equivalent to a circulating current where the macroscopic 'force field' within the magnet is as shown in Figure 11b with the 'force field' being continuous from the inside of the magnet out through the end faces.

But if we do accept that the Equation (3) for the Lorentz force is entirely correct in all situations, then this is consistent only with the 'force field' of any sort of magnet being as shown in Figure 11b. All known magnets, therefore, must be equivalent to a circulating current (even for those originating from elementary particle spin which we cannot actually describe by a classical specific circulating current). Only circulating microscopic currents can sum to produce such a macroscopic picture. 55

Table 4: Summary of electromagnetic field names in use with proposed replacements. The existing names in bold are the current IEC nomenclature in which, problematically, magnetic polarization has a separate definition from magnetization. It refers to μ_0M as opposed to M, a subtlety that offers up yet another potential ambiguity.

Field	Existing Names	Proposed Names
E	**Electric Field Strength** Electric Field Electric Field Intensity	**Electric Field**
P	**Electric Polarization** Polarization Dielectric Polarization	**Electric Polarization**
D	**Electric Flux Density** Displacement Electric Displacement Dielectric Displacement [24, p4.6; 76]	**Auxiliary Electric Field**
B	**Magnetic Flux Density** Magnetic Induction	**Magnetic Field**
H	**Magnetic Field Strength** Magnetic Field Magnetic Intensity	**Auxiliary Magnetic Field**
M	**Magnetization** Magnetic Polarization	**Magnetization**

54

Finally, since the Lorentz force is entirely consistent with the theory of special relativity as applied to the electric field, there can be no remaining doubt that the term **B** therein really does mean **B** in all contexts.

6.8 Energy Density

In the literature, the energy density of an electromagnetic field is often expressed in different forms, either the free-space version, Equation (38a) below [67; 45, pp. 70 and 241], the macroscopic form such as (38b) which applies only to linear media [22, pp. 189 and 205] (almost always applicable to dielectric, diamagnetic and paramagnetic media, but not ferromagnets), and the general form (38c) which can be integrated in principle to find the energy density in any medium [15, p. 8]

$$\mathcal{E} = \frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2\mu_0}B^2 \qquad (38a)$$

$$\mathcal{E} = \frac{1}{2}\mathbf{D}\cdot\mathbf{E} + \frac{1}{2}\mathbf{B}\cdot\mathbf{H} \tag{38b}$$

$$d\mathcal{E} = \mathbf{E}\cdot d\mathbf{D} + \mathbf{H}\cdot d\mathbf{B} \tag{38c}$$

The energy density free space is unambiguously given by Equation (38a). Equation (38b) is a fairly clear extension of Equation (38a) which requires no particular interpretation. However in Equation (38c) things are not so obvious because in the electric contribution the differential applies to the *auxiliary* field, whereas in the magnetic contribution it applies to the *primary* field. We understand that for the electric field, **E** is the force field and that the work is done by the movement of charge as represented within $d\mathbf{D}$. With the magnetic field, however, if we believe that **B** is responsible for the force while **H** embodies the movement of magnetic dipoles being affected by that force, why do we not have $\mathbf{B}\cdot d\mathbf{H}$ rather than $\mathbf{H}\cdot d\mathbf{B}$? $\mathbf{H}\cdot d\mathbf{B}$ is what we would expect from a polar theory of magnetism, and so surely it should be the other way round in the modern theory. But this is not the case.

If we do work in the process of changing an electric field, we do so by moving free charges about. As discussed in Section 6.5 above, bound charges are not accessible to be moved directly and they are moved only as a consequence of moving the free charges, since this changes the electric field throughout the system. The bound charges are affected implicitly since not only do they follow the changes in the field, they contribute to them in turn. Provided we make any change sufficiently slowly, however, all the bound charges will remain in equilibrium throughout the change process and the work done in their movement is therefore zero. The work we do in moving the free charge, however, must still equate to the resulting change in energy stored in the entire system. There is therefore a considerable difference between the concept of energy density in free space and within a macroscopic medium. Let us look at Figure 10 in the entirely fresh context of a parallel plate capacitor filled with some dielectric material. In 10a, if the charge density σ on the parallel plate capacitor is made up of free and bound charge densities, σ_f and σ_b respectively, the electric field between plates is $(\sigma_f - \sigma_b)/\varepsilon_0$ and so the work done per unit area in moving a quantity of free charge $\delta\sigma_f$ from one plate to the other is $(\sigma_f - \sigma_b)/\varepsilon_0 \cdot \delta\sigma_f$. If the bound charge is zero, this integrates to give the electric part of Equation (38a), otherwise it gives the electric part of Equation (38c), since we can identify σ_f with D, σ_b with P and $(\sigma_f - \sigma_b)/\varepsilon_0$ with E.

As to magnetic energy, there are no free poles to move about and all magnetic dipoles can be considered as being bound. The nature of the Lorentz force itself is that no work is done on a moving charge that is undergoing a force of the form $\mathbf{v}\times\mathbf{B}$, as such a force has no component in the direction of motion. A change in magnetic energy, however, can be brought about through Faraday's law of induction as a result of a change in current. Consider now Figure 10b and let us allow the possibility of the dielectric being magnetic. The current density, K, in the top plate crosses the magnetic field, **B**, perpendicularly. A change purely in the magnitude of the current density will change the field, which in turn will generate an electric field **E** through Faraday's law of induction. This field will tend to oppose the change in current and so will be in the *x*-direction, while $-\frac{\partial \mathbf{B}}{\partial t}$ must be in the *y*-direction.

Consequently $\nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}$ reduces to $\frac{\partial E_x}{\partial z}=-\frac{dB}{dt}$. We need only be concerned with E_x within the plate itself because the apparent current density K_{free} exists only there. Furthermore, as **B** vanishes entirely outside the plate then so must E_x. We must therefore have $\frac{0-E_x}{\delta z}\rightarrow -\frac{dB}{dt}$ at the plate, where δz is the 'thickness' of the current sheet. In other words $E_x=\delta z\frac{dB}{dt}$. The bulk current density, J_{free} , however, is simply $K_{free}/\delta z$. We then have the rate of doing work, per unit volume, given by

$$
\begin{aligned}
\frac{dW}{dt} &= \mathbf{J}_{free}\cdot\mathbf{E}\\
&= \frac{K_{free}}{\delta z}E_x\\
&= \frac{K_{free}}{\delta z}\cdot\delta z\frac{dB}{dt}\\
\Rightarrow dW &= K_{free}dB
\end{aligned}
\qquad (39)
$$

Bearing in mind that the magnetic field **B** itself is due to the total current including any magnetic current, for our configuration we have in the steady state $B=\mu_0(K_{free}+K_{mag})$, which can be seen to be the equivalent of $B=\mu_0(H+M)$. When the medium between the plates is non-magnetic we have $K_{mag}=0$ and the result $dW=K_{free}dB$ may be integrated using the relation $B=\mu_0K_{free}$ to obtain a magnetic energy density $\mathcal{E}=\frac{1}{2\mu_0}B^2$, which establishes the second term of Equation (38a).

Otherwise, since K_{free} equals H, we have the more general form $dW=H\cdot dB$ when the medium is magnetic. This is consistent with the magnetic part of the energy density differential being as in Equation (38c). The energy change arises out of moving the free current against an electric field that originates from a change dB in the magnetic field. Consequently *it has nothing at all to do with any movement of currents or magnetic dipoles against a magnetic force*. The magnetism in a medium only comes into play indirectly through the magnitude of the electric force that is produced by, and in turn resists, a given change in current.

While for the electrostatic energy $E\cdot dD$ may represent electric field × (charge density × dz), $H\cdot dB$ effectively represents (charge density × velocity) × (electric field × dt). For the electrostatic part we are moving free charge against an electric field and integrating along *the path* of the charge. For the magnetic part, current is moving against an induced electric field and we are integrating over *the time* during which the field changes. Given that electrostatic energy is a form of potential energy, the comparison has been drawn between magnetostatic energy and kinetic energy [23, p. 230; 3, Vol. 2, pp. 197-198 and 271-276]. If the mechanical analogue of Equation (38a) is taken as $\mathcal{E}_{mech}=\frac{1}{2}kx^2+\frac{1}{2}mv^2$, then Equation (38c) becomes $d\mathcal{E}_{mech}=F\cdot dx+v\cdot dp$ where F is force, x is distance, v is velocity and p is momentum. If the magnetostatic energy were associated with poles in the same way that electrostatic energy is associated with charges, both would constitute forms of potential energy and integrating to find the work done would be along the spatial path in both cases. We know that this is not the case, but it does lead to a tempting but fallacious interpretation of $H\cdot dB$ as being magnetic field × (pole density × dz), analogous with electric field × (charge density × dz) as in the case of $E{\cdot}dD$. The major difficulty with this analogy is, of course, the identification of H with a force field similar to E. In addition, it would imply that the electromagnetic energy density $\mathcal{E}$ was entirely in the form of potential energy rather than potential energy plus kinetic energy, *which is of course necessary for any wave equation to exist*. Equation (38c), therefore, is to be treated with care. It implies no such analogy or pairing between **E** and **H** or with **D** and **B**.

6.9 Poynting Vector and Momentum Density Vector

The Poynting vector, defined by **S** = **E**×**H** is generally accepted as defining a flow of electromagnetic energy density [44, pp. 321-322; 9, pp. 131-135; 22, pp.189-190] on the assumption that Equation (38c) holds good for time-dependent fields. On the other hand **p** = ε_0**E**×**B** has been interpreted as an electromagnetic momentum density as discussed in Section 3.1 above. Note that the Poynting vector is normally taken to apply to the energy flow in waves, while the electromagnetic momentum density can also be considered to be associated with moving charges even when they do not radiate.

In free space we have **S** = **E**×**B**/μ_0 and so there is ample scope for confusion. The reason why **H** appears in the one form and **B** in the other is hard to argue in fundamental terms, as both **S** and **p** are mathematical constructs based on Maxwell's equations. In the derivation of the Poynting vector, however, $\mathbf{J}_{free}\cdot\mathbf{E}$ is taken to be the rate of doing work in the movement of charge. For the momentum density vector, however, the Lorentz force, is taken as the rate of change of mechanical momentum. Written in the form **F** = ρ**E**+ **J**×**B**, ρ and **J** are the *total* charge and current respectively, involving all bound and magnetic quantities. This certainly gives a clue as to why **H** appears in **S** while **B** appears in **p**, but it is hardly an intuitive matter. Not all issues regarding the correct use of **H** and **B** are therefore easily resolved and, as with the Lorentz force, care must be taken to use the correct definition and to avoid confusing the issue with convenient alternative forms because we simply prefer to use **H** rather than **B**, or *vice versa*, in various circumstances.

6.10 B-H Loop

The B-H loop, as it is called, is a typical method of measuring the basic magnetic characteristics of ferromagnetic materials [22, pp. 153-154; 9, pp. 125-126; 45, pp. 282-283]. A coil of wire is wound around a toroidal sample of the material under investigation. A low frequency AC current is applied to the coil and the voltage induced across the coil is measured.

Given all that has been said, it seems perplexing that in this measurement we vary **H**, the auxiliary field, while measuring the resulting value of **B**, the primary field, rather than the other way around. In the electric counterpart, where we would apply **E** to a dielectric-filled capacitor and measure the resulting field **D** in terms of the amount of free charge displaced, we can generate **E** by applying an AC voltage and infer **D** from measuring the current that flows, and so why can we not measure **H** versus **B** in a similar manner? The temptation is of course to assume that **B**, being the actual magnetic field, should be the independent variable, but unlike the electric counterpart, in order to generate the field we must start with a current. While the current does produce a magnetic field **B**, we cannot say directly what the resulting magnitude of **B** is, as it depends not only on the applied current but on the induced and permanent[17] magnetic currents as well. We can, however, infer **H** directly from the integral form of Maxwell's fourth equation because it depends *only on the free current in the coil* and the given geometry. On the other hand, from Maxwell's second equation we can measure **B** directly in terms of the back EMF generated as a result of the changing magnetic field. In effect, we can readily measure **B** versus **H**, but not so the other way around, just as we can we can readily measure **D** versus **E**, but not so easily the other way around, *simply because we are applying and measuring voltages and currents*. If we apply a voltage we vary **E** directly, and if we apply a current we vary **H** directly. The use of **E** and **H** as the 'variables' is therefore unrelated as to which is the principal field and which is the auxiliary one.

Within a typical B-H loop, however, we can see that the material magnetically saturates when the value of the applied field, **H**, is high enough. Here, **B** ceases to increase rapidly with increasing **H**. While it is difficult to separate cause from effect in a linear medium where **B** and **H** or **D** and **E** are simply proportional, the saturation of **B** seems to suggest that **H** represents the driving force and **B** the

resulting state of the material. What we really mean is that **H** represents the cause and **B** the effect. Inside the sample, the magnetic field that is acting to produce a torque on the magnetic dipoles is still none other than **B**, the only difference being that the value of **B** depends not only on the applied current but on the internal state of the material as determined by the magnetization **M**.

In the case of the dielectric example, the applied electric field depends on the applied voltage and given geometry. The sample will change its internal state, or polarization **P**, as a result of the free charge arriving on the capacitor plates, but in the steady state the free charge and polarization must come to an equilibrium such that that the resultant electric field within the capacitor is equal and opposite to the applied field, which we can of course measure directly from the voltage and given geometry. The situations are not dissimilar, the distinction lies in the variables that we can control directly and indirectly, as per the discussions of free and bound quantities and energy density in Sections 6.5 and 6.8 above.

To conclude, it is worth noting that in the case of the capacitor, it is possible to reverse the situation and directly control the free charge, and therefore **D**, by applying a known current for a given time, while measuring the resultant **E** through the voltage between the capacitor plates. Similarly, it is possible with the B-H loop to apply a voltage pulse to the coil and measure the resulting current so as to vary **B** while measuring **H**, but there is the considerable practical difficulty that as the available change in magnetization decreases, the current becomes very large! We point this out, however, simply to emphasize that there is an arbitrariness about which variable is independent or dependent and that the conceptual association of the independent variable with the force field is somewhat of a red herring.

---

[17] In this situation the permanent magnetisation itself may be varying, but it nevertheless 'permanent' can be taken to mean 'still present when all free and induced current is set to zero'.

6.11 Flux and Lines of Force

From the earliest lessons in the science of electricity and magnetism, lines of force are introduced in order to make the unfamiliar notion of a field easier to grasp. Indeed, the concept arose in the early history of this science as a means of describing something, like the wind, that was known to exist and the effects of which could be measured, but which could not be seen. This was later given a more rigorous footing by defining the number of lines of force that emanate from a unit charge or pole, 4π to be precise [13, p.40]. The original idea for envisaging the field is still valid, but there is probably no continuing need to have a rigorous defined relationship between the number of lines and the actual field strength, which, after all, is really quite arbitrary. A problem arises, however, if it is taught that lines of force emanate from a point. This may be true for an electric field, but it should not be for a magnetic one, in spite of the fact that the easiest way to demonstrate the idea is to dust iron filings over a surface such as a piece of stiff board under which is held a bar magnet. The next step in this analogy between the alignment of the filings and the field is to deduce that the magnet has 'poles'. These poles, which correspond to the convergence of the lines at the ends of the magnet, do not exist – as we would realize if we used a hollow solenoid rather than a permanent magnet. In fact, if we wound a large enough cylindrical solenoid *through* the board, with the axis of the solenoid lying in the plane of the board, the trick with the iron filings would reveal no poles. The apparent lines of force would enter into the 'magnet' at one end and pass out through the other without converging on any such poles.

In the iron filing experiment, each filing acts as an induced dipole, the nature of which is like a miniature compass needle. As in Section 2.9.3 above, we can still identify the magnetic 'force field' with the alignment of a test dipole, rather than with the supposed path of a free pole. It is perhaps unfortunate that the term pole is the root of the word dipole, but the term dipole is universal, even in electric and radio terminology. At least with dipoles, however, we can keep the same visual picture while giving it a basis that is free from the concept of individual poles. Beyond their value as visual aid to understanding, lines of force are of course a useful form of graphical representation, *e.g.* of fundamental modes. Even so, for computer modeling of fields the generation of lines of force is far from trivial. It is usually easier to generate a vector plot, however, which simply indicates the strength and direction of the field over a grid of points. Visually, the results are similar. Given this and a variety of other ways of representing fields, *e.g.* by introducing colors and contours, it is doubtful that there is a continuing need for lines of force as a rigorous and quantitative, rather than a qualitative, concept.

Flux is a concept similar to lines of force, but it tends to be introduced at a more advanced stage. Using the visual analogy of a density of lines, the number of lines that pass through a given area is the flux that it encloses. As the word suggests, flux depicts a flow which, integrated over an area gives the total flow through it. And so the lines used to visualize a flux density are different from lines of force, rather they are lines of flow. Making an analogy with fluid dynamics, lines of force would correspond to the pressure causing the flow, while the flux density would correspond to the rate of flow.

Historically, the fields **D** and **B** have been associated with the electric and magnetic flux densities respectively. The concept of **D** as a flux is at least consistent Maxwell's fourth equation where its time derivative appears along with **J,** the current density, which represents a real flow. The integral of the flux over a closed surface yields the free charge enclosed within in the case of $\int \mathbf{D}\cdot d\mathbf{A}$, whereas $\int \mathbf{B}\cdot d\mathbf{A}$ must be zero. While reference to **D** as a flux density is now little used, despite the official SI term electric flux density, for **B** flux density is still very much part of the present day terminology. Flux in the electromagnetic context is generally taken to mean magnetic flux unless stated otherwise, and both the term itself and the concept endure as a result of Farday's law of induction. But there is no real flow, as there is with, say, **J** the current density. Maxwell felt that **B** was associated with a kind of momentum [23, p. 230; 3, vol. 2, pp. 197-198 and pp. 271-276], which is consistent with the concept of a flow. Moreover, in terms of energy density, which we discussed in Section 6.8 above, force times flow is the rate of doing work. Given Equation (38c) in the form $dW = \mathbf{E}\cdot d\mathbf{D} + \mathbf{H}\cdot d\mathbf{B}$, if **E** and **H** are force fields then this is consistent with **D** and **B** being fluxes. But if **H** is not truly a force field we now have difficulties with **H** and **B** in this context. Besides, with **B** it could be argued that any flow or 'momentum' involved is more really associated with the current giving rise to the magnetic field rather than the field itself. For example, in the Einstein de Haas experiment it has been demonstrated that angular momentum is transferred from the magnetization to the sample itself when its magnetization is reversed [14, p.167].

Putting aside for a moment the issues concerning whether **H** can be considered to be a force field and turning back to Maxwell's equations, Equations (25-26), as they would be in a pole description taking **E** and **H** as the basis fields, then in the third equation a term $\mathbf{Y} = \frac{\partial \mathbf{M}}{\partial t}$ arises along with the time

derivative of **H**. We have said that **Y** may be thought of as a magnetic current density, a flow of bound poles, in the same way that $\mathbf{J}_{bound} = \frac{\partial \mathbf{P}}{\partial t}$ is a flow of bound charge. In associating **B** with a flux, then, we are making a similar link between **B** and **Y** as holds between **D** and **J**.

In truth, in considering **H** as a force field and **B** as a flux, we are simply making an analogy on an operational basis rather than on a physical one. The time derivatives of **D** and **B** appear in Maxwell's equations together with the curls of **H** and **E** respectively. In their integral form, these equations then relate the integral of **H** (or **E**) around a given loop with the integral of **D** (or **B**) over the area enclosed within the loop, which is taken to be the flux. With **E** and **D** the notion of force field and flux is admissible, whereas with **H** and **B** it is based on their roles within the equations being similar to those of **E** and **D** (unless, of course, we are willing to accept poles as a reality). This situation is similar to Lagrangian mechanics where generalized coordinates, velocities and forces are employed. Within the Lagrangian formulation these quantities have an analogous role to their physical counterparts, but in general the analogy is mathematical rather than physical.

The SI term magnetic flux density and its unit Wb/m^2 emphasize the notion of a flux. Tesla is the preferred form for Wb/m^2 and does not, of itself at least, imply a flux. Many of us may find it helpful to remember the dimensions of **B** through *electric field / velocity* = [Vm^{-1}]/[ms^{-1}] = Vsm^{-2} rather than by *flux/area* as this altogether avoids the issue of flux by relating directly to the Lorentz force.

6.12 The Local Field

So far in this article we have deliberately treated all media in the classical view, as continuums devoid of any structure on a molecular scale. Having assumed the simplest of frameworks, the conclusions that were reached must be independent of specific microscopic detail and are therefore quite general as a phenomenological treatment. But in order to make the step from the dielectric or magnetic susceptibility of individual molecules to the macroscopic 'continuum' picture, however, a complication arises, as discussed in Section 2.12 above. The field at an individual molecule is not equal to the background field that pervades the medium treated as a continuum, for we must discount the molecule itself in order to establish this. For dielectric materials, this is the basis of the Lorentz-Lorenz treatment [14, pp. 89-95; 15, pp. 100-104; 16, pp. 150-158; 49, pp. 137-139; 22, pp. 116-119] in which the molecule under consideration is envisaged to be in a small cavity, generally spherical, within the medium. If we consider only the dielectric case, as the magnetic case would seem to be mathematically quite analogous, the field $\mathbf{E}_{loc}$ seen by the single molecule is equivalent to the sum of four components. Following Kittel [14, pp 89-95] these are:

- $\mathbf{E}_0$, the external field applied to the body (as derived from some distant free charge distribution)
- $\mathbf{E}_1$, the depolarization field, due to the bound charge induced on the outside surface of the body. This is a function of the shape of the body.
- $\mathbf{E}_2$, the Lorentz field, due to the induced bound charge seen on the surface of the internal cavity
- $\mathbf{E}_3$, the field due to the surrounding molecules that were bounded by the cavity, before they were taken away

Note that all these four fields are generally of different magnitudes so that the local field is different from both the applied field and the prevailing field on the interior of the body. While at first there would seem to be some arbitrariness about the chosen size and shape of the selected cavity, for media that are effectively isotropic a sphere is appropriate. Finally, if the molecules within the cavity have at least cubic symmetry, their contribution sums to zero. The method due to Ewald and Oseen [15, pp. 84-87 and pp. 100-104] provides a rigorous derivation of the local field, while in a crystalline structure the local field may be calculated directly from a lattice sum of the dipole fields [14, pp. 122-123 and 350-351] over all the molecules in the body except for the one in question, so that the ploy of using a cavity introduced by Lorentz not actually essential to the result.

The key point, however, is that the selected molecule within either the lattice or cavity is now being treated as if it were *in free space rather than the medium itself*, as all the contributions to its local field have been resolved in terms of fields due all charge distributions, ρ_{free} and ρ_{bound} . In other words, we are now back in the *microscopic* picture.

This situation, where we resolve the local field *in vacuo*, must not be confused with the macroscopic fields within dielectric and magnetic media that have been defined so that we can deal with them without explicit reference to magnetic currents or bound charges, but in terms of dielectric constant and magnetic permeability. The macroscopic fields are given by the sum of the applied field and depolarization field alone, *i.e.* in the electric case $\mathbf{E} = \mathbf{E}_0+\mathbf{E}_1$.

For the local field in the magnetic case we may run through the same analysis as above but with $\mathbf{H} = \mathbf{H}_0+\mathbf{H}_1+\mathbf{H}_2+\mathbf{H}_3$ replacing the corresponding terms in **E** for the contributions to the field. In employing **H** rather than **B** in this process we are simply following the mathematical analogy between **E** and **H** for the sake of expediency, and this does not represent a momentary lapse in philosophy. Surprisingly, while there are references to the local electric field throughout the literature, the local magnetic field is rarely mentioned in a similar context. In fact, as discussed in Section 2.1 above, when the torque on a magnetic dipole is quoted in the literature, a more significant issue is whether the field involved is either **B** or **H**! Surely this lack of clarity compared with the literature on molecular and macroscopic dielectric polarizabilities would throw up some erroneous results? Let us therefore enquire further.

Almost all problems in magnetic materials which involve the molecular scale, *e.g.* NMR and ESR, deal with either very weak or very strong magnetism. In the case of diamagnetic and paramagnetic materials, the magnitude of the magnetic susceptibility is of the order of 10^{-3} or less, so that the local field generally represents a fairly trivial correction to the macroscopic field. On the other hand, in ferromagnetic and ferrimagnetic materials where $1<<\mu$, individual magnetic moments interact strongly with their neighbors and tend to behave as a highly correlated ensemble. Any applied field can therefore be considered as acting on $\overline{\mathbf{m}}$, the average magnetic moment of the ensemble as a whole, rather than on each $\mathbf{m}_i$ separately. Kittel [68, 69] and Collin [70], who do pause to remark on the local field, argue that the Lorentz field correction, being proportional to **M**, will produce no net torque

on $\overline{\mathbf{m}}$, since any term of the form $\overline{\mathbf{m}}\times\mathbf{M}$ simply vanishes given that $\overline{\mathbf{m}}$ is clearly proportional to **M**. Unlike typical dielectric cases where the Lorentz field and the depolarization field are of similar magnitudes, in highly magnetic materials the depolarization field alone is therefore the most significant factor in relating the applied field to the effective internal field.

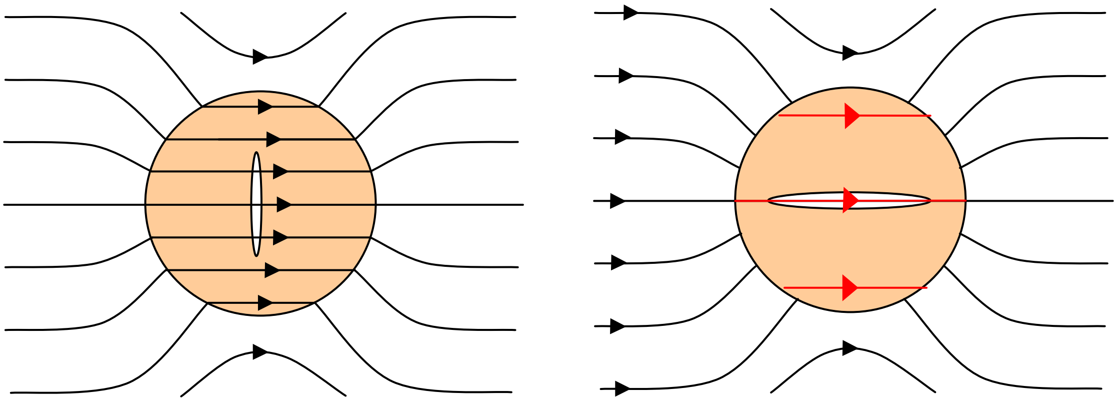
Particularly frequent throughout the literature on spin resonance and its applications, the use of $\mu_0\mathbf{m}\times\mathbf{H}$ or $\mathbf{m}\times\mathbf{H}$ rather than $\mathbf{m}\times\mathbf{B}$ for the torque on a magnetic dipole may be unhelpful or misleading. Such casual usage, however, generally succeeds in achieving the required result because it turns out that it does not matter whether we use **B** or **H**, since these are different only by the term $\mu_0\mathbf{M}$ which, as in the case of the Lorentz field, can produce no net contribution to the torque, $\overline{\mathbf{m}}\times\mathbf{B}$, as discussed further in Appendix 2.1 below

Before leaving this issue, however, to demonstrate that we are not just dealing with a point of semantics, consider what happens when we substitute $\mu\mathbf{H}$ for **B** in the torque $\mathbf{m}\times\mathbf{B}$. We get $\mu\mathbf{m}\times\mathbf{H}$ for the torque, contrary to the value $\mu_0\mathbf{m}\times\mathbf{H}$ that we have just reasoned. The problem is, of course, that **H** and **B** are not always parallel even in an isotropic medium. If they were, then **H** and **M** would also be parallel, resulting in no net torque at all on the magnetic moment. In spin resonance, the applied RF field is orthogonal to the static magnetization, while the magnetic moment takes on a time dependent variation that is orthogonal to the applied static field and resulting magnetization. Both of these pairings, being orthogonal, do contribute to the torque that drives the interaction between field and magnetic moment.

6.13 Cavity Definitions of B and H

Lord Kelvin proposed definitions of the macroscopic fields based on the fields that would exist in a small ellipsoidal cavity excised from the medium [49, pp. 137-139; 9, pp. 213-214]. This is shown in Figure 12 for the case of the magnetic field. In one case the ellipsoid is taken as a flat disk, the minor axis being parallel to the field, and in the other case a long needle shape, with the major axis parallel to the field. In the case of the needle-like cavity, **E** and **H** within the cavity are the same as within the medium, while in the case of the disc-like cavity, **D** and **B** are the same in the cavity as in the medium. In trying to identify the force that would act on a moving charge within a continuous medium, we cannot make use of any such cavity, since with $\mu = \mu_0$, $\mathbf{F} = q\mathbf{v}\times\mathbf{B}_{\mathrm{cav}}$ and $\mathbf{F} = q\mathbf{v}\times\mu\mathbf{H}_{\mathrm{cav}}$ are equally valid therein. By choosing the shape of the cavity we are only making $\mathbf{B}_{\mathrm{cav}}$ (or equally well $\mu_0\mathbf{H}_{\mathrm{cav}}$) equal to $\mathbf{B}_0$, $\mathbf{B}_{int}$, $\mu_0\mathbf{H}_{\mathrm{int}}$, or indeed anything in between.

In truth, these 'definitions' are restatements of the well known field boundary conditions summarized in Equations (9) above. While these notions distinguish between the fields, they no more define them than do the boundary conditions. It gives us no guidance as to the meaning of the fields or to their roles in respect of electromagnetic force. The problem, therefore, is the notion that these 'definitions' might be of some help in sorting out the fields, whereas in reality, they are not definitions at all.



(a) **(b)**

Figure 12 : Cavity 'definitions' of the magnetic field within a magnetic sphere placed in a uniform background field $\mathbf{B}_0$, *in vacuo*. In (a) we have lines depicting B which are continuous. An ellipsoidal cavity within the sphere contains a field, $\mathbf{B}_{cav}$, which strongly depends on its shape. For a spherical cavity, the field will equal the background field $\mathbf{B}_0$, while for a thin disc as shown it will be the same as $\mathbf{B}_{int}$ within the magnetic material. In (b) the lines now depict H. Outside the sphere, H is indistinguishable from $\mathbf{B}/\mu$, but within $\mathbf{H}_{int}$ is much reduced in magnitude. The field lines of H cannot be continuous across the surface of the sphere when the density of lines represents its magnitude. For the needle-like cavity shown, $\mathbf{H}_{cav}$ equals $\mathbf{H}_{int}$, but this varies with cavity shape exactly as for $\mathbf{B}_{cav}$, given that $\mathbf{B}_{cav} = \mu_0\mathbf{H}_{cav}$.

7 Conclusions

Maxwell's equations have been the basis for the description and analysis of all electromagnetic phenomena to date. Provided they are treated on a modern footing, they afford a description that is fundamentally correct despite their phenomenological and macroscopic origins. They do not of themselves, however, rule out the existence of magnetic poles, rather they provide a basis for magnetism that does not require them. Poles may be included, but they are ruled out on the basis that they do not give a correct description for the magnetic force and that special relativity together with Coulomb's law fully accounts for magnetism without them.

Much of the confusion that arises between the old pole based and modern current based theories of magnetism lies in the fact that pole theory mimics the true theory so well, while its apparently greater simplicity is both intriguing and appealing. There are many situations in which relationships that are formally correct within the new theory are at first sight more consistent with the old, *e.g.* for the B-H loop we have $B = B(H)$ and for the energy density differential we have $d\mathcal{E} = \mathbf{H}\cdot d\mathbf{B}$. Perhaps this has helped the old picture to linger on, but equally there has not been sufficient effort to abolish or at least update old definitions and terminology, and instances of the casual use of H and B out of their proper contexts are plentiful. Although the SI system as represented in IEC 60050 provides an effective basis for the modern view, there is still room for objection to some of the names, but more problematically in some important definitions there is a lack of harmony between it and the derivative standards IEEE Std270 and ISO 31-5.

This paper has scratched the surface of the historical development of the subject, and then only to the extent that has been necessary to explain where we now stand with regard to some outmoded conventions and ideas that have lingered on. While these may have a historical place, they no longer fit in with the logical development of the subject. It would seem to be very worthwhile if someone were to take up the challenge of producing a concise, definitive and up-to date chronology of the historical development, pinning down all the cornerstones in terms of the key observations and theoretical formulation along the way - that is to say, not just the way things are now, but the how and the why. If it were widely accessible it would do much to assist the teaching of the subject, and no doubt all of the information is available, it simply needs to be gathered, distilled and reassembled.

To the same end, those of us who study the subject, or who produce works based on it, should consider avoiding those traditional approaches that are no longer in its best interests. In fact, simple avoidance may not be enough, pointing out the pitfalls may be equally appropriate. Perhaps it will be also be necessary to find novel ways of introducing students to the fundamentals in order to avoid making things too complex or mathematically demanding, but nevertheless it would be of benefit if all those who have found this paper to have been of interest should try to set an example by adhering only to the modern concepts and letting the old ones pass into history.

The current age of advanced computer technology and sophisticated problem-solving packages for electromagnetics bestows mixed blessings. On the one hand it may well bring the risk that the software will be used blindly without real understanding of the fundamentals, but on the other hand it does bring more freedom from the difficult and often highly complex technical aspects of solving electromagnetic problems. This should allow more time in which to study and understand the very basics of the subject, which, as we have seen here, has several aspects that can appear to be cryptic and can often be perplexing. Even if special relativity is beyond the reach of students in the early stages or within certain disciplines, introducing a modified Coulomb's law in the form $\mathbf{F}_{12} = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2}\left(\hat{\mathbf{r}} + \frac{1}{c^2}\mathbf{v}_1 \times \left(\mathbf{v}_2 \times \hat{\mathbf{r}}\right)\right)$ appropriate to moving charges is only a little more difficult than the usual static form where the possibility of the charges actually being in motion is overlooked. Students would then be given the idea that magnetism is something related to Coulomb's law rather than something altogether apart from it. Ampere's force law, of course, follows directly and not as a separate entity, and the electromagnetic fields **E** and **B** drop out as a consequence, as does the Lorentz force. Having thus defined **E** and **B**, there can be no doubt or ambiguity as to the fields responsible for electromagnetic force, and in particular that **H**, like **D**, is involved in only an auxiliary role.

It is to be hoped that we have been able to satisfy the reader with reasoned answers and comment on the numerous and often troublesome issues raised in our introduction. A sounder footing on the basics of the subject will help to promote understanding, while some encouragement with regard to terminology and usage will help to promote good communication and to minimize confusion. It is further to be hoped that the problem issues that we have discussed will begin to disappear sooner rather than later.

8 Acknowledgements

The author would like to express his thanks to Professors Peter Grant and Alan Murray at the School of Engineering and Electronics, University of Edinburgh, Professor Alan Sangster at the Department of Computing and Electronics, Heriot-Watt University, Dr Simon Watts, Thales Defence UK and Professor David Vass for reading earlier drafts of this article and offering comments and encouragement. Dr Richard Dougal of the James Clerk Maxwell Foundation and the Department of Physics and Astronomy, University of Edinburgh, provided similar assistance and contributed some of the reference material. The advice of Professor Hugh Griffiths of the Department of Electronic and Electrical Engineering at UCL and Professor Tapan K. Sarkar of Syracuse University, NY, is also appreciated.

9 Appendix 1:
The Essential Electromagnetic Equations

As far as essential equations are concerned, it is always possible to put together a variety of equivalent forms. The equations summarized below in Table 5 provide at least a basis. They are numbered as in the main text and have been selected based on -

- Maxwell's Equations, in microscopic and macroscopic form.
- Definition of all the required microscopic source terms.

- Definition of **D** and **H** as convenient auxiliary fields.
- Constitutive relations and definition of **ε** and **μ** where linear relationships apply.
- The Lorentz force.

- Definition of an infinitesimal magnetic dipole moment in terms of a circulating current.
- Coulomb's law, as extended to moving point charges.
- Symbols are have their conventional meanings as referred to in this article. Other well-known forms of many equations may be reconstituted, *e.g.* Coulomb's law, Ampere's laws, Biot and Savart's law.

Table 5. The essential Electromagnetic equations (see Section 9).

Equation	Comment	Text Reference
$\nabla\cdot\varepsilon_o\mathbf{E}=\rho_{total}$ $\nabla\cdot\mathbf{B}=0$ $\nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0$ $\nabla\times\frac{\mathbf{B}}{\mu_o}-\frac{\partial\varepsilon_o\mathbf{E}}{\partial t}=\mathbf{J}_{total}$	Microscopic form of Maxwell's Equations	(23)
$\rho_{total}=\rho_{free}+\rho_{cond}+\rho_{bound}$ $\rho_{bound}=\nabla\cdot\mathbf{P}$ $\mathbf{J}_{total}=\mathbf{J}_{free}+\mathbf{J}_{bound}+\mathbf{J}_{mag}$ $\mathbf{J}_{bound}=\frac{\partial\mathbf{P}}{\partial t}$ $\mathbf{J}_{mag}=\nabla\times\mathbf{M}$	In free space we have: $\rho_{total}=\rho_{free}$ $\rho_{cond}=\rho_{bound}=0$ $\mathbf{J}_{total}=\mathbf{J}_{free}$ $\mathbf{J}_{bound}=0$ $\mathbf{J}_{mag}=0$ (except for isolated magnetic dipoles)	(24)
$\mathbf{D}=\varepsilon_o\mathbf{E}+\mathbf{P}$ $\mathbf{H}=\frac{\mathbf{B}}{\mu_o}-\mathbf{M}$ $\mathbf{D}=\boldsymbol{\varepsilon}\cdot\mathbf{E}$ $\mathbf{H}=\boldsymbol{\mu}^{-1}\cdot\mathbf{B}$	Constitutive Relations (general and linear cases)	(6), (8)
$\nabla\cdot\mathbf{B}=0$ $\nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0$ $\nabla\cdot\mathbf{D}=\rho_{free}$ $\nabla\times\mathbf{H}-\frac{\partial\mathbf{D}}{\partial t}=\mathbf{J}_{free}$	Standard Form of Maxwell's Equations (Macroscopic)	(33)
$\mathbf{F}=q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$	Lorentz Force	(3)
$\mathbf{m}=I\hat{\mathbf{n}}dA$	Magnetic Moment of Infinitesimal Current Loop	(16)
$\mathbf{F}_{12}=\frac{q_1q_2}{4\pi\varepsilon_o r_{21}^2}\left(\hat{\mathbf{r}}_{21}+\frac{1}{c^2}\mathbf{v}_1\times(\mathbf{v}_2\times\hat{\mathbf{r}}_{21})\right)$	Modified Coulomb's Law	(13)

10 Appendix 2: Examples

10.1 Magnetic Spin Resonance

Magnetic spin resonance in ferrites conforms to a semi-classical description [14, pp. 152-155]. It is commonly employed as the basis of non-reciprocal microwave devices such as isolators and circulators [71].

Depending on the preference of the author of those papers and textbooks which discuss the effect [14, p. 155 and pp. 167-171; 72; 71, p.10; 39, pp. 456-460], the torque on a magnetic dipole **m** is taken variously as either

$$\mathbf{\Gamma} = \mathbf{m} \times \mathbf{B} \tag{40a}$$

or

$$\mathbf{\Gamma} = \mathbf{m} \times \mathbf{H} \tag{40b}$$

Note that when we see the form involving **H**, this generally implies that emu or Gaussian Units are being employed, for in SI units it would even be dimensionally incorrect without the constant μ_0 .

Often the use of Equation (40) is simply implied in the ‘equations of motion’ for the magnetization

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} \tag{41a}$$

or

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{H}, \tag{41b}$$

where γ is the gyromagnetic ratio. If we assert that at least one of these equations is incorrect, then should there not have been some wrong results? Fortunately this is not the case for we can go between the equations as follows:

$$\begin{array}{ll} \text{SI} & \text{emu} \\ \mathbf{\Gamma} = \mathbf{m} \times \mathbf{B} & \mathbf{\Gamma} = \mathbf{m} \times \mathbf{B} \\ \quad = \mathbf{m} \times \mu_o (\mathbf{H} + \mathbf{M}) & \quad = \mathbf{m} \times (\mathbf{H} + 4\pi\mathbf{M}) \\ \Rightarrow \overline{\mathbf{\Gamma}} = \overline{\mathbf{m}} \times \mu_o \mathbf{H} + \overline{\mathbf{m}} \times \mu_o \mathbf{M} & \Rightarrow \overline{\mathbf{\Gamma}} = \overline{\mathbf{m}} \times \mathbf{H} + \overline{\mathbf{m}} \times 4\pi\mathbf{M} \\ \Rightarrow \dfrac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mu_o \mathbf{H} & \Rightarrow \dfrac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{H} \end{array} \tag{42}$$

That is to say, while the torque on individual dipoles depends on **B** rather than **H**, as an ensemble the average torque $\overline{\mathbf{\Gamma}}$, can be equally said to depend on **H** since the average dipole moment, $\overline{\mathbf{m}}$, is parallel to **M** with the result that the term $\overline{\mathbf{m}} \times \mathbf{M}$ vanishes. Consequently, it does not matter whether either **B** or **H** (in SI, μ_0**H**) is used in Equations (40) as a starting point for the elementary theory of spin resonance, the result is the same. What we can say, however, is that substituting μ**H** for **B** would be an easy mistake to make as it would give a completely erroneous answer by a factor of μ/μ_0! The reason for this is that if **B** and **H** were completely parallel, then the effect would vanish altogether since **M** and **H** would also be parallel so yielding a vanishing cross product in Equations (42). This example very clear illustrates why the terms **B**, **H** and μ**H** should not be used casually for convenience, they must be treated knowledgeably.

Taking $\mathbf{\Gamma} = [\mu_0]\mathbf{m}\times\mathbf{H}$ for of the force on a magnetic dipole is incorrect in principle, since if it is ever correct this is only within a restricted context, as in the example above. In much of the work in ESR, where the Gaussian system is commonplace, $\mathbf{H}_0$ is regularly used to refer to the applied field, and often simply referred to as **H**. The approach of Kittel, who carried out much of the early work on ESR in ferrites, is fairly typical [69; 14, p. 155 and 167-171]. One of his well known results, applicable to a sample with a plane surface, is that the resonance frequency is given by $\omega_0 = \gamma(BH)^{1/2}$. This result, however, is only a special case of a general formula that includes the dependency of the internal field on the shape of the sample. It should not be interpreted as implying yet another possible form of Equations (40).

10.2 The Hall Effect in a Magnetic Conductor

The Hall effect in a magnetic conductor would be expected to depend primarily on **B** rather than **H**. Historically, however, the Hall coefficient is defined in terms of *E/JH*, where *E* is the induced electric field relative to a current density *J* in the transverse field *H* [14, pp. 241-242]. Again, the old legacy is evident in that *H* appears in the literature rather than *B*. Most of the available data on the Hall Effect applies to non-magnetic metals and semiconductors where the difference between **B** and μ_0**H** is insignificant. Within magnetic conductors, however, the so-called spontaneous or extraordinary Hall effect cannot be treated semi-classically [73, 74]. In such conductors there are two distinct Hall coefficients and the Hall field is described by

$$E_H = R_o H + R_1 M \tag{43}$$

In principle, the right hand side of this equation should correspond to the magnitude of the magnetic part of the Lorentz force. Indeed, the term R_0H does correspond to the Lorentz force and is consequently termed the ordinary Hall effect, which predominates above T_c, the Curie temperature, above which the possibility of permanent magnetization ceases. On the other hand, R_1M corresponds to the so-called extraordinary effect that predominates below T_c.

Now here the appearance of *H* rather than *B* in Equation (43) is rather troublesome. One is left to question whether the use of *H* refers to the applied external field, or whether it should mean the field within the sample. Moreover, one could ask whether the term in *M* accounts for this, so that it really means

$$\begin{aligned} E_H &= R_o(H+M) - R_o M + R_1 M \\ &= R_o B + (R_1 - R_o)M \end{aligned} \tag{44}$$

In nonmagnetic materials this is of almost no consequence and the habit has been to use *H*. One reference [75], however, does quote the equation for the Hall field in the unambiguous form

$$E_H = R_o B + R_1 M \quad (45)$$

stating B to be "the magnetic induction in the material", which puts it beyond all doubt. Now, it turns out that the extraordinary term, which depends on M alone, normally tends to be dominant and does not actually arise from the Lorentz force but from scattering due to spin-orbit coupling. Nevertheless, R_0 has been measured in Iron, Cobalt and Nickel, but even here the results do not help, as its sign is positive for iron while it is negative for the other two. In short, the effect is sufficiently complex in magnetic materials that it cannot be taken as clear-cut evidence for the true nature of the Lorentz force without additional detailed information about the behavior of the conducting charges. This aside, it is another situation where we often encounter references to H rather than B in the literature without due explanation.

10.3 The Force on a Current-carrying Magnetic Conductor

The force on a current-carrying magnetic conductor in the presence of a uniform magnetic field is analyzed by Stratton [9, pp. 258-262] who assumes a cylindrical wire of permeability μ_1 carrying a current I with the wire running parallel to z and embedded in a medium of permeability μ_2 within which the applied field B_0 is uniform and directed along x. After a complicated analysis, Stratton's result is simply

$$\begin{aligned} F_y &= \mu_2 H_o I \\ &= B_o I \end{aligned} \quad (46)$$

Taken at face value, this means that the force on the wire is dependent on the external field B_0 alone and not, as we might initially expect, on the field B_{int} experienced by the current carriers within the wire. But the reason for this is simply that any demagnetization field within the wire only causes forces between the free current carried by the wire and the induced distribution of magnetization

current running around the surface of the wire. These forces are purely internal and so contribute nothing to the force between the wire itself and the external field. If, however, the medium supporting the wire were a magnetic fluid, then Stratton's result implies that the correct value of the force will be by $B_0 I$ rather than $\mu_o H_0 I$. Since the ratio of these terms is μ_2/μ_o, a greater force is measured in a magnetic medium and the so correct form of the Lorentz force can be confirmed experimentally.

11 References

1. E. T. Whittaker, "A History of the Theories of Aether & Electricity: The Classical Theories" and "The Modern Theories 1900-1926" (both volumes bound as one), Dover Classics, 1990
2. J. C. Maxwell, "A Dynamical Theory of the Electromagnetic Field", Trans Roy. Soc, **155**, pp. 459-512
3. J. C. Maxwell, "A Treatise on Electricity and Magnetism", 2 vols., Clarendon Press, Oxford, Unabridged 3rd edn., 1891 & 1998

4. J. C. Oersted, “Experiments on the Effect of a Current of Electricity on a Magnetic Needle”, Annales de Chimie et de Physique, **14**, 1820, pp417-425; also in Annals of Philosophy, **16**, 1820, pp273-277

5. A-M. Ampere “Notice sur les Experiences de Mm Oersted, Ampere, Arago and Biot”, Annales des Mines…, **V**, 1820, pp535-558, reproduced in “Receuils d’Observations Electrodynamiques…”, Crochard, Paris, 1822, available at http://www.ampere.cnrs.fr, pp83-84

6. N. Feather, “Electricity and Matter: an Introductory Survey”, Edinburgh University Press, 1968

7. W. Benenson, J. W. Harris *et al*, “Handbook of Physics”, Springer, NY, 2003

8. L. Page and N. I. Adams, ”A Proposed Reformulation of the Electromagnetic Equations and Revision of Units“, Journal of The Franklin Institute, **218**, No 5, Nov 1934, pp. 517-531

9. J. A. Stratton, “Electromagnetic Theory”, McGraw-Hill, NY, 1941

10. O. Heaviside, “Electrical Papers”, McMillan, London, 1892, reprinted by Chelsea, NY, 1970

11. D. H. Menzel D H, “Mathematical Physics”, Dover, NY, 1961

12. J. C. G. Field, “Electromagnetism and Semantics”, IEE Proc. A, **135**, No 8, pp523-527, Nov 1988

13. J. B. Clark, “Mathematical and Physical Tables” (Revised), Oliver and Boyd, Edinburgh, 1969

14. C. Kittel, “Introduction to Solid State Physics”, John Wiley, NY, 2nd edn., 1954

15. M. Born and E. Wolf, “Principles of Optics”, Pergammon Press, 3rd edn., 1964

16. G. Scharf, “From Electrostatics to Optics”, Springer-Verlag, 1994

17. R. S. Elliott, “Electromagnetics: History, Theory and Applications” IEEE Press NY/ Oxford University Press, Oxford, 1993

18. C. W. Oseen, "The Alternating Effect Between Two Electrical Dipoles and the Rotation of the Plane of Polarization in Crystals and Liquids", Ann. Der Physik, **48**, pp1-56, 1915

19. P. P. Ewald, "The Analysis of Crystal Optics", Ann. Der Physik, **49**, pp1-38, 1916

20. L. Lorenz, “Über die Refractionsconstante” Ann. Phys. Und Chem., **11**, 1881, pp70-103

21. H. A. Lorentz, “Über die Beziehungzwischen der Fortpflanzungsgeschwindigkeit des Lichtes der Körperdichte,” Ann. Phys. Und Chem. **9**, 1880, pp641-665

22. J. D. Jackson, “Classical Electrodynamics”, John Wiley, NY, 1962

23. J. Hendry, “James Clerk Maxwell and the Theory of the Electromagnetic Field”, , Adam Hilger Ltd, Bristol, 1986

24. E. U. Condon, “Basic Electromagnetic Phenomena”, in “Handbook of Physics”, eds. E. U. Condon and H. Odishaw, McGraw-Hill, NY, 1967

25. C. Kittel, “Introduction to Solid State Physics”, 3rd edn., John Wiley, NY, 1966

26. C. P. Slichter, “Principles of Nuclear Magnetic Resonance”, p10, Harper & Row, NY, 1963

27. R. M. Eisberg, “Fundamentals of Modern Physics”, p330, John Wiley, NY, 1961

28. J. M. Ziman, “Principles of the Theory of Solids”, p211, Cambridge, 1969

29. L. D. Landau, and E. M. Lifshitz, “The Classical Theory of Fields”, p48, trans. Hamermesh M, Pergamon Press, Oxford, 4th revised English edn., 1975

30. J. Thewlis, ed., “Encyclopaedic Dictionary of Physics”, vol4, p352, Pergamon Press, Oxford, 1961

31. N. E. Burris, “Maxwell’s Equations”, in “The RF and Microwave Handbook”, p9.3, ed Golio M, CRC Press, NY, 2001

32. J. D. Kraus, “Electromagnetics”, 4th edn., pp219 and 239, McGraw-Hill NY, 1992

33. “International Electrotechnical Vocabulary”, IEC 60050, IEV 221-04-16 and 221-04-18, International Electrotechnical Commission, http://www.electropedia.org

34. G. Bate and M. H. Kryder, “Magnetism and Magnetic Fields” in “The Electrical Engineering Handbook”, ch34, p813, ed. R. C. Dorf, CRC Press, 1993

35. A. Kitaigorodsky, “Introduction to Physics”, §103, p272-3, Dover Publications, NY, 1965, trans. O. Smith

36. H. G. Jerrard and D. B. McNeill, “Dictionary of Scientific Units”, p115, Chapman and Hall, 1992

37. L. Page and N. I. Adams, “Electrodynamics”, Dover Publications, NY, 1965, pp518 and 520

38. R. Becker, “Electromagnetic Fields and Interactions”, ed. F. Sauter, trans., A W Knudsen, Volume 1, Blackie, London, 1964

39. R. E. Collin, “Foundations of Microwave Engineering”, McGraw-Hill NY, 2nd edn.

40. A. J. W. Sommerfeld, “Lectures in Theoretical Physics Volume 3: Electrodynamics”, p238, trans. E. C. Ramberg, Academic Press, NY, 1952

41. J. F. Nye, “Physical Properties of Crystals”, Clarendon Press, Oxford, 1972

42. W. F. Brown, “Tutorial Paper on Dimensions and Units”, IEEE Transactions on Magnetics, **MAG20**, No 1, 1984, pp112-117

43. Molière (J. B. Poquelin), “Le Bourgeois Gentilhomme”, 1670

44. E. L. Hill, “Theory of Relativity”, in E. U. Condon and H. Odishaw (eds.), “Handbook of Physics” , Part 2, Chapter 6, p2.46, McGraw-Hill, NY, 1967

45. D. Corson and P. Lorrain, “Introduction to Electromagnetic Fields and Waves”, W H Freeman, SF, 1962

46. L. Page and N. I. Adams, “Action and Reaction Between Moving Charges”, Am. J. Phys, **13**, 1945, pp141-147

47. A. Shadowitz, “The Electromagnetic Field”, McGraw-Hill, NY, 1975

48. H. A. Lorentz, “Electromagnetic Phenomena in a System Moving with any Velocity Smaller than the Speed of Light”, Proc. Acad. Science Amsterdam, **VI**, 1904, pp809-830

49. H. A. Lorentz, “The Theory of Electrons” 2nd edn., Dover NY, 1915

50. G. H. Wannier, “The Deflection of Charged Particles in a Ferromagnetic Medium”,

52. J. L. Anderson, “Principles of Relativity Physics”, ch. 8, p257, Academic Press, 1967

53. E. U. Condon, "Electromagnetic Waves", in E. U. Condon and H. Odishaw, (eds.), "Handbook of Physics", p6.5, McGraw-Hill, NY, 1967

54. J. A. Kong, "Electromagnetic Fields", in "The Electrical Engineering Handbook", *ibid,* ch33, p803

55. J. A. Kong, "Electromagnetic Wave Theory", ch1, Wiley InterScience, NY, 1990

56. P. Penfield Jr, and H. A. Haus, "Electrodynamics of Moving Media", Research Monograph No 40, MIT Press, Cambridge MA, 1967

57. T. M. Helliwell, "Introduction to Relativity", Allyn and Bacon, Boston,1966

58. A. Einstein, "The Meaning of Relativity", Princeton University Press, 5th edn., 1956

59. A. Einstein, "On the Electrodynamics of Moving Bodies", Annalen der Physik, **17**, pp891-921 (see www.fourmilab.ch/etexts/einstein/specrel/www/ for a free to use English Translation)

60. H. Fizeau, "On Hypotheses of the Lumeniferous Ether and a Result which Appears to Demonstrate that the Movement of a Body Changes the Speed of Light Propagating Therein", Comptes Rendus, **33**, 1851, pp349-355

61. J. W. Arthur, "The Equivalence of Electromagnetic Field Solutions Based on Poles and Currents" , to be published

62. L. Page and N.I. Adams, "Electrodynamics", p519, Dover Publications, NY, 1965

63. "The International System of Units (SI)", Bureau International des Poids et Mesures, 8th edn., Paris, 2006

64. International Organization for Standardization, "ISO 1000:1992 SI Units and Recommendations for the Use of their Multiples and of Certain Other Units", Geneva, 1992

65. International Organization for Standardization, "ISO 31-5:1992 Quantities and Units - Part 5 Electricity and Magnetism", Geneva, 1992

66. "Standard Definitions for Selected Quantities, Units, and Related Terms, with Special Attention to the International System of Units (SI)", IEEE Standard 270-2006, IEEE Press, 2006

68. C. Kittel, "Interpretation of Anomalous Larmor Frequencies in Ferromagnetic Resonance Experiment", Physical Review, **71**, pp. 270-271, 1947

69. C. Kittel, "On the Theory of Ferromagnetic Resonance absorption", Physical Review, **73**, pp. 155-172, 1948

70. R. E. Collin, "Field Theory of Guided Waves", §3.8, pp. 110-111, McGraw Hill, NY, 1960

71. J. Helsajn, "Nonreciprocal Junctions and Devices", John Willey, NY, 1975

72. D. P. Ames, "Magnetic Resonance" in E. U. Condon and H. Odishaw, (eds.), "Handbook of Physics" , p. 8.114, eq. 9.16, McGraw-Hill, NY, 1967

73. A. C. Ehrlich, "The Hall Effect", in R. C. Dorf (ed.)"The Electrical Engineering Handbook", p1106, CRC Press, Boca Raton,1993

74. L. Berger and G. Bergmann, "The Hall Effect and its Applications", p55, eds. C. L. Chien and R. Westlake, Plenum Press, NY, 1980

75. D. E. Gray, ed, “American Institute of Physics Handbook”, pp. 5-214 to 5-215, McGraw-Hill, NY, 3rd edn., 1972

76. A. Von Hippel, “Dielectrics” in E. U. Condon and H. Odishaw (eds.), “Handbook of Physics” , Part 4, p4.104, McGraw-Hill, NY, 1967

Introducing the Feature Article Author

John Arthur trained as a physicist, gaining a BSc from the University of Toronto and PhD from Edinburgh University. After initially working on light scattering from lattice vibrations in crystals, he took up a career in 1976 as an engineer, first researching on the applications of CCD FIR filters. He then moved to an industrial post with Thales-MESL Ltd to develop applications for SAW Devices in radar, communications and EW. While there, he also worked in the fields of electronics, signal processing and microwaves. He became a board member in 1990 and Technical Director in 1997, and lead the teams that gained the company a Queen's Award for Technology and a Design Council Millennium Product Award. During this time he was also a council member of the Scottish Optoelectronic Association and served with local universities as an external examiner and a member of industrial advisory boards.

In 2003, he became an independent technical consultant, and in 2005 was appointed an Honorary Fellow in the School of Engineering and Electronics at Edinburgh University.

During his career he has published a number of papers in various areas of physics and engineering and has instigated several patents. His interests as a consultant centre on innovation. A senior member of the IEEE, he was elected as a fellow of the Royal Society of Edinburgh and of the Royal Academy of Engineering in 2002. Dr Arthur is also a fellow of the Institute of Engineering and Technology and the Institute of Physics (UK).

<u>**NOTE**</u>

As result of a problem with the transcription of references from the original manuscript, the following correction for the published article was printed on page 65 of the August issue of IEEE APS Magazine, Volume 50, No 4, 2008.

Please note these corrections do not affect this version at **http://www.JohnWArthur.com.**

Correction

The following changes should be made to the article by John W. Arthur, "The Fundamentals of Electromagnetic Theory Revisited," *IEEE Antennas and Propagation Magazine*, **50**, 1, February 2008, pp. 19-65. In the case of citations, the corrected citation is shown.

Location	**Change**
p. 19, Author's affiliation	Include "University of Edinburgh"
§1.1, p. 20, ¶3, ℓ15	[9; pp. 241-242; 10, vol. 1, pp. 179-181]
§1.2, p. 21, ¶3, ℓ4	[14, pp. 89-95; 15 pp. 84-87 and 100-104; 16, pp. 150-158]
§2.1, p. 23, ¶1, ℓ8	[23, pp. 262-263; 24, p4.11; 14, pp. 155 and 241; 25, p. 503; 26-30; 31, p. 9.3]
§2.12.5, p. 29, ¶2, ℓ8	[39, p.18 and p. 27; 22, p. 153; 45, p. 276]
§2.5, p. 23, ¶1, ℓ5	[10, vol. 2, p. 27]
§2.9.4, p. 26, ¶3, ℓ6	[3, vol. 2, p. 275]
§3.1.1, p. 33, ¶2, ℓ3	[23; 48; 49, p. 14;1, vol. 1, p. 422]
§3.1.1, p. 34, ¶1, ℓ11	[3, vol. 2, p. 240].
§4, p. 40, ¶1, ℓ9	[10, vol. 2, pp. 1-23],
§5, p. 48, ¶2, ℓ3	[22, pp. 380-382; 9, pp. 78-80; 47, pp. 486-495; 56].
§6.6, p. 54, ¶1, ℓ 6 and ℓ 15	[33, IEV 121-11-19]
§6.6, p. 54, ¶1, ℓ8	[33, IEV 121-11-69]
§6.6, p. 54, ¶1, ℓ14 & ¶4, ℓ5	[33, IEV 121-11-56]
§6.6, p. 54, ¶3, ℓ7	[7, p. 453]
§6.6, p. 54, ¶3, ℓ7 and ¶4, ℓ10	[33, IEV 121-11-40]
§6.8, p. 55, ¶1, ℓ3	[67; 45, pp. 70 and 241],
§6.8, p. 56, ¶3, ℓ9	[23, p. 230; 3, vol. 2, pp. 197-198 and 271-276].
§6.13, p. 59, Figure 12, ℓ7	B_{cav} should read B_{int}
§10.3, p. 62, ¶1, "Equation (45)"	Replace with "Equation (46)"
p. 63, references 54 and 73	Remove **34**