Basic Electromagnetic Theory

Basic Electromagnetic Theory

Field Theory Foundations and Structure

JAMES BABINGTON

Mercury Learning and Information

Dulles, Virginia
Boston, Massachusetts
New Delhi

Publisher: David Pallai
Mercury Learning and Information
22841 Quicksilver Drive
Dulles, VA 20166
info@merclearning.com
www.merclearning.com
800-232-0223

This book is printed on acid-free paper.

James Babington. *Basic Electromagnetic Theory.*
ISBN: 978-1-942270-74-4

Library of Congress Control Number: 2016935946

161718321 Printed in the United States of America

CONTENTS

PREFACE

This book is intended as a short introduction to the foundations of electromagnetic field theory. It has been based on my experience both as a former field theorist (working on quantum field theories, quantum electrodynamics and Casimir physics) and currently as an applied optical physicist. Indeed, my thoughts were in the last couple of years to simply write down what I thought I knew and try to critically examine this. What has been very interesting to see is that, in my opinion, an appreciation of its history together with the characters involved is a necessity to get to grips with the foundations of the subject, and therefore understanding at a deeper level.

I was fortunate enough as an undergraduate to be exposed to some quite advanced electromagnetic field theory early on. My then tutor, Robin Devenish, at Hertford College, Oxford, set me some very interesting problems on electromagnetic fields written in the four-vector notation of special relativity at the end of my first year. This was real classical field theory that one would typically study before moving on to quantum fields. It came about because I actually wanted to study the general theory of relativity, and I had asked him how to get into it. In hindsight, these were exactly the right problems to work on over a summer vacation, and I thank him greatly for his direction here.

My prejudice is clearly going to be skewed towards the field theory aspects. However, and importantly, I have not shied away from experimental and measurement issues that ultimately are the reason for introducing a field theory in the first place. In particular, the historical demonstration of a principle, or the quantification of a law takes first place. The mathematics should be there to capture the essence of any experimental studies and provide a language with which to ask further questions. With this in mind, I am assuming on the part of the reader a genuine interest in the subject and present the material from this standpoint.

I have included some problems at the end with solutions; these are meant to illuminate parts of the text with either a definite calculation method, or to just see some numbers. In addition to these,

exercises are scattered throughout the text to give a little poke to the reader to interact.

Finally, I would like to dedicate this book to my wife Serena and my son Albert with the very greatest affection.

James Babington

June 2016

CHAPTER 1

INTRODUCTION

Electromagnetism is perhaps the most basic theory we see around us in the natural world. It is responsible for the color we see, the forces that bind us together and for any technology beyond the steam engine. It is the theory *par excellence* in that it is responsible for the macroscopic world around us that we are part of and interact with. But it goes far beyond this as it is also the gateway to the subatomic world that tries to ultimately explain the existence of matter and forces. From these concepts must emerge the structures such as the atomic nuclei that are necessary to get any form of life going. Modern particle physics theories that try to explain the nature of quarks and the very small, and the evolution and dynamics of the universe on the other, take as their starting point something that looks quite similar to electromagnetism (that is the central role of gauge theories and symmetry as a governing principle).

One of the main thrusts of this book was to see how the theory of electromagnetism is (or should be) consistent, together with making no *a priori* assumptions about the nature of matter and its interactions. In particular, I have taken the opposite approach from that of the infamous Bleaney and Bleaney [5] and abandoned the modern atomic viewpoint from the outset. Obviously one has to have good reason for doing this and it is the following question; how can you be sure that what you are measuring is the genuine physical variable and not something that requires some extra theory on top or refers back to itself? This is an important point because I suspect many seeming inconsistencies that one encounters in advanced work would simply

not appear if the corresponding measurement theory had been correctly setup and applied. All this being said, there is obviously great merit in taking the atomic viewpoint on good faith from scratch as it speeds up the time to be able to perform calculations and get to grips with more material. Both [8] and [5] are classic works on electromagnetism and essential reading; the current book simply tries to aim at a macroscopic and foundational level, and should therefore be considered as complementary material.

The outline of this book is as follows. In the current chapter, we describe in fairly general terms field theory with a certain amount of flowery language to try and give the bigger picture. Following on from this in Chapter 2, we list in an overview like way the necessary math that will be used later on throughout the book. We do not prove theorem's here, but rather state them. The form is somewhat terse but gives the notation and conventions used throughout. It gives the necessary language with which to be able to understand the latter chapters.

Chapter 3 derives the first two of Maxwell's equations. Both the electric and magnetic fields are introduced and then an equation of motion for each is deduced. This is a constructive chapter and aims firstly to start the construction of a field theory, and secondly, to show the parallel between electric and magnetic phenomena.

Chapter 4 formulates the remaining Maxwell equations, together with the continuity equation and the Lorentz force law for dynamic charges moving in background fields. Currents are included, both the more standard electric current and the displacement current deduced by Maxwell from the consistency of the field theory. With these in place on has the governing dynamics of the theory.

Chapter 5 describes the attributes of the fields from a mathematical standpoint. This involves solutions to the wave equations in different coordinate systems. From here one finds that polarization is a necessary consequence to keep track of the genuine degrees of freedom. It also surfaces then with the realization that what we are dealing with is a gauge theory and so gauge transformations and fixing are discussed.

Chapter 6 gives an account of what we can measure and the corresponding quantities in the field theory. This is a complicated topic,

but essential understanding for the foundations of electromagnetism as well as applications (such as photometry, polarimetry, etc).

In Chapter 7, we discuss how fields interact with real matter. The two distinct aspects here how charges assemble themselves into multipoles and how the dynamics of these give rise to response functions. A summary and omissions chapter then follows.

I have included short exercises throughout the text as a way for the reader to try and assist assimilation. These are not difficult problems, rather they are just meant to fill in the gaps typically with the derivation of a result. At the end of the book there are some problems with worked solutions. These are a bit more difficult and cover some material that was not included in the main chapters.

A word of caution about notation used here is also appropriate. I have found over the years that one tends to switch between notation, and this is certainly what the student will encounter in skipping between different books and papers. My advice on this front is that it should be clear from the context in which the equations are presented. Thus one should not be alarmed at seeing x, $\mathbf{x}$, x^a all meaning the same position vector because it depends on the setting which one gives the best aesthetic. Likewise, seeing $\mathbf{E}(x)$ written down shouldn't lead the reader to think the field is only a function of the x Cartesian coordinate. Common sense should prevail here and if there is any consternation, the best solution is to write down oneself the offending expression in a more pleasing form.

1.1 GENERAL REMARKS

Field theory is a complicated business. The original conception due to Michael Faraday and the later mathematical synthesis due to James Clerk Maxwell has remained fairly well intact for over a century. In its current form, the electromagnetic field equations serve as the basis for investigating any micro or nano structure, at the applied level, and is the DNA blueprint for modern theory of the sub-atomic world.

The impression one gets, certainly for school level (in particular A-level in the UK), is that electromagnetism is quite a neat and tidy

affair, and that the subject doesn't require any further investigation (save for in applications). I would like to challenge this view, that instead not just because of its history, but because as you start to appreciate its foundations, it is richer but more complex.

There are couple of points that I would like the reader to have at the back of the mind while using this book. One is of measurables and the other is of disturbances. Both of these are strongly interconnected and cannot really be considered by themselves. Since this is a quantitative science, it is obviously necessary to be able to calculate quantities that can then be measured. In fact this is the purpose of the theory in some real sense (this depends very much philosophical school one has subscribed to). The measurables will therefore require a set of devices that are sensitive to and can respond accordingly to the quantity being measured. This means that there is coupling between the device and the electromagnetic field quantity being measured that necessarily has a certain amount of extra structure and assumptions built into it. Clearly this will disturb the original configuration, which may well be small, but there is necessity to be able to quantify this. Admittedly, this is a difficult programme to pursue rigorously and a heuristic approach will often be required to make progress. Nevertheless, for those with the predisposition, I would like to advocate a rigorous approach as a topic for further research. My hope with the current book is that it can stimulate thoughts in this direction, as well as introducing the basic subject matter at an introductory level.

A closing comment concerns the history of the subject. Obviously with a history as long as electricity and magnetism, this is going to be varied, full and sometimes confusing. It seemed a shame not to give some of the original protagonists a little bit more character. To this end, I have made some historical inclusions where I felt it illuminating or worthy acknowledgement that fits the general development of the topics addressed herein.

CHAPTER 2

MATHEMATICAL TOPICS

In this first chapter, we will provide essentially a look up table of mathematical formulae and concepts that feature prominently throughout the book. It will consist of basically a definition or a statement about mathematical quantities that are needed in the description and manipulation of the field equations. The key topics are vector calculus, a bit on Fourier transforms, and coordinate systems. A few side topics, such as group theory, are also included to provide some extra language for some of the math we will encounter later on. This is intended only to broaden the vocabulary of the reader. In fact each of these topics sit in a more general theory, which has in part been borrowed from. For example, vector calculus and coordinate systems naturally lie in differential geometry, the mathematical structure that for one underpins general relativity. The objective of this chapter is therefore to outline the notation used throughout the following chapters and to collect together relevant mathematical results. At the end we provide a table of some useful numbers and notational descriptions that are fairly standard. See [15] for good introductory material and [2] for a more sophisticated treatment.

2.1 SUMMARY OF NECESSARY MATHEMATICAL FORMULAE

We will focus on here fields that live naturally on our normal three dimensional space. All of the results generalize straightforwardly to

an n-dimensional space. A *scalar field* is defined simply as some function $f(x)$, where $x \in \mathbb{R}^3$ and $f(x) \in \mathbb{R}$. A *vector field* is defined as a vector valued function $\mathbf{V}(x)$, or in terms of its three components $\mathbf{V}^a(x)$. The three components can be specified with respect to an orthonormal basis of vectors (or for that matter any vectors that span the space), $\mathbf{e}_a(x)$, such that

$$\mathbf{V}(x) = \sum_{a=1}^{3} \mathbf{V}^a(x)\mathbf{e}_a(x) \coloneqq \mathbf{V}^a(x)\mathbf{e}_a(x), \tag{2.1}$$

where in the last line the Einstein *summation* convention been adopted. This means that when there is one upstairs and one downstairs index that are the same letter, they are summed over (also known as being contracted). We will assume in this book that the orthonormal basis vectors are self evident and from the stand point of performing operations on vector fields, simply make the identification

$$\mathbf{V}(x) \leftrightarrow \mathbf{V}^a(x). \tag{2.2}$$

This is just a notational convenience. There is an important distinction between an upstairs index and an downstairs index. An upper index is referred to as being a *tangent* vector index (or simply a vector index), while a downstairs index is referred to as *covector* index. Note that this is true for the components rather than the basis vectors index.

If an object has more than one index then it is referred to as a *tensor* field and in general will look like $T^{abc\cdots}_{de\cdots}(x)$.

Two important second rank tensors that are the underpinnings of all operations are the *Kronecker* delta and the *metric* tensor. They are defined in the following way

The Kronecker Delta δ^a_b:

$$\begin{aligned} \delta^a_b &= 1, \quad \text{if } a = b, \\ &= 0, \quad \text{if } a \neq b. \end{aligned} \tag{2.3}$$

This can be thought of as an identity matrix (and in our case is 3×3).

The metric tensor g_{ab}:

$$ds^2 = g_{ab}dx^a dx^b, \tag{2.4}$$

where ds^2 is the square distance of a small line element and dx^a are small distances in the respective coordinate. A metric is then a tensor field that allows one to compute distances between points. Thinking of this also as a matrix allows us to define its inverse, g^{ab} with now two indices upstairs,

$$g^{ac}g_{cb} \coloneqq \delta^a_b. \tag{2.5}$$

The key property of the metric and its inverse is that it is a machine that allows one to raise or lower indices on any tensor field.

The scalar product between two vector fields is defined by

$$\mathbf{V}(x) \cdot \mathbf{W}(x) \coloneqq \mathbf{V}^a(x)\mathbf{W}_a(x) \tag{2.6}$$

$$= g_{ab}(x)\mathbf{V}^a(x)\mathbf{W}^b(x). \tag{2.7}$$

Note that the scalar product only makes sense when the two vector fields are evaluated at the same point.

The *divergence* of a vector is defined as

$$\nabla \cdot \mathbf{V} \coloneqq \nabla_a \mathbf{V}^a, \tag{2.8}$$

where the gradient operator is given by the partial derivatives $\nabla_a = (\partial/\partial x^1, \partial/\partial x^2, \partial/\partial x^3)$. More exactly, this type of derivative is referred to as a covariant derivative, which means it should not act on any factor of the metric (the metric is said to be covariantly constant). So the partial derivative has to promoted to accommodate this. This would be relevant for example if you wanted to do your calculus on the surface of a sphere.

The wedge (or cross) product of two covector fields is given by

$$\mathbf{V} \wedge \mathbf{W} \coloneqq \mathbf{V}_a \mathbf{W}_b \epsilon^{abc} \mathbf{e}_c, \tag{2.9}$$

and the curl of a covector field is defined as

$$\nabla \wedge \mathbf{V} \coloneqq \nabla_a \mathbf{V}_b \epsilon^{abc} \mathbf{e}_c, \tag{2.10}$$

where ϵ_{abc} is the *Levi-Civita tensor*. It is defined by

$$\epsilon_{abc} = -\epsilon_{bac} = \epsilon_{bca}, \tag{2.11}$$

$$\epsilon_{abc} = 1, \quad \text{for } a = 1, b = 2, c = 3 \text{ are the same.} \tag{2.12}$$

Note that this definition requires that the Levi-Civita tensor has its indices raised and therefore involves three factors of the inverse metric. It is an unfortunate accident of three dimensions that the Levi-Civita tensor provides one with a map from what appear to be vectors to vectors, i.e. we take the curl of a vector and get another vector. As with the covariant derivative, there is more taking place. The derivative of a covector field gives a two index tensor (with indices downstairs) and then contracting this with the Levi-Civita tensor (with indices) upstairs (known as taking the Hodge dual in differential geometry) leads to a tensor with one index upstairs, i.e. a vector. This is why we seemingly get another vector quantity.

Considering a two index tensor $\mathbf{F}_{ab}$ for simplicity, a tensor is said to be symmetric if

$$\mathbf{F}_{ab} = \mathbf{F}_{ba}, \tag{2.13}$$

and antisymmetric if

$$\mathbf{F}_{ab} = -\mathbf{F}_{ba}. \tag{2.14}$$

The metric tensor is a symmetric tensor ($g_{ab} = g_{ba}$), while the Levi-Civita is totally anti-symmetric on its indices. If a tensor satisfies such conditions, then the number of independent components is reduced. For example a second rank tensor has $3 = 9$ independent components, whereas a symmetric tensor has $3 \times 2 = 6$ independent components.

2.1.1 Some Basic Group Theory

The ideas of group theory are so useful that it is worth collecting here a few basic concepts and definitions. This will add clarity in particular to some of the mathematical issues encountered with Maxwell's equations.

A group is defined as a set of objects G on which we can impose a rule combining pairs of elements $a, b \in G$ to form other objects $c = a \cdot b \in G$. This rule of combination must satisfy certain axioms to make it useful. It might be helpful to keep in mind rotations of a body in three dimensional space. The set of all such is then the group of rotations. The axioms are:

1. Combing elements is an associative operation, which means that $\forall\, a, b, c \in G$ it is such that $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
2. For all $g \in G$ there exists a unit element $e \in G$ defined by $g \cdot e = e \cdot g = g$.
3. There exists for every $g \in G$ an inverse element $g^{-1} \in G$ defined by the property that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

As an example of a group, the group of rotations in three dimensions is written as $G = SO(3)$ which means in words "special orthogonal transformations in three dimensions". Elements of this group are simply 3×3 matrices R that perform a rotation on a position vector. The orthogonal bit means that they satisfy the constraint that any rotation preserves the length of the position vector and the special bit discounts reflections (or parity transformations). Another group that occurs in electromagnetism is the group $G = U(1)$ which is the one dimensional unitary group. This is an example of an abelian group, where the order of operations is irrelevant, while $SO(3)$ is a non-abelian group; performing two rotations about different axes is different depending on which one is done first. This has the technical name of commutation. The rotations just described are examples of operators if we start thinking more generally. For two operators $\mathcal{O}$ and $\mathcal{P}$, one can form another operator given by the *commutator* that is defined by

$$[\mathcal{O}, \mathcal{P}] := \mathcal{O}\mathcal{P} - \mathcal{P}\mathcal{O}. \tag{2.15}$$

Two operators are said to commute if $[\mathcal{O}, \mathcal{P}] = 0$. So abelian groups have commutators that are zero, and non-abelian groups have nonzero commutators. One last remark; the groups we are considering here are all continuous groups rather than discrete groups (an example of a discrete group are the permutations of a set of numbers).

2.1.2 Operators and Eigenvalue Business

In general, an operator is anything that acts on another object, for example a partial derivative on a function or a matrix on a vector. The eigenvalues λ_n and eigenvectors (or eigenfunctions) V_n of an operator $\hat{\mathcal{O}}$ are defined by the relation

$$\hat{\mathcal{O}} V_n = \lambda_n V_n. \tag{2.16}$$

Eigenvalues are useful in connection with group theory because they can be used to label how the group theory is realized at the practical level of finding solutions (this goes by the name of representation theory). In turn, the eigenvectors are the indivisible units that the group theory forbids from turning into one another.

The eigenvalues and eigenvectors (or eigenfunctions) satisfy very neat relations in perhaps the most useful case when the operator is *Hermitian.* An Hermitian operator is given by the condition $\hat{\mathcal{O}}^\dagger = \hat{\mathcal{O}}$, where the dagger indicates the Hermitian conjugate operation. In this the eigenvalues are real $\lambda_n \in \mathbb{R}$. If the eigenvectors and functions are normalized to unity, then they both satisfy the orthonormality conditions

$$(V_m)^* \cdot V_n = \delta_{mn}, \tag{2.17}$$

for eigenvectors and

$$\int_a^b dx (V_m(x))^* \cdot V_n(x) = \delta_{mn}, \tag{2.18}$$

for eigenfunctions defined on the interval $[a, b]$. Not for both functions and vectors, the concept of vector space can be applied to both which the allows eigenvalue problems to be viewed from a unified perspective.

2.1.3 Differential Vector Calculus

We now list a set of identities that are useful in the simplification of vector calculus equations. For scalar fields $f(x)$ and $g(x)$ and vector fields $\mathbf{V}^a(x)$ and $\mathbf{W}^a(x)$, there are a set of one derivative - two field relations

$$\nabla(fg) = f(\nabla g) + g(\nabla f), \quad \text{(Leibnitz property)} \tag{2.19}$$

$$\nabla(f\,\mathbf{V}) = f\nabla\cdot\mathbf{V} + \mathbf{V}\cdot\nabla f, \tag{2.20}$$

$$\nabla\wedge(f\,\mathbf{V}) = f\nabla\wedge\mathbf{V} + (\nabla f)\wedge\mathbf{V}, \tag{2.21}$$

$$\nabla\wedge(\mathbf{V}\cdot\mathbf{W}) = \mathbf{V}\wedge(\nabla\wedge\mathbf{W}) + (\mathbf{V}\cdot\nabla)\mathbf{W} + \mathbf{W}\wedge(\nabla\wedge\mathbf{V}) + (\mathbf{W}\cdot\nabla)\mathbf{V}, \tag{2.22}$$

$$\nabla\cdot(\mathbf{V}\wedge\mathbf{W}) = \mathbf{V}\cdot(\nabla\wedge\mathbf{W}) - \mathbf{W}\cdot(\nabla\wedge\mathbf{V}), \tag{2.23}$$

$$\nabla\wedge(\mathbf{V}\wedge\mathbf{W}) = \mathbf{V}(\nabla\cdot\mathbf{W}) - \mathbf{W}(\nabla\cdot\mathbf{V}) + (\mathbf{W}\cdot\nabla)\mathbf{V} - (\mathbf{V}\cdot\nabla)\mathbf{W}. \tag{2.24}$$

There are also a useful set of two derivative - one field relations

$$\nabla\cdot(\nabla f) = \nabla^2 f, \tag{2.25}$$

$$\nabla\wedge(\nabla f) \equiv 0, \tag{2.26}$$

$$\nabla\cdot(\nabla\wedge\mathbf{V}) \equiv 0, \tag{2.27}$$

$$\nabla\wedge(\nabla\wedge\mathbf{V}) = \nabla(\nabla\cdot\mathbf{V}) - \nabla^2\mathbf{V}. \tag{2.28}$$

An interesting point to note is that in the above the resulting expressions will depend explicitly on the metric used.

One can express the total time derivative of a function in terms of partial derivatives and the instantaneous velocity

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{dx^a}{dt}\nabla_a. \tag{2.29}$$

2.1.4 Integral Vector Calculus

It is necessary to be able to integrate over different types of surfaces. These are one dimensional curves, two dimensional surfaces, and three dimensional volumes. Therefore a knowledge of

the different integration measures is required so that one can switch between different coordinates and metrics with ease.

The length of the curve $\mathcal{C}$ is obtained from the line element in Equation (2.4) by

$$L = \int_{\mathcal{C}} ds = \int_{\mathcal{C}} \sqrt{g_{ab} dx^a dx^b}. \tag{2.30}$$

It is in this form that the metric can be seen as the means by which distances between points are calculated.

The following integration measures are defined with respect to the Levi-Civita tensor,

$$d^3x := \frac{\sqrt{g}}{3!} \epsilon_{abc} dx^a dx^b dx^c, \tag{2.31}$$

$$d^2\Sigma_a := \frac{\sqrt{h}}{2!} \epsilon_{abc} dx^a dx^b, \tag{2.32}$$

where the tensor h_{ab} is obtained from the metric g_{ab} by restricting it to the two-dimensional subspace where the area integration is to be carried out. This may be as simple as choosing one of the coordinate values to be a constant value.

There are two key integral relations for vector fields. They allow one to transform between volume and surface integrals, and separately between surface integrals and line integrals when there are derivatives involved. Because of the different dimensionality of the integration measure, derivatives of the vector fields feature in these relations

For a volume $\mathcal{V}$ that is bounded by the surface $\partial\mathcal{V}$ the *divergence theorem* is given by

$$\int_{\mathcal{V}} d^3x \nabla_a \mathbf{V}^a(x) = \int_{\partial\mathcal{V}} d^2\Sigma_a \mathbf{V}^a(x). \tag{2.33}$$

For an area $\mathcal{A}$ bounded by the loop $\partial\mathcal{A}$, *Stokes's theorem* is given by

$$\int_{\mathcal{A}} d^2\Sigma_a \cdot (\nabla \wedge \mathbf{V}(x))^a = \int_{\partial\mathcal{A}} dx^a V_a(x). \tag{2.34}$$

2.1.5 Functions, Distributions, and all Things Fourier

Next we consider some basic properties of functions and distributions. This is necessary because having written the field equations from the previous vector calculus techniques, we need to be able to analyze and extract information out of them.

A *Fourier series* is an infinite sum of trigonometric functions given by the following expression

$$f(x) = \sum_{n=0}^{\infty} a_n \sin(nx) + b_n \cos(nx), \tag{2.35}$$

where $n \in \mathbb{Z}$. In particular, the $f(x)$ here will be a periodic function, so that for some L, $f(x+L) = f(x)$.

A *Fourier transform* of a function in one dimensional space, say $f(x)$ where $x \in \mathbb{R}$, to another function in reciprocal space (given by k) is defined by

$$F(k) = \int dx e^{ikx} f(x) \tag{2.36}$$

$$f(x) = \int \frac{dk}{(2\pi)} e^{ikx} F(k). \tag{2.37}$$

The Fourier transform is the continuum version of a Fourier series, where we let the integer that defines the summation turn into a continuous variable, i.e. $n \to k$. It is a simple matter to perform a Fourier transform in three dimensions (or indeed in any number of dimensions). The above then become

$$F(k) = \int d^3x e^{ik\cdot x} f(x) \tag{2.38}$$

$$f(x) = \int \frac{d^3k}{(2\pi)^3} e^{ik\cdot x} F(k). \tag{2.39}$$

The *Dirac delta* function, $\delta^3(x)$, in three dimensions is defined with respect to an arbitrary function as

$$f(x') := \int_{-\infty}^{\infty} d^3x \delta^3(x' - x) f(x). \tag{2.40}$$

It is not really a function but rather a distribution which means that it should be sat inside an integral. It is used so often that one cannot really do without it. An integral representation of the the delta function in three dimensions is given by

$$\delta^3(x) = \int_{-\infty}^{\infty} \frac{d^3k}{(2\pi)^3} e^{ik\cdot x}. \tag{2.41}$$

The convolution of two functions $f(x)$ and $g(x)$ is defined by

$$f(x) * g(x) := \int dy\, f(y) g(x-y). \tag{2.42}$$

Using Fourier transforms in this definition results in the convolution theorem; if $F(k)$ and $G(k)$ are the Fourier transforms of $f(x)$ and $g(x)$, then their convolution satisfies the simple multiplicative property

$$\int dx e^{ikx} f(x) \star g(x) := F(k)G(k). \tag{2.43}$$

Two useful expressions for what are Green functions of the Laplace equation are

$$\nabla_a \left(\frac{1}{|x-x'|} \right) = -\frac{x^a - x^{a\prime}}{|x-x'|^3}, \tag{2.44}$$

$$\nabla^2 \left(\frac{1}{|x-x'|} \right) = -4\pi\delta^3(x-x'). \tag{2.45}$$

$$\nabla_b (1/|x-x'|) = -\nabla'_b (1/|x-x'|) \tag{2.46}$$

An integral representation of the above Green function is given by

$$\frac{1}{|x-x'|} = \frac{-1}{4\pi} \int \frac{d^3k}{(2\pi)^3} e^{ik\cdot(x-x')} \frac{1}{k^2}. \tag{2.47}$$

A function $f(x)$ is said to be *square-normalizable* if

$$\int_{-\infty}^{+\infty} d^3x\, |f(x)|^2 < \infty. \tag{2.48}$$

This shows up time to time when it is necessary to ensure that physical fields are not ending up with infinite energy.

2.2 COORDINATE SYSTEMS AND TRANSFORMATIONS

The vectors and tensor quantities encountered so far require us to specify both a set of coordinates x^a which labels points in the space we are working in and a vector space with which a tensor can be expanded in. The components of a tensor transform under a coordinate transformation according to how differentials or derivatives transform by application of the chain rule of differentiation. For example, the line element dx^a can serve as a basis of orthogonal vectors (if the metric is diagonal) in which a co-vector field can be expanded in. In the new coordinates x'^a we have the transformation law

$$dx'^a = \frac{\partial x'^a}{\partial x^b} dx^b. \tag{2.49}$$

Since we have an upstairs index here, this serves as the transformation law for a vector field

$$\mathbf{V}'^a = \frac{\partial x'^a}{\partial x^b} \mathbf{V}^b. \tag{2.50}$$

In a similar fashion, a general coordinate transformation on a derivative operator is given again by the chain rule as

$$\nabla'_a = \frac{\partial x^b}{\partial x'^a} \nabla_b. \tag{2.51}$$

The lower index here gives now the transformation law for a co-vector field as

$$\mathbf{V}'_a = \frac{\partial x^b}{\partial x'^a} \mathbf{V}_b. \tag{2.52}$$

Given then a particular coordinate transformation the transformation matrix $\partial x^b / \partial x'^a$ can then be calculated explicitly. While this gives a set of vectors that are at least locally orthogonal, it remains to normalize them to unity in order to obtain an orthonormal basis. The majority of expressions encountered at least at a basic level are expressed in an orthonormal basis.

Standard Cartesian coordinates are given by (x, y, z) together with their line element. With our numbering convention these are given by

$$x^1 = x, \tag{2.53}$$
$$x^2 = y, \tag{2.54}$$
$$x^3 = z. \tag{2.55}$$

The line element is just the familiar one based on Pythagoras's theorem

$$ds^2 = g_{ab}dx^a dx^b = dx^2 + dy^2 + dz^2. \tag{2.56}$$

The Laplacian operator ∇^2 is given by

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}. \tag{2.57}$$

The position vector $\mathbf{x}$ can be written in terms of an orthonormal basis as

$$\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z. \tag{2.58}$$

The components of the wedge product with respect to the orthonormal basis $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ are

$$(\nabla \wedge \mathbf{V})_x = \frac{\partial V^z}{\partial y} - \frac{\partial V^y}{\partial z}, \tag{2.59}$$

$$(\nabla \wedge \mathbf{V})_y = \frac{\partial V^x}{\partial z} - \frac{\partial V^z}{\partial x}, \tag{2.60}$$

$$(\nabla \wedge \mathbf{V})_z = \frac{\partial V^y}{\partial x} - \frac{\partial V^x}{\partial y}. \tag{2.61}$$

Spherical coordinates (r, θ, ϕ) are defined by the coordinate transformation from Cartesian coordinates by

$$x^1 = r\sin\theta\cos\phi, \tag{2.62}$$

$$x^2 = r\sin\theta\sin\phi, \tag{2.63}$$

$$x^3 = r\cos\theta, \tag{2.64}$$

with $r \geq 0, 0 < \theta < \pi, 0 < \phi, 2\pi$. Notice that there is a degeneracy in the coordinates at the origin; at $r = 0, \theta$ and ϕ can be anything and still label the same point. The corresponding line element is

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2. \tag{2.65}$$

Following from this the Laplacian operator is given by

$$\nabla^2 = \left(\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \right). \tag{2.66}$$

The components of the wedge product with respect to the orthonormal basis $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ are

$$(\nabla \wedge \mathbf{V})_r = \frac{1}{r\sin\theta}\frac{\partial(\sin\theta V^\phi)}{\partial\theta} - \frac{1}{r\sin\theta}\frac{\partial V^\theta}{\partial\phi}, \tag{2.67}$$

$$(\nabla \wedge \mathbf{V})_\theta = \frac{1}{r\sin\theta}\frac{\partial V^r}{\partial\phi} - \frac{1}{r}\frac{\partial(rV^\phi)}{\partial r}, \tag{2.68}$$

$$(\nabla \wedge \mathbf{V})_\phi = \frac{1}{r}\frac{\partial(rV^\theta)}{\partial r} - \frac{1}{r}\frac{\partial V^r}{\partial\theta}. \tag{2.69}$$

Cylindrical coordinates (r, θ, z) are defined by another coordinate transformation from Cartesian coordinates by

$$x^1 = r\sin\theta, \tag{2.70}$$

$$x^2 = r\cos\theta, \tag{2.71}$$

$$x^3 = z, \tag{2.72}$$

with $r \geq 0, 0 < \theta < 2\pi, 0 < z < \infty$. The line element is given by

$$ds^2 = dr^2 + r^2 d\theta^2 + dz^2. \tag{2.73}$$

The Laplacian operator is then

$$\nabla^2 = \left(\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}\right). \tag{2.74}$$

The components of the wedge product with respect to the orthonormal basis $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z)$ are

$$(\nabla \wedge \mathbf{V})_r = \frac{1}{r}\frac{\partial V^z}{\partial \theta} - \frac{\partial V^\theta}{\partial z}, \tag{2.75}$$

$$(\nabla \wedge \mathbf{V})_\theta = \frac{\partial V^r}{\partial z} - \frac{\partial V^z}{\partial r}, \tag{2.76}$$

$$(\nabla \wedge \mathbf{V})_z = \frac{1}{r}\frac{\partial (rV^\theta)}{\partial r} - \frac{1}{r}\frac{\partial V^r}{\partial \theta}. \tag{2.77}$$

These are the most often used coordinate systems in electromagnetism, but there are others. It largely depends on the problem at hand that is trying to be solved.

2.3 PHYSICAL CONSTANTS AND USEFUL NUMBERS

In this last section, we list a number of physical constants, useful numbers, and some mathematical notation used throughout. The standard SI units are used whenever a numerical calculation is performed.

Physical Constant, Property/Notation	**Symbol**	**Value**
Speed of light	c	$2.998 \times 10^8 \mathrm{ms}^{-1}$
Permittivity of free space	ϵ_0	$8.854 \times 10^{-12} \mathrm{C}^2\mathrm{m}^{-2}\mathrm{N}^{-1}$
Permeability of free space	μ_0	$4\pi \times 10^{-7} \mathrm{NA}^{-2}$
Elementary electric charge	e	$1.602 \times 10^{-19} \mathrm{C}$
Gravity of Earth	g	$9.81 \mathrm{ms}^{-2}$
Earth's magnetic field at equator	$\lvert\mathbf{B}\rvert_E$	$3 \times 10^{-5} \mathrm{T}$

Physical Constant, Property/Notation	Symbol	Value
Intensity of solar radiation at Earth's surface	$I^{\odot}(\oplus)$	$\approx 1.0\text{kWm}^{-2}$
Typical refractive index of glass	$n_{typical}$	≈ 1.5
Typical laboratory electric field	$\mathbf{E}_{typical}$	$\approx 10^{4}\text{Vm}^{-1}$
Typical laboratory magnetic field	$\mathbf{B}_{typical}$	$\approx 10\text{T}$
Optical frequency	$f_{opt.}$	$\approx 10^{15}\text{Hz}$
The integers	$\mathbb{Z}$	$0, \pm 1, \pm 2, \cdots$
The real numbers	$\mathbb{R}$	$1.3, -2.567, \pi, \cdots$
The complex numbers	$\mathbb{C}$	$1 + 3i, e^{i\pi/4}, \cdots$

CHAPTER 3

MAXWELL'S EQUATIONS PART I

In this chapter, we will think about the formulation of two of the field equations from the larger set collectively known as "Maxwell's equations". They are Gauss's law and the magnetic version which is sometimes referred to as Gauss's magnetic law. This is where the electric field $\mathbf{E}^a$ and the magnetic flux density $\mathbf{B}^a$ will make their first appearances. The electric field is intertwined from the outset with the notion of electric charge, while the magnetic flux density has no such counterpart. These quantities are introduced as a way of accounting for observed static forces $\mathbf{F}^a$ on macroscopic bodies. For the electric field it is due to acquiring a net charge by some means; for the magnetic flux density it is more complicated.

3.1 THE LAW OF COULOMB - ELECTRIC CHARGE, FORCE, AND BASIC MEASUREMENT

Let us consider from first principles a number of observations and experiments that lead to the idea of electric charge and the electric field. Much of this may be considered to be somewhat elementary material. It is interesting, however, to have an historical perspective and to clearly delineate the foundations of the subject. Ultimately we must be able to trace back any useful definitions or

theoretical abstractions to a set of basic physical measurables. For an illuminating and useful discussion of this history and the basic set of experiments see [11].

Experiment 1 (Greek antiquity - circa 1700s) *Take two bodies (for example glass and rubber) that are initially in mechanical equilibrium. Next apply frictional forces to their surfaces with a third material (for example rub them with a silk cloth or a piece of fur). What happens at some later time to the two bodies?*

Observation 1 *At some later time they are no longer in equilibrium and will feel attractive or repulsive forces* $\mathbf{F}^a$ *to one another by some action at a distance agent.*

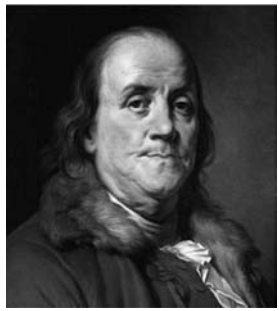
Normally to produce a mechanical force on a body there would have to be some physical contact for this to happen. This is not the case in the above (the same is also true for gravity - much of historical electrostatics took its inspiration form Newtonian gravity).

Thales of Miletus (ca 624 BC–547 BC), according to Bertrand Russell, "Western philosophy begins with Thales", and so does the path to electromagnetism. Some of the earliest observations of static electricity were due to Thales (around 600 BC). The basic observation in this epoch was that the fossil material amber would attract objects when it had been rubbed by wool.

This first experiment and observation gives us the idea that some quantity can be transferred between bodies and depending on how much and its type will lead to different amounts of observable forces. We are thus led to the quantity which we call the *electric charge* of a body which might be in the form of discrete point like units that are sat at particular points, or distributed continuously in some fashion throughout a body given by a charge density. A key observation of this behavior was made by Benjamin Franklin. He observed a fundamental book keeping principle taking place which we now call the *law of conservation of charge:-*

Principle 1 - *In any system, electric charge is neither created nor destroyed but is simply redistributed amongst the subsystems.*

Reconsidering our first experiment where the frictional forces cause charge to be transferred from one body to the other, we see that each must have an equal but opposite amount to satisfy the conservation of charge (given that the net charge at the start was zero). Note also that we haven't specified anything about the dynamics of how the charge moved, or what type of quantity it is (one should think scalar, vector, tensor… here).



Benjamin Franklin (1706–1790), the quintessential polymath, dabbling in a huge number of areas such as printing, authoring, science, and political theory, as well being President of the State of Philadelphia and one of the founding fathers. It is amazing he had any time to do any experimental physics at all.

It is now necessary to put the observed forces on a more quantitative footing by systematically varying the previously described systems. The basic parameters are the amount of charge on the two bodies, their separation and shapes and also the types of material they are made of. What we will be able to measure then is the macroscopic force on the bodies that was previously encountered qualitatively. This was investigated by Coulomb in the late 1700s which resulted in the now well known Coulomb's law of electrostatics:-

Experiment 2 (Coulomb) *Taking two spherical bodies (for simplicity), vary the separation between them and the amount of the quantity transferred in the previous experiment.*

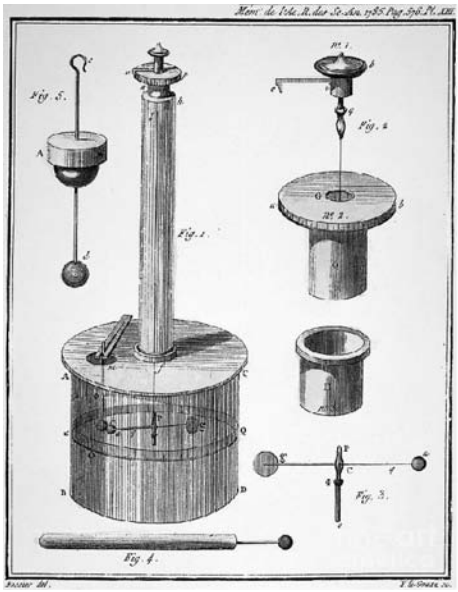
Observation 2 *The force between them is found vary as the inverse square of the separation and is proportional to the amount of the quantity residing on/in the body due to the previous transfer process.*

A necessary point to appreciate here is how this force can be measured. One way to measure the electric force on a body is by

balancing it with a known mechanical force. We can also do the same balancing with torques if the geometry of the apparatus is arranged appropriately. The form originally used by Coulomb is just such a system, the apparatus being known as a torsion balance (see Figure 3.1).

Charles-Augustin de Coulomb (1736–1806), as a Captain returning to France from the West Indies, Coulomb discovered the inverse square law for charged bodies (1785). In addition to this he also established for conductors that the charge resided on the surface.

This method was used earlier by Cavendish to measure gravitational forces between bodies. It comprises of a bar suspended at its midpoint by a thin fiber. At one end is a body which can be charged



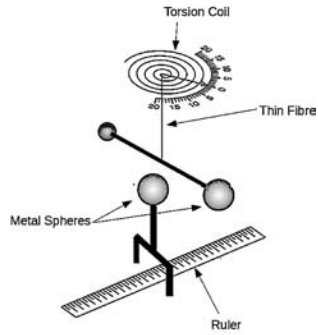


FIGURE 3.1: **Coulomb's torsion balance experiment (1785)** The left picture is taken from Coulomb's 1785 memoirs of the French academy of sciences. The right figure shows the principles of operation, as described in the main text.

to a known reference level. One such way is to charge an identical body outside and then bring the two into contact - by geometry the charge will distributed evenly between the two and we have thus defined the reference point. When a second charged body is brought into proximity repulsion or attraction results. The bar twists and because the fiber has a weak spring constant one can obtain a large deflection angle of the bar, thus ensuring that weak forces map to large easily observed angles. When the torque due to the electric force balances the restoring torque of the fiber, attaining an equilibrium, we can read off the observed force by the angular deflection. Therefore the torsion balance physically measures the force and from this an amount of charge can be deduced, with respect to the reference level.

It is straightforward to codify Coulomb's law mathematically. Suppose we have two bodies labelled by I and J and that they have positions $x^a(I)$ and $x^a(J)$. By positions we either mean their center of mass, for finite size objects, or regard them as point like objects when their separation is much larger than their respective length scales. It is observed that charge is a scalar quantity - varying the orientation has no effect on the mutual force. Then Coulomb's law states that the force on charge q_I due to charge q_J is given by

$$\mathbf{F}^a(I\,|\,J) = \left(\frac{q_I q_J}{4\pi\epsilon_0}\right) \frac{x^a(I) - x^a(J)}{|\,x(I) - x(J)\,|^3}. \tag{3.1}$$

A constant of proportionality has been introduced and we need to think about the units of measure and how more generally we should measure the electric charge on a body. In fact this turns out to be a subtle issue because we will in addition need to introduce a dynamical element. Clearly, both the observed forces and separations are honest to goodness measurable quantities. If we can identify the correct symmetry in a problem that has been setup with this in mind, we will be able find the ratio of the charges. But ultimately the observed forces will have to be made with respect to the undetermined constant ϵ_0, which is called the *permittivity* of free space. If charge was infinitely divisible we would need to define a basic unit, whereas if (as is the actual case for electrons and protons) it is found to have a minimum discrete unit, we can use this as our basic

natural unit of charge. The upshot is that, at this point, we will not ascribe to ϵ_0 a particular number and return to its numerical value later on.

A natural question to ask is what is the effect of a number of charged bodies on one another. Let us do another experiment.

Experiment 3 *Suppose we introduce other electrically charged bodies in addition to the previous two already considered. What is the effect on one of the bodies?*

Observation 3 (The principle of linear superposition) *After this introduction and a long enough time for all the bodies to settle down, the resultant force on the body is the vector sum due to all the other bodies.*

In connection with this principle, it is useful to introduce the idea of a *test charge* as this is what we were alluding to in the discussion of Coulomb's law. Suppose we have a number of bodies all charged in some static configuration. If we introduce another body that doesn't change the existing configuration, we then have a measure of the local force felt there; it doesn't disrupt the configuration and only feels the measurable force acting on it. We will assume that for any macroscopic configuration of bodies considered, a test charge can always be found.

Building on Coulomb's law, it is necessary to think about how to measure electric charge on bodies in more detail and quantitatively. To that end, we consider some very basic material properties and make a statement about the types of materials involved, as part of their definition and grouping relies on the movement of charge (which will anticipate the following chapter, in particular the consequences of currents). We will not worry at this point about the details of how charge moves apart from saying that it does. A material is said to be a *conductor* if charge can move freely in it and an *insulator* if it is impeded and so remains static. For insulators, charge can be built up in local regions on a body whereas for conductors the charge is redistributed on the surface of the body. The notion of the movement of charge is a very important one as it is the mechanism by which systems can find new electrostatic equilibria when we change the system. Of course, this is a sliding scale; we know charge does move and can expect different materials to display

differing amounts of resistance to charge movement. The definitions introduced are simply a convenient division. These definitions result from how different materials are able to transfer charge from one another e.g. for two conductors bringing one charged into contact with the other uncharged will result in the total charge being shared easily between the two. We lastly define *ground* to be the infinite reservoir of charge that when a body is connected by a conductor to the earth (terra firma), the charge flows from the body to the reservoir thus becoming charge neutral - this relies on the earth therefore being conductive.

One can measure the electric charge on a body by using an electrometer or an electroscope. At its simplest level an electroscope consists of a strip of metal foil that is hung over a metal bar which is in turn connected to the outside world. Charge is transferred onto the foil which can then move upwards symmetrically until the electrostatic force is balanced by the downward gravitational force. The angle of separation is then a measure of the electric force and since the separation is known, we can calculate the charge. An electrometer on the the other hand is often a mechanical device similar in spirit to the torsion balance that requires an auxiliary store of charge to operate. In Figure 3.2, a quadrant electrometer is shown. Its principle feature is that the charge can be determined to a high degree of accuracy.

We perform one last experiment before getting down to some math and collectively building up a scheme with which to perform calculations. It was mentioned earlier that for conductors with a net charge, all the charge resides on the surface. Given that it is now possible to measure electric charge, we are now in a position to observe this.

Experiment 4 *Suppose we lower a charged metal sphere suspend by an insulating thread into an uncharged hollow conductor that is insulated from the ground. What happens to the hollow conductor?*

Observation 4 *It is found that charge is induced onto the surface of the hollow conductor but not into its interior.*

This basic observation was originally found by a number of scientists (Franklin, Coloumb, and Faraday) in slightly different ways. An electrometer is connected to the outside of the metal container

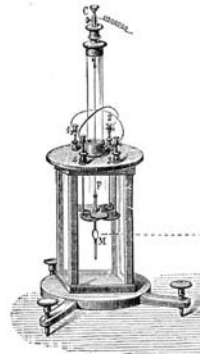


FIGURE 3.2: **The quadrant electrometer** developed by Lord Kelvin in the 1860s. An aluminium vane with a known amount of charge (that is stored via a Leyden jar at the bottom of the apparatus) is suspended by a thin fiber which has a mirror attached. The vane in turn is suspended inside a circular metal box like structure that is divided into quadrants. Each quadrant is connected to the diagonally opposite one with a conductor so that they will have the same charge. One pair of quadrants are positively charged, the remaining two are grounded. As the quadrants are charged due to an external source that is being investigated, the vane rotates and therefore so does the mirror. Shining a beam of light off the mirror provides a measure of the deflection action thus establishing the amount of charge on the sample.

as in Figure 3.3. As the sphere is lowered inside, the electrometer registers a charge. This persists when the sphere touches the inside of the container. Taking the sphere out, one finds now that it no longer has any charge on it. The inside of the container also has lost its charge, whereas the outside still retains its original charge which is equal to the original charge on the sphere. Experimentally then we have the basic fact that for conductors, any charge must reside on its outer surface when the system is in equilibrium. The interested reader can find in [17] a good commentary about these points. We

FIGURE 3.3: **Faraday's ice pail experiment**, for demonstrating that electrostatic charge resides on the surface of a conductor, not in its interior. It consists of a metal container into which a charged sphere is lowered. The metal container is connected to an electrometer, shown on the right.

will return to this point later on in this chapter when we are thinking in terms of fields.

3.2 THE LAW OF GAUSS - THE ELECTRIC FIELD AND SOME BASIC MATH

Forces are what we can measure, as previously seen, but what exactly is there located at the point where the charge *is*? A conceptual leap is required here to assert that the other charged body produces *something* at the point where the original charge is located, regardless of whether the charge is there or not. A coupling is therefore taking place between the electric charge (which we have already said is a scalar quantity) and a vector quantity due to the other charged body, such that one just multiplies the other. We call this vector quantity the electric field $\mathbf{E}^a$ and it is given by the defining relation

$$\mathbf{F}^a(x) = q\mathbf{E}^a(x), \tag{3.2}$$

for a discrete charge q located at x. We can extend this to a collection of N charges by use of the linear superposition principle, so that one has by simple summation the total force given by

$$\mathbf{F}^a := \sum_{I=1}^{N} \mathbf{F}^a(x(I)) = \sum_{I=1}^{N} q_I \mathbf{E}^a(x(I)). \tag{3.3}$$

This readily extends to a continuous distribution of charge as

$$\mathbf{F}^a|_{\mathcal{V}} = \int_{\mathcal{V}} d^3x \rho(x) \mathbf{E}^a(x), \tag{3.4}$$

where $\rho(x)$ is the charge density of a body of volume V. These are the defining relations for the concept of the electric field. There is an interesting interplay between the continuous or discrete nature of the electric charge[1] which will depend on the length scale we are choosing to do our measurements or perform calculations. For example, if we are working with discrete charges that have some characteristic separation d, and working on a length scale $L >> d$ then we can treat the discrete set of charges as a continuum i.e. $q_I(x(I)) \to \rho(x)$.

The use of the test charge introduced earlier should now be apparent. It provides a local measure of the electric field strength at a point, since it will not disrupt the local field there. Taking the idea of a test charge further, let us assume we have some charge on a small insulating ball. Now imagine we make the radius smaller and smaller and we decrease the charge on there. Either we will reach a practical limit or a fundamental limit to how many times this reduction can be done[2]. When this is done Coulomb's law seems to hold at each reduction. This is fortunate as we are now able to give expressions for the electric field in terms of its source charge. By using Equation (3.1) together with the linear superposition principle, the field at x is then

$$\mathbf{E}^a(x) = \sum_{I=1}^{N} \frac{q_I}{4\pi\epsilon_0 |\, x - x(I)\,|^3} (x^a - x^a(I)). \tag{3.5}$$

It is straightforward to convert this to a continuum version as well

$$\mathbf{E}^a(x) = \int_{\mathcal{V}} d^3x' \frac{\rho(x')}{4\pi\epsilon_0 |\, x - x'\,|^3} (x^a - x^{a\prime}). \tag{3.6}$$

The reader will notice just how similar these expressions are to what is encountered in Newtonian gravity. In fact the static electric field is also a *central* force field and it is easily verified that it satisfies

$$\nabla \wedge \mathbf{E} = 0, \quad \text{in vector notation,} \tag{3.7}$$

$$\epsilon_{abc} \nabla^b \mathbf{E}^c = 0, \quad \text{in components.} \tag{3.8}$$

(Exercise: verify this)

As we will see in the next chapter, this only holds for static fields.

Together, Equations (3.6) and (3.7) constitute the equations of *electrostatics.* To solve these equations we have to specify for a given system its geometry and the boundary conditions the field must satisfy. We can write Equation (3.7) also in integral form

$$\int_{\mathcal{A}} d^2\Sigma \cdot (\nabla \wedge \mathbf{E}) = 0, \tag{3.9}$$

$$\int_{\mathcal{A}} d^2\Sigma^a \epsilon_{abc} \nabla^b \mathbf{E}^c = 0, \tag{3.10}$$

for a two dimensional area $\mathcal{A}$ (with the topology of a disk). This can be converted to a line integral by the use of Stokes theorem Equation (2.34) such that

$$\oint_{\partial\mathcal{A}} dx^a \mathbf{E}_a = 0. \tag{3.11}$$

From the definition of the electric field, we see that this is a statement about the work done in moving our test charge around a closed loop. In this case no work is done, and in particular, leads to the result that the work done by the electrostatic force between two points is path *independent* for a central force. It is a mathematical result that any such vector field can be written as the gradient of a scalar function so that the electric field can be recast as

$$\mathbf{E}_a(x) = -\nabla_a \phi(x). \tag{3.12}$$

(Exercise: verify this is path independent)

The scalar function $\phi(x)$ is known as the electrostatic *potential*. This quantity will be important when we start thinking about voltages as an observable. The reader is more likely to be familiar with the idea of voltages from standard school physics and electronics courses.

Having introduced the concept of an electric field, we can now turn the question around and ask for a statement about how the electric field is produced or sourced by other electric charges. We know already that it is due to the body having a net charge of the same type as the first. That they should be proportional is also evident. From Coulomb's observation of an inverse square law, one can see that if we were to surround the source body with a sphere that was centered on where the source body was, then the product of the area of the sphere and the size of the electric field should be a numerical constant. In fact it should also be proportional to how the charge is distributed on the source body. This is all encapsulated in *Gauss's law*.

Carl Friedrich Gauss (1777–1855), one of the strongest mathematicians of the 19th century (he was named the "Prince of Mathematicians"). He worked on a huge number of different areas in mathematics and physics. Gauss published this particular result in 1838; however it had already been discovered earlier by George Green in 1828 and was later rediscovered by Lord Kelvin in 1842.

We define the electric flux Φ_E through a closed two dimensional surface ∂V which is the boundary of the volume $\mathcal{V}$ to be:-

$$\Phi_E := \int_{\partial \mathcal{V}} d^2\Sigma_a \mathbf{E}^a. \tag{3.13}$$

In integral form, Gauss's law states that the electric flux for a closed surface area that encloses a certain amount of charge is proportional to the enclosed charge:-

$$\int_{\partial \mathcal{V}} d^2\Sigma^a \mathbf{E}_a = \frac{1}{\epsilon_0} \sum_{I=1}^{N} q^I. \tag{3.14}$$

The constant ϵ_0 is the permittivity of free space mentioned earlier so that when we consider a single point charge and its electric field at some distance we get back Coulomb's law Equation (3.1).

(Exercise: verify this)

Since it is an integral equation we are making a global statement about the system. However, we can also make a local statement (that is, recast the equation so that it holds at each point in space) by using one of our useful integral relations. The two-dimensional surface integral can be converted to the volume integral by the divergence theorem Equation (2.33) whence,

$$\int_{\mathcal{V}} d^3x \nabla_a \mathbf{E}^a = \frac{1}{\epsilon_0} \int_{\mathcal{V}} d^3x \rho(x). \tag{3.15}$$

We have now replaced here the discrete charge summation with a continuous distribution of charge density integrated, as in earlier

examples. By dropping the integrals Gauss's law becomes *local*, so it has a differential form. One finds

$$\nabla_a \mathbf{E}^a(x) = \frac{\rho(x)}{\epsilon_0}, \tag{3.16}$$

$$\nabla \cdot \mathbf{E}(x) = \frac{\rho(x)}{\epsilon_0}. \quad \textbf{(Maxwell I)}. \tag{3.17}$$

Gauss's law is the first of the four of Maxwell's equation and we shall refer to it throughout the text as **Maxwell I**. We can also derive Gauss's Law simply by taking the divergence of Equation (3.6) and using Equations (2.44) and (2.46) so that

$$\begin{aligned}
\nabla_a \mathbf{E}^a(x) &= \int_v d^3x' \nabla_a \frac{\rho(x')}{4\pi\epsilon_0 |x - x'|^3}(x^a - x'^a) \\
&= -\frac{1}{4\pi\epsilon_0}\int_v d^3x' \nabla_a \rho(x') \nabla^a \frac{1}{|x - x'|} \\
&= -\frac{1}{4\pi\epsilon_0}\int_v d^3x' \rho(x') \nabla^2 \frac{1}{|x - x'|} \\
&= \frac{\rho(x)}{\epsilon_0}.
\end{aligned} \tag{3.18}$$

Note also from the nature of the experiments and observations previously discussed that we are dealing initially with *static* phenomena; the bodies that have been subjected to forces via electric charges on one another have all been allowed to "settle down" after having been disturbed and so we do not see any time dependence in our equations.

Two important equations follow by the use of the scalar potential that is valid when dealing with static phenomena. If we use the result Equation (3.12), one finds that the scalar field $\phi(x)$ satisfies by virtue of Equation (3.17)

$$\nabla^2 \phi(x) = -\frac{\rho(x)}{\epsilon_0}, \tag{3.19}$$

which is known as the *Poisson equation*. This is for a region of space where there is a distribution of charge. In fact, this equation can be

inverted to express the potential as a function of charge distribution. It results in a scalar form of Equation (3.6) by combining it with Equations (2.44) and (3.12) to give

$$\phi(x) = \frac{1}{4\pi\epsilon_0}\int_v d^3x' \frac{\rho(x')}{|x - x'|}. \tag{3.20}$$

If there is no charge present such that $\rho(x) = 0$ then

$$\nabla^2\phi(x) = 0, \tag{3.21}$$

which is called the *Laplace equation*. The solution of these two second order partial differential equations is one of the key aspects in studying electrostatic phenomena.

It is worth while putting a picture to the ideas of electric vector fields and charges that act as sources. Figure 3.4 shows how iron filings align themselves between a positive and negative charge and also two positive charges. The filings serve to illuminate (not literally) the direction of the electric vector field at a particular point in space (it doesn't say anything about its magnitude because this would require some type of movement to show the force). It is conventional to attach an arrow to one of these field lines that points away outward for a positive charge and inwards for a negative charge. A field line therefore either starts on a charge or terminates on it. The corresponding case for magnetism is quite different!

As a final remark as concerns Figure 3.4, there is a neat relationship with the scalar potential. Suppose we consider an equipotential surface given by $\phi(x) = 0$. Then, from the definition of

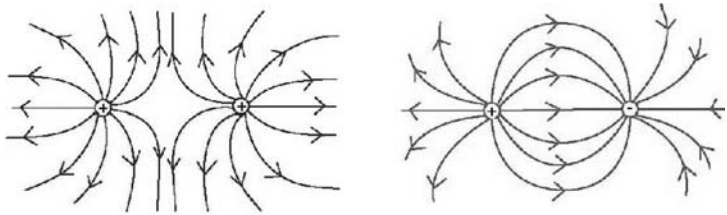


FIGURE 3.4: A pictorial representation of the electric field in the region of two charges.

the electric field as a gradient of the scalar potential, the electric field lines must be orthogonal to the equipotential surfaces. To see this, first take the differential of the scalar potential, which is $d\phi(x) = (\nabla_a \phi) dx^a$. Now, on an equipotential surface $d\phi(x)$ must also equal zero and dx^a must lie in the local area element of the equipotential surface. Therefore ∇_ϕ must be normal to the equipotential surface. In this way in Figure 3.4 we could draw in lines that cut field lines normally.

3.3 THE "MAGNETIC" LAW OR THE LAW WITH NO NAME

At this point we have arrived at the idea of electric charges that produce and couple to electric fields. These then do the same in return. These systems are therefore always coupled and the interaction cannot be turned off[3]. We now turn our attention to the phenomena of magnetism. Some materials (e.g. iron or magnetite) when suspended in free air will naturally align themselves in a preferential direction directed from the North to the South pole along a great circle. This basic property is the principle by which the magnetic compass operates, a piece of technology that has been around as a navigational aid since the 11th century. A material that shows this alignment along a great circle is said to be *magnetized*. It is also possible to magnetize a body which previously did not display any magnetization (that is, for those materials which can display magnetization). This leads to the following experiment:-

Experiment 5 *Take two magnetic bodies in the form of iron bar magnets. For a single bar magnet label one end positive and the other negative so that the positive points to the North Pole and the negative to the South Pole. Now bring them into close proximity. What happens to the bar magnets some time later?*

Observation 5 *The bar magnets ends either repel if a positive (negative) end from one bar is close to a positive (negative) of the other end, or they attract one another if the ends are a positive and a negative. That is like sign ends repel while unlike ends attract.*

This looks very much like the phenomena encountered earlier in electrostatics (at least qualitatively) where the positive or negative ends corresponds to the the two types of electric charge. Indeed, looking at Figure 3.5 that shows a bar magnet, one sees that the fields lines displayed by the iron filings look quite similar to the electric field lines between the two equal and opposite electric charges shown in Figure 3.4. A natural question arises now as to whether we can isolate the corresponding "charges" at either end of the bar magnet, such that we might try and build up a picture similar to that of electric charges. Consider the following experiment.

Experiment 6 *Take one such bar magnet and divide it into two by cutting it across its axis where the positive and negative charges reside. What happens?*

Observation 6 *We find two new (smaller) bar magnets have been created. Where we have performed the cut we find a positive and negative charge on either side that gives rise to the two new bar magnets.*

No matter how many times we subdivide the bar magnet we will continually find smaller and smaller bar magnets. We are of course trying to investigate magnetism in an analogous way to the electrical forces previously found. Having introduced an electric charge earlier we have just found that their appears to be no comparable "magnetic" charge for magnetized bodies[4]. However, they do appear to share a common form of interaction between themselves.

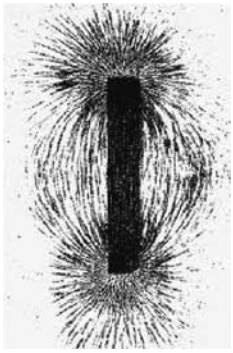
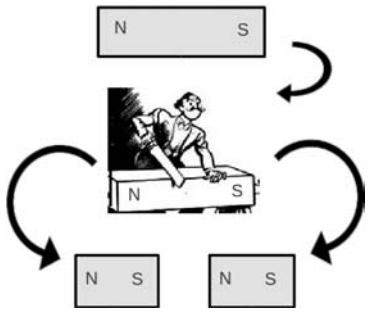


FIGURE 3.5: Iron Filings lining up around a bar magnet. The filings show the local direction of the magnetic field in the vicinity of the magnet's poles.



It's frustrating cutting bar magnets.

Given that the idea and introduction of an electric field proved so useful beforehand to account for observable static forces, it is clear we should then do something similar for magnetism. To this end we introduce a magnetic field $\mathbf{B}^a(x)$ that a bar magnet will couple to such that it feels a force and aligns itself with. Not only does it couple to the magnetic field but it is also a source of a magnetic field. Thus the Earth produces a magnetic field, which we can regard as a background field in a similar way to the bar magnets, albeit on a much larger scale. Note also that while the electric charge is a scalar quantity, for the bar magnet a vector type charge can be associated connecting the north and south poles.

It is necessary now to setup some basic field theory that is an equivalent of Gauss's law applicable to magnetic bodies. We define therefore the magnetic flux as

$$\Phi_M := \int_{\partial \mathcal{V}} d^2\Sigma_a \mathbf{B}^a. \tag{3.22}$$

In the electric case the electric flux through a closed surface was found to be proportional to the enclosed charge. Since there appears to be no free magnetic charge the magnetic flux must be zero

$$\int_{\partial \mathcal{V}} d^2\Sigma_a \mathbf{B}^a = 0. \tag{3.23}$$

To see this geometrically take the bar magnet shown in Figure 3.5 and surround it by a sphere. For every small area on the sphere where the magnetic field pierces it pointing outwards, there is another small area somewhere else where the magnetic field pierces

the sphere pointing *inwards*. As before, the divergence theorem can be applied to this, resulting in an integral over the volume. Clearly the differential version is just

$$\nabla_a \mathbf{B}^a(x) = 0. \quad \textbf{(Maxwell II)} \tag{3.24}$$

(Exercise: verify this)

We shall refer to this as **Maxwell II** and is the second of the four Maxwell's equations. On an historical note as the title of this section suggests, Equation (3.24) does not have a name as the others do. It seems Maxwell wrote it down first based on earlier work of Lord Kelvin and Faraday's idea of closed magnetic lines of force [6, 17]. It might better be called the Maxwell-Kelvin-Faraday law, though this is a bit of a mouthful.

At this juncture, we have introduced the idea of two distinct fields and one source. **Maxwell I** tells us that a charge density $\rho(x)$ is a source of the electric field and that it will feel a force due to the presence of an electric field there. The magnetic field however does not have a single source term but rather a positive and negative end separated by some distance. To account for the observed forces on electrically or magnetically charged bodies, an electric field has to vanish on a charge; there is a definite single termination point. A magnetic field always comes as loop; there is no start nor end point.

3.4 VOLTAGE AND CAPACITANCE

It is now possible to define two basic quantities for conductors that are already probably familiar to the reader, but are essential in any form of measurement theory and the construction of simple electronic circuits. They are *voltage* and *capacitance*.

Definition 1 *The voltage V of a conductor is the potential there with respect to some reference point or space*.

For example, if we think of an electrical circuit where wires are essentially straight lines, the voltage is the potential difference between two points, typically with a component in between. Compare this with

the case of a charged metal sphere where we often choose the difference to be between the surface of the metal sphere and a sphere at infinity. The potential can also be measured by use of the electrometer encountered earlier that was used to measure charge; it acts as a local probe of the potential at some point. This follows because having determined the charge on a body, we just need to measure the distance to the local probe to find the potential. We assume in addition that we can identify the traditional term voltage associated with a battery with this potential difference and will therefore regard them as synonymous. The reader can consult [17] and [11] for more on its equivalence and historical finding. From now on we will take the voltage to be a macroscopic physical measurable that can used to measure other quantities. Now let us turn to the capacitance.

Definition 2 *The capacitance of a body is the ability of a conductor or collection of conductors to store electrical charge. It is defined as the ratio of the electrical charge on a conductor to the voltage on it. One writes it schematically as* $C \sim Q/V$.

For the case of a single conductor this is a constant that depends on its size and geometry. It is evident, however, that for a more general situation of many bodies it is more complicated. This is because the introduction of a new body (remember it is not a test body!) disrupts the previous electrostatic equilibrium; all the charges are shifted around as the new body is introduced and a new electrostatic equilibrium has to be found. For example, suppose we have some conductor with a charge on it and we know what it's equilibrium potential is. Now bring a second conductor into proximity with the first body. Charges will be induced in the second. Obviously it has required work to separate the charges, which can only come from the potential of the original body. Therefore the potential must have lowered, thereby increasing the capacitance since the charge on it has remained constant. Saying it in slightly different way, for a collection of conductors imagine we put known amounts of charge on them and then electrically isolate each one. Clearly there is potential due to this distribution of charge and this will depend on the overall geometry of the system. We have already encountered an early form of capacitor in connection with the electrometer. The Leyden jar employed there was used for the simple purpose of storing a known amount of electric charge - this was the original form

of the capacitor [11]. To get an idea of the form a capacitance takes, consider the following simple example.

Example 3.1 *A solid metal sphere of radius R carries a charge Q on it. Calculate the capacitance of the sphere.*

Since this problem involves spherical geometry we clearly need to work in spherical polar coordinates. We can apply **Maxwell I** to this in the form of Equation (3.14). By spherical symmetry the electric field only has a radial component and is given by

$$\mathbf{E}_r = \frac{Q}{4\pi\epsilon_0 r^2}, r > R, \tag{3.25}$$

$$\mathbf{E}_r = 0, \leq R. \tag{3.26}$$

From this we can calculate the potential difference between the surface of the sphere and a sphere at infinity,

$$V_R - V_\infty = -\int_\infty^R dr\mathbf{E}_r \tag{3.27}$$

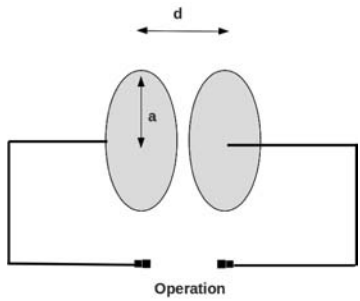
$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{R} - \frac{1}{\infty}\right) \tag{3.28}$$

$$= \frac{Q}{4\pi\epsilon_0 R}. \tag{3.29}$$

The capacitance is therefore

$$C = \frac{Q}{V_R - V_\infty} = 4\pi\epsilon_0 R. \tag{3.30}$$

Another important example that we would like to start thinking about is the parallel plate capacitor in the form of two disks each of radius a separated by a small distance d. What is the capacitance of this system? It is in fact quite a difficult problem to calculate exactly because the edge of the disks cause a complication in the math. One should note here that this system will play an important part for other aspects of the field theory we are investigating. We will return to it again and again as there is really a lot of physics taking place under the hood.



The standard geometry of the parallel plate disk capacitor. It consists of two metal disks, each of radius *a* and a separation *d*. Each has been connected to a wire and ends on some type of module labelled as "Operation" This could be a simple switch, a battery cell or something else.

Notwithstanding the above, let us calculate an expression for the capacitance (per unit area) in the limit that $d << a$. This amounts to saying that we have two infinite plates and that on each we have a constant charge density σ. One plate has $+\sigma$ the other has $-\sigma$. What then is the potential difference between the two plates? Let the separation d be in the z coordinate of the standard Cartesian coordinates so that the plates lie in the x, y plane. By symmetry the electric field is constant - using **Maxwell I** one finds

$$E_z = +\frac{\sigma}{2\epsilon_0}, \quad \text{for plate 1,} \tag{3.31}$$

$$E_z = -\frac{\sigma}{2\epsilon_0}, \quad \text{for plate 2.} \tag{3.32}$$

Using the linear superposition principle the electric field in the two regions *outside* the plates cancel, whereas for the region *in between* the fields add together. The potential difference between the two surfaces is

$$V_d - V_0 = -\int_0^d dr 2\mathbf{E}_z \tag{3.33}$$

$$= \frac{\sigma}{\epsilon_0}(d-0) \tag{3.34}$$

$$= \frac{\sigma d}{\epsilon_0}. \tag{3.35}$$

From this it is straightforward to derive the capacitance per unit area as

$$C = \frac{\sigma}{V_d - V_0} = \frac{\epsilon_0}{d}. \tag{3.36}$$

This is a useful expression to bear in mind as it gives us at least a first approximation to the finite geometry disk case. The complications set in for finite geometries because of fringing effects at the boundaries and typically this will require a numerical type of solution.

3.5 SUMMARY

We have seen in this chapter that to account for observations of static forces between bodies, charges and fields have had to be introduced. By measuring a mechanical force we can in turn measure an electric charge on another body with respect to some known reference amount. The introduction of a scalar potential together with Gauss's law turns electrostatics into a precise mathematical theory where one needs to solve either the Poisson equation in the presence of static charges, or the Laplace equation in free space. For magnetic interactions we have found a physical field but not a simple equivalent of electric charge. This would have made life simpler (or at least more symmetric) as one would perhaps like to have a similar situation to the electric case. Standard references on electrostatics are; [8] very complete and rigorous; [5] useful but with a condensed matter/atomic bent; and [7], always a joy to read and full of many interesting side lines. For an historical point of view on **Maxwell II**, the reader can consult [17] as a starting point.

NOTES

[1]Of course we know that there is a lower limit where electric charge becomes discrete at the atomic level and that it can not be further reduced. The oil drop experiment due to Millikan gives a clear demonstration of this but relies on extra structure so far not introduced. We are just trying to keep each matter logically separate

[2]This follows on from the previous point. The answer to this in our enlightened age is well established and standard school textbook facts that are committed to memory. We simply wish to keep things as macroscopic and bottom up as possible to begin with.

[3]One can see this is a difficult path to tread; if we can't turn the interactions off, can we be sure we are measuring the right thing? And does the very act of measuring the system disturb the system too much such that we do not obtain genuine results? Not surprisingly there is a large overlap here with quantum mechanics.

[4]This in particular makes Maxwell's equations asymmetrical with regard to their matter content; we shall discuss this more later.

CHAPTER 4

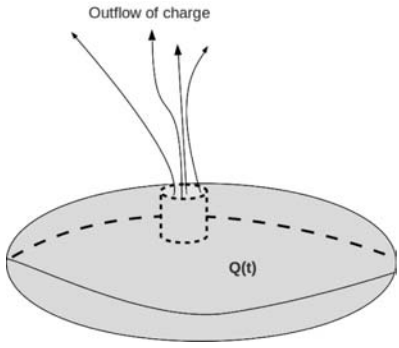
MAXWELL'S EQUATIONS PART II

We have encountered the two fundamental physical fields $\mathbf{E}^a$ and $\mathbf{B}^a$ that arise in static experimental situations. They are constructs that are used to quantify the forces found in electric and magnetic situations after they have settled down into an equilibrium state. The electric charge distributions and magnetic bars have been static and therefore by definition, not been moving. A natural question to ask is exactly this :- what if they do move in some simple fashion (constant velocity or constant acceleration)? By studying the nature of this time dependent phenomena, we will arrive at the astonishing fact that the electric and magnetic fields are intertwined. Not only that they are really two sides of the same coin. As one varies in time and space there is always the other present and can never be switched off. It is in this sense the topic is truly called electromagnetism. Originally electricity and magnetism were considered separate phenomena and it was only with the discovery and introduction of the galvanic cell that currents were linked to magnetism and unity was brought forward. For some historical background on this point, see [11] and [6].

4.1 THE CONSERVATION OF CHARGE

In the previous chapter, we encountered the law of conservation of charge. It is a book keeping device that demands that a final

system in some state must have the same charge as the system in its initial state. To formulate this more generally we need to think about the movement of charge which naturally leads to the idea of a *current*. We will write it in the form of what is known as a continuity equation. Consider a volume $\mathcal{V}$ bounded by the surface area $\partial\mathcal{V}$ and within which sits a certain amount of charge $Q(t)$ at time t. A current $I(t)$ is simply the flow of charge per unit time through some surface $I(t) = dQ(t)/dt$. This could consist of a set of discrete charges moving on their own trajectories or a continuum where there is a flow at each point. Suppose this charge now starts to move about. How much charge will there be in $\mathcal{V}$ at some later time? What we need to think about is the rate of change of charge $dQ(t)/dt$ within the volume as a function of time. Let us also cut up the boundary $\partial\mathcal{V}$ into lots of small areas $\Delta\Sigma_a$. If we know the velocity $\mathbf{v}^a(t, x)$ of a small amount of charge $\Delta Q(t)$ in a small volume ΔV that has part of its boundaries as one of the areas, then we can calculate the outflow of charge through this area.



A volume $\mathcal{V}$ with a boundary $\partial\mathcal{V}$ encloses a charge $Q(t)$ which is a function of time. The dashed cylinder indicates the small volume ΔV inside the main volume and just below the boundary surface. One of the cylinders end faces (with area $\Delta\Sigma_a$) forms a small part of the boundary.

We define a *current density* $\mathbf{J}^a(t, x)$ to be the flow of charge per unit area per unit time. Then the dynamic statement of the conservation of charge is

$$\frac{dQ(t)}{dt} = -\int_{\partial\mathcal{V}} d^2\Sigma_a \mathbf{J}^a(t, x). \tag{4.1}$$

Given a velocity vector field $\mathbf{v}^a(t, x)$ and a charge density we can write an expression for the current density in terms of these variables, that is

$$\mathbf{J}^a(t,x) = \rho(t,x)\mathbf{v}^a(t,x). \tag{4.2}$$

(Exercise: check you agree with the dimensions of this expression.)

As with the earlier **Maxwell's I and II** equations, this can be written in differential form when the charge distribution is continuous $Q(t) \to \rho(t,x)$

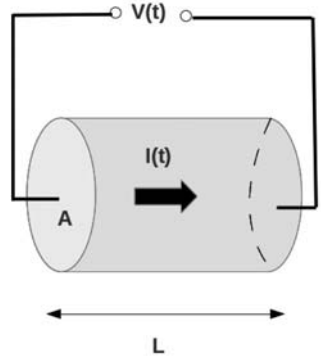
$$\begin{aligned}\frac{d}{dt}\int_{\mathcal{V}} d^3x\rho(t,x) &= -\int_{\partial\mathcal{V}} d^2\Sigma_a \mathbf{J}^a(t,x) \\ &= -\int_{\mathcal{V}} d^3x \nabla_a \mathbf{J}^a(t,x) \\ &\Rightarrow \\ \partial_t \rho(t,x) + \nabla_a \mathbf{J}^a(t,x) &= 0.\end{aligned} \tag{4.3}$$

This is the equation of continuity and although it is applying here to electric charge, it also shows up in many other areas of physics (notably fluid dynamics). A further point that we will discuss in the next chapter is to do with symmetry. The conservation of charge has a mathematical consequence that the field equations have a symmetry built into them.

How does one generate a current in the first place? There are two straightforward ways that follow from its definition. One is simply to move away from a charged object at a constant velocity. Relative to you the charge will thus appear to be moving. The second is the use of a simple battery cell. This involves some chemistry that converts chemical energy into electrical energy in the form of a sustained potential difference at two terminals - negative charge builds up at one end, positive charge at the other [11]. As long as there is a potential difference in a region of space then we know that a charge will couple to the local electric field and feel a force. It will therefore be in a local state of acceleration. Although discharging a capacitor does indeed produce a current, it is not very useful because it manifests itself as a sudden burst of uncontrolled charge movement.

At this juncture we are interested in steady currents meaning constant velocities so we must think about a further element. In defining the electrical properties of materials we defined an effective

division between conductors and insulators. Of course this boundary is not sharp and there is a whole spectrum of behavior. Depending on the material, the charge will be subjected to different amounts of resistive forces when a potential difference is applied across two regions. When the two balance out, we will attain a steady flow and hence a steady current. It is useful therefore to consider how the electrical current and voltage are related. The relationship considered here is an effective description that we are establishing phenomenologically as a useful way of characterising materials which in turn can be used in applying electromagnetism to other systems. This is because a priori we do not know what happens on a small length scale. The basic idea is that we want an equation of motion for the electric charge, but not to keep track of each individual charge. Since this type of scenario invariably involves wires, the equation of motion should be for the current, rather than an individual electric charge, so that we are always working at a coarse grained level where we have a chance of making decent measurements.



A cross-sectional view of a piece of wire that has been magnified to show the basic material configuration. It takes the form of a metallic cylinder which has a potential $V(t)$ applied across its end faces of area A. This then must set up a movement of charge within the volume, the constant velocity contribution being the macroscopic current $I(t)$.

For a potential difference $V(t)$ between two end faces of material we make a guess at a general polynomial of order N type behavior for the equation of motion

$$\sum_{n=1}^{N} L_n \left[\frac{dI(t)}{dt} \right]^n + R_n I(t)^n = V(t). \tag{4.4}$$

Here we haven't included any higher order derivative terms simply because equations of motion tend to have only a second time

derivative as their highest - higher derivatives leads to all sorts of mathematical and physical difficulties, notably improperly behaved solutions and ghosts. Obviously in order to map out which terms are important in Equation (4.4) we need to be able to measure currents as well as voltages. In principle, we could use the electrometer to measure both the electric charge and its rate of flow, but one might anticipate practical difficulties in this. It would therefore be good to have an independent means with which to measure just the current. To do this we need to bring back onto the scene the magnetic compass and its transformation into a galvanometer.

Hans Christian Oersted (1777–1851), a professor at Copenhagen, submitted a set of insightful observations on currents and magnetism to a leading group of physicists which were largely ignored. Working in both physics and chemistry, he was the first physicist to name and explicitly describe the "thought experiment" concept.

A key observation was made by Hans Christian Oersted in 1820 (during a lecture), that a compass needle was deflected when the current from a battery was switched on or off. More generally he found that the compass needle was deflected when it was placed in the vicinity of a current carrying wire. Building on this Ampère aligned a current carrying wire (produced via a battery) in a North-South direction and then simply placed a magnetic compass close to it having found that the needles' plane of rotation should be perpendicular to the Earth's magnetic flux density for it not to be influenced by it. The degree of the deflection is thus a measure of the local current flowing in the conductor. If the deflection settles down to a constant value the local current must have reached a steady state. We have already introduced the magnetic flux density in the last chapter. What must be taking place here is a balance between the torque produced due to the background magnetic flux density of the Earth and a local magnetic flux density produced by the current carrying wire in general - currents are a source of magnetic flux density! Thus we have a consistent picture and a local probe with which to perform measurements.

Andrè-Marie Ampère (1775–1836), Maxwell described him as the "Newton" of electricity. With a flurry of experimental and theoretical activity in 1820 Ampère linked firmly the magnetic field to steady current sources. He is also credited with the basic invention of the "Galvanometer".

Given that we can now measure current as well as voltage by two independent means, we can return to Equation (4.4) and try to establish a phenomenological relationship. This is exactly what Ohm did and established the well known phenomenological law

$$V(t) = RI(t). \tag{4.5}$$

Georg Simon Ohm (1789–1854), found the familiar Ohm's law by varying the length, area and material of different wire type configurations. This was then further developed by the use of Fourier series analysis and analogies to thermal heat flow.

For the diagram depicting the cross-sectional view of a piece of wire one can see that there is a new macroscopic component being conceptualised - the resistor. This simply limits the amount of current that can flow in a circuit that would otherwise flow without hindrance. Since the materials resistance has a dissipative effect on the current flow, the component will naturally turn this absorbed energy into heat, thus transforming the energy in some fashion.

4.2 ELECTRIC CURRENTS AND THE LAW OF AMPÈRE

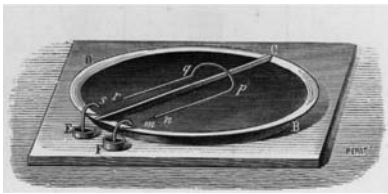
Now that we know what current is and how to measure it, we are going to try and play the same game as we did for electric charge. To that end, we will think about how steady currents interact and that

the magnetic field is the agent by which forces are affected. See [6] for more details on the history and experimental discussion.

Experiment 7 ***(Ampère, 1820–25)*** *- Take two wires carrying steady currents in the form of two separate current loops and bring them into proximity with one another. What happens to the loops of wire?*

Observation 7 *Each of them feels a force and can be either attractive or repulsive depending upon the orientation of the two loops relative to one another. It is proportional to each of the currents and varies inversely with the square of the separation.*

One should in fact be careful here about the square of the separation observation because with two different current loops there are obviously other length scales that can enter. This will be clearer when we write the force down. We now try to codify a similar equation for the currents as we did in Equation (3.1) for the electric charges. There is a striking difference that can be immediately seen between charges and currents. While charges referred to small local regions in space, steady currents require loops of a conductor to sustain themselves (i.e. we need charge to be continually moving through some region).



One of Ampère's experimental arrangements. This one gives the repulsion of collinear electric currents.

The equivalent of Coulomb's law for electric currents is *Ampères law* for current loops M and N:-

$$\mathbf{F}(M|N) = \frac{\mu_0 I^M I^N}{4\pi} \oint_{C_M} \oint_{C_N} \frac{dx(M) \wedge dx(N) \wedge [x(M) - x(N)]}{|[(x(M) - x(N)]|^3}. \qquad (4.6)$$

As with Coulomb's law, an arbitrary constant has been introduced and will depend on the system of units used. We will use Equation (4.6) to define the system of units and thus the reference point. Note also the plethora of indices - it is quite easy for these expressions to become unwieldy. The ampère will be the standard unit of current, so that one ampère is written as 1A. It is of course an

arbitrary definition, just as the kilogram or meter are simply defined to be something. This allows us to define the unit of charge as the coulomb so that $1\mathrm{A} = 1\mathrm{Cs}^{-1}$. Then the constant of proportionality is *defined* to be $\mu_0 = 4\pi \times 10^{-7}\mathrm{NA}^{-2}$, and is called the permeability of free space. Using Ampère's law, this tells us that two parallel wires each carrying 1A of current 1m apart will feel a force per unit length of $2 \times 10^{-7}\mathrm{N}$. So a base unit of current has been defined in terms of a measurable mechanical force.

(Exercise: verify this.)

4.2.1 The Magnetic Flux Density and Current Coupling

Since we are playing the same game as we did for electric fields, we want to understand how the current loop is a source of magnetic field. The observations of Oersted gave us the basic coupling between the field and the current and Ampère's law gives us the coupling between the currents. Let us take a step back and ask how we should try to run our magnetic investigation in a similar way to the the electric case. Suppose we do another experiment;

Experiment 8 *Replace one of the current loops with a bar magnet. What happens to the current loop and bar magnet?*

Observation 8 *Each of them feels a force that can be either attractive or repulsive depending upon the orientation of the two and basically behaves in a similar manner as if we had the previous current loops (at large enough separations).*

The current loops and bar magnets interact in apparently the same way, so we conclude that they are the *same* sources - currents (the movement of charge). What then is the magnetic equivalent of Equation (3.2)? Clearly we need a current loop from Ampère's law so the basic coupling must look like

$$\mathbf{F}(\mathcal{C}) = \oint_{\mathcal{C}} I(x)(dx \wedge \mathbf{B}(x)). \tag{4.7}$$

Note the difference with the electric case, that we really need think in terms of current loops. For the sustained and constant current, a loop of wire and a battery cell are required. Therefore the observable effect is necessarily on the entire loop and will effectively be at its center of mass - the effect is non-local. If the geometry and configuration permits it, we may be able to drop the integral such that

$$\delta \mathbf{F}(x) = I(x)(dx \wedge \mathbf{B}(x)), \tag{4.8}$$

where there is a current in a small element of length, but this form will not always make sense.

As it stands, the currents are naturally confined to the wires in which they have been set up. If we can identify individual charges as in Equation (3.2), that is of a point like nature, then we can rewrite Equation (4.7) in terms of its velocity $\mathbf{v}(x)$ of the charge q so that

$$\mathbf{F}(x) = q\mathbf{v}(x) \wedge \mathbf{B}(x). \tag{4.9}$$

Note here another difference between a magnetic and electric field. An electric field is a *vector* quantity while the magnetic field is an *axial-vector*. In brief this means that under a reflection the electric field points in the opposite direction, while the axial vector remains unchanged. One can see this from the wedge product - the velocity and the wedge product would yield a force that was an axial vector, were it not for the fact that the magnetic flux density is also an axial vector. How these vectors are classified is based on the group theory for how the fields transform under symmetry transformations. In particular, these definitions follow from considering how the fields transform under the rotation group and the parity group (written as $O(3)$).

(⇐ Left) Jean - Baptiste Biot (1774–1862) - a critic of Ampère and extremely conservative - more Laplacian than Laplace himself!

(Right ⇒) Felix Savart (1791–1841) - went to Paris in 1819 to see Biot and ended up putting forward the Biot-Savart law in 1820.

The experiments of Biot and Savart established the dependency of the the magnetic flux density on the current [6]. Returning to Equations (4.6) and (4.7) we see that the magnetic flux density due to a steady current loop is therefore

$$\mathbf{B}(x) = \frac{\mu_0 I}{4\pi} \oint_C \frac{dx' \wedge (x - x')}{|(x - x')|^3}. \tag{4.10}$$

When the charge is continuous so that the current becomes a current density (as is used in the continuity equation), we must form a continuum version of the above with the result

$$\mathbf{B}(x) = \frac{\mu_0}{4\pi} \int_v d^3x' \left(\frac{\mathbf{J}(x') \wedge (x - x')}{|(x - x')|^3} \right). \tag{4.11}$$

Remember that although the current I is the physical measurable, any conducting wire will have finite cross sectional area. This means that while it will be the current density that enters our equations, this will have to be integrated over the wire area to give the physical measurable I. The units of the magnetic flux density are the Tesla, T, where $1\mathrm{T} = 1\mathrm{NA}^{-1}\mathrm{m}^{-1}$.

It is also worth writing this in components

$$\mathbf{B}_a(x) = \frac{\mu_0}{4\pi} \int_v d^3x' \epsilon_{abc} \mathbf{J}^b(x') \frac{(x - x')^c}{|x - x'|^3}, \tag{4.12}$$

and drawing attention to a slightly mathematical point. In this form it is clear that the magnetic flux density is metric dependent and similarly for the equivalent electric field expression, since distances enter explicitly. However, these correspond to a particular solution of the field equations. What *is* independent of the metric is the absence of any magnetic monopoles. This is reflected in the form of **Maxwell II** where there is no source term present that is akin to the electric charge.

4.3 THE LAW OF FARADAY

Michael Faraday (1791–1867). The influence of Faraday's work on physics and electromagnetism cannot be understated; electrolysis, polarization and electromagnetic induction to name just a few. His achievements are all the more remarkable as he started his career as a bookbinder with little formal education.

Up until this point, we have investigated electric and magnetic phenomena that have mostly been decoupled. This is because historically the two originated separately and while electrostatics was seemingly well understood mathematically and experimentally, sources of current were less so; they were shrouded in the ethos of galvanism [11, 17, 6]. The observation of Oersted of changing currents producing magnetic fields of course still had a voltage (from the battery) to produce the current in the first place, so the state of decoupling is not absolute. The type of coupling manifests itself in the force laws

$$\mathbf{F}_a^{Electric} = \int_{\mathcal{V}} d^3x \rho(x) \mathbf{E}_a(x), \tag{4.13}$$

$$\mathbf{F}_a^{Magnetic} = \int_{\mathcal{V}} d^3x \epsilon_{abc} \mathbf{J}^b(x) \mathbf{B}^c(x), \tag{4.14}$$

for continuous charge distributions. The corresponding force law for a the point like test charge q is given by the *Lorentz force law*

$$\mathbf{F}(x) = q\mathbf{E}(x) + q\mathbf{v} \wedge \mathbf{B}(x). \tag{4.15}$$

By a coupling we typically mean a simple multiplication of the test body charge or current and the field at that point due to something else. The first of these is a simple scalar coupling, while the second (for the magnetic case) is a tensor coupling. Further, the magnetic field has no point source, arising as they do from the steady movement of charge. It is this very fact that suggests the following: given that they originate from the same basic source (electrical charge), it

might well be that the magnetic and electric fields are two sides of the same coin. If I change one in time, will I see a manifestation of the other? This is exactly what happens. Time for another experiment.

Experiment 9 (Faraday, 1831) *Suppose we move a bar magnet through a loop $\mathcal{C}$ of conducting wire. What happens in the loop?*

Observation 9 *We observe a time dependent current is induced in the loop. It is found that the induced voltage that produces this current is proportional to the rate of change of magnetic flux.*

Faraday's original experiment used another coil of wire instead of a bar magnet through which a time varying current passed. This had the effect of creating a time varying magnetic flux, some of which cut the coils interior in a similar fashion to the above moving bar magnet.

We can capture the above experiment in terms of the electric and magnetic fields and link them directly for the first time. The voltage $V(t)$ in our loop is just the line integral of the electric field which itself must now be time dependent

$$V(t) = \oint_{\mathcal{C}} dx^a \mathbf{E}_a(t, x). \tag{4.16}$$

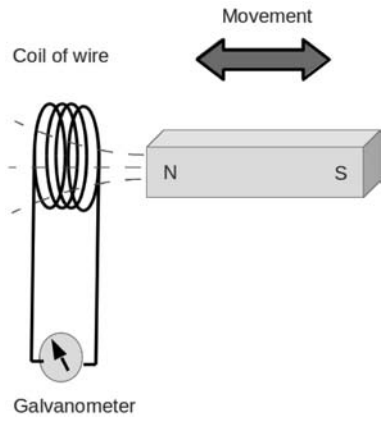


FIGURE 4.1: **Faraday's induction experiment, 1820**. When the bar magnet is moved towards or away from the coil (in its vicinity) a time varying magnetic flux is generated from standpoint of the coil. This cuts the the loop formed by the coil and the galvanometer is observed to deflect. This is the same observable effect as when a time varying voltage is setup in the coil.

This satisfies based on the above observation

$$V(t) \propto \frac{d\Phi_M}{dt}. \tag{4.17}$$

The question at this point is what the constant of proportionality should be. Remember that up until this experiment the electric and magnetic fields have been considered separate. Electric charge has been measured with respect to an arbitrary source, and current has been measured with a particular compass. The next input on this law was to fix the relative sign by essentially performing a number of experiments. This was done by Emil Lenz [6] and can be summarised as follows:-

Principle 2 (*Lenz's law*) - *the induced current takes the route such that the magnetic field it produces opposes the change in flux through the loop that initiates it.*

The best that we can do at this point therefore is to write down

$$V(t) = -\chi \frac{d\Phi_M}{dt}, \tag{4.18}$$

where the minus sign is capturing the result of Lenz' law.

Heinrich Friedrich Emil Lenz (1804–1865). Lenz arrived at his law based on the examples of induction that Faraday had considered, and some of his own experiments.

The fact that the constant of proportionality $-\chi$ appears to be undetermined requires us to impose some further constraints or symmetries. We have defined the unit of current (the Ampère) as being a flow of $1Cs^{-1}$. So we *now* take the Coulomb to be the unit of charge used in Coulomb's force law Equation (3.1). This fixes the units such that χ becomes a dimensionless number and for the permittivity of free space ϵ_0 to have units that feature the Coloumb.

(Exercise: verify the constant of proportionality is a dimensionless number and check the dimensions of the ϵ_0.)

We can go further and evaluate χ by thinking about the *relative* motion of the loop with respect to the bar magnet and the consequence of assuming Galilean invariance.

Principle 3 *Physical laws must be invariant under a Galilean transformation. In particular, forces felt by bodies must be the same in any inertial frame that is moving relative to one another with constant velocity.*

A Galilean transformation is a time dependent mapping between two coordinate systems that are moving relative to one another with some constant velocity u^a. If (t, x^a) and (t', x'^a) are coordinates in the original and moving frames of reference, then the transformation is given by the defining relations

$$x'^a = x^a + tu^a, \tag{4.19}$$

$$t' = t. \tag{4.20}$$

From this, it is simple to see how the velocity in the two frames are related, viz

$$\frac{dx'^a}{dt'} = \frac{dx^a}{dt} + u^a. \tag{4.21}$$

Now consider again Equation (4.15) in view of Galilean invariance. Suppose that a test charge is at rest and subjected to fields $\mathbf{E}(x)$ and $\mathbf{B}(x)$ in the unprimed reference frame[1]. In the primed system, since the fields are vector quantities, we expect them to also undergo a Galilean transformation to the fields $\mathbf{E}'(x')$ and $\mathbf{B}'(x')$. Galilean invariance says that the force should be the same in both reference frames so that

$$\begin{aligned}\mathbf{F}'(x') &= q\mathbf{E}'(x') - q\mathbf{u} \wedge \mathbf{B}'(x') \\ &= \mathbf{F}(x) \\ &= q\mathbf{E}(x). \\ &\Rightarrow \\ \mathbf{E}'(x') &= \mathbf{E}(x) + \mathbf{u} \wedge \mathbf{B}(x).\end{aligned} \tag{4.22}$$

To complete the transformation laws we should also calculate what the $\mathbf{B}(x)$ transforms too. This, however, is not necessary for the

task at hand of determining χ^2. Writing Equation (4.18) in integral form we have

$$\int_{\partial\mathcal{A}} dx^a \mathbf{E}_a + \chi \frac{d}{dt}\int_{\mathcal{A}} d\Sigma_a \mathbf{B}^a = 0, \tag{4.23}$$

where the loop of wire is along the contour $\partial\mathcal{A}$ and the magnetic flux density of the bar magnet cuts the area $\mathcal{A}$. The reader can probably see the road ahead. If the bar magnet is moving so that the loop is stationary, the time derivative just acts on the field. In the Galilean transformed frame however, the loop has some constant relative velocity (and so therefore does the area), such that we need to invoke Equation (2.29). On this point we should remember the Leibnitz rule for differentiating under the integral sign

$$\frac{d}{dt}\int_{u(t)}^{v(t)} dx f(t,x) = \int_{u(t)}^{v(t)} dt \frac{\partial f}{\partial t} + f(t,v)\frac{dv}{dt} - f(t,u)\frac{du}{dt}. \tag{4.24}$$

It becomes

$$\int_{\partial\mathcal{A}} dx^a \mathbf{E}'_a + \chi \int_{\mathcal{A}} d\Sigma_a (\partial_t \mathbf{B}^a + u^b \nabla_b \mathbf{B}^a) = 0. \tag{4.25}$$

This can be simplified by using Equation (2.24) and **Maxwell II** to

$$\int_{\partial\mathcal{A}} dx^a \mathbf{E}'_a + \chi \int_{\mathcal{A}} d\Sigma_a (\partial_t \mathbf{B}^a - (\nabla \wedge (u \wedge \mathbf{B}))^a) = 0. \tag{4.26}$$

Finally, by using Stoke's theorem the last term can be converted to a line integral and regrouped with the electric field

$$\int_{\partial\mathcal{A}} dx^a (\mathbf{E}'_a - \chi(u \wedge \mathbf{B}))^a) + \chi \int_{\mathcal{A}} d\Sigma_a (\partial_t \mathbf{B}^a) = 0. \tag{4.27}$$

For Galilean invariance to hold the line integral quantity in the above must be equal to the transformed electric field found in Equation (4.22) so that $\chi = 1$. Equation (4.27) becomes

$$\int_{\partial\mathcal{A}} dx^a \mathbf{E}_a + \int_{\mathcal{A}} d\Sigma_a \, \partial_t \mathbf{B}^a = 0. \tag{4.28}$$

So the constant of proportionality has been determined.

For the situation where the loop itself is not varying in time, we can write Faraday's law in local differential form, as we did for Gauss's law. By using Stoke's theorem Equation (2.34) we can convert the line integral into a surface integral

$$\int_{\mathcal{A}} d\Sigma_a (\nabla \wedge \mathbf{E})^a + \int_{\mathcal{A}} d\Sigma_a \, \partial_t \mathbf{B}^a = 0. \tag{4.29}$$

This reduces to

$$\nabla \wedge \mathbf{E} + \partial_t \mathbf{B} = 0. \quad (\textbf{Maxwell III}) \tag{4.30}$$

It is worth drawing the readers attention here to a couple of points about this equation. The first is that it is a statement about the basic electromagnetic fields with no matter content and all the action is happening in the vacuum of free space. The electric and magnetic fields are always coupled in this way regardless of what charges and currents are on the scene. The second point is a mathematical one. This equation is what is termed in differential geometry as *exact*, which means that locally we can replace the electric and magnetic fields with potentials (see for example [2] for a fuller discussion of these types of mathematical issues). We have already partly done this with the scalar potential for the electric field. In the next chapter we will look at this more fully and how it works for both fields.

Faraday's law is of crucial importance because it links electric and magnetic fields for the first time. If you have a magnetic field that varies in time, there is an electric field present at the same time. We will refer to this equation as **Maxwell III**. It should not come as too much of a surprise, however, since the original experimental means of detecting a current was from the magnetic deflection of a compass, that was initiated with an electric potential! A very readable discussion of Faraday's law can be found in [7].

One immediate use of **Maxwell III** is in the operation of an inductor and its inductance. Not surprisingly this features a coiled conducting structure, taking the form of something like a squashed helix. The key aspect to this component is that we have a conducting pathway that circulates about an axis so that a magnetic field can be produced by a steady current. What then is the phenomenological

parameter that is like the resistance for a resistor and the capacitance for a capacitor?

The basic geometry of an inductor consists of a loop type spiral structure. This allows charge to circulate around a common orthogonal axis and thereby couple to a magnetic flux density in this direction.

Considering then a simple inductor geometry where we apply a time varying potential across it, **Maxwell III** takes the form

$$V(t) = \int_{\mathcal{A}} d\Sigma_a \, \partial_t \mathbf{B}^a, \tag{4.31}$$

together with the Biot-Savart Law Equation (4.10) that the magnetic flux density is proportional to the current $I(t)$, we can write the above as

$$V(t) = \frac{\mu_0}{4\pi} \frac{dI(t)}{dt} \int_{\mathcal{A}} d\Sigma \cdot \oint_{\partial\mathcal{A}} \frac{dx' \wedge (x - x')}{|(x - x')|^3}, \tag{4.32}$$

$$= L \frac{dI(t)}{dt}, \text{ where} \tag{4.33}$$

$$L = \frac{\mu_0}{4\pi} \int_{\mathcal{A}} d\Sigma \cdot \oint_{\partial\mathcal{A}} \frac{dx' \wedge (x - x')}{|(x - x')|^3}. \tag{4.34}$$

In the above, L is the inductance for the inducting component. Typically the loop integral is evaluated by overlaying a number of circles on top of one another leading N number of coil turns, rather than evaluating the complicated helical arrangement. However, for certain applications this distinction may be necessary.

4.4 THE DISPLACEMENT CURRENT

We now return to Ampère's study of the magnetic field produced by a steady current in a long straight wire. Let us reconsider the expression Equation (4.10) which gives the magnetic field at some

point due to a steady current. Clearly the magnetic field is in a direction normal to the current flow. It is also normal to the difference vector which, together with the direction of current flow, define a plane. The magnetic field has only one component and it is normal to this plane. If we rotate the difference vector about the axis of the current, the magnetic field is unchanged and we have an honest to goodness symmetry.

Our aim here is to try and invert this relationship - we want the current in terms of the magnetic field. The simplest way to proceed is with Equation (4.12); we want to get rid of the integral over volume on the right hand side. To that end we know that if we can extract a Dirac delta function, the integral will collapse and we will have another local equation. This is further supported by recalling Gauss's law, where we have a partial derivative acting on the electric field that is equal to the charge density. So consider taking the curl of Equation (4.12);

$$\nabla \wedge \mathbf{B}(x) = \frac{\mu_0}{4\pi} \int_v d^3x' \nabla \wedge \left(\frac{\mathbf{J}(x') \wedge (x - x')}{|(x - x')|^3} \right). \tag{4.35}$$

To make progress on simplifying the right hand side we use Equation (2.44) so that we can rewrite Equation (4.35) as

$$\nabla \wedge \mathbf{B}(x) = -\frac{\mu_0}{4\pi} \nabla \wedge \nabla \wedge \int_v d^3x' \mathbf{J}(x') \left(\frac{1}{|x - x'|} \right). \tag{4.36}$$

We can simplify the above double curl using the vector identity, where we now write our expressions in components

$$(\nabla \wedge \nabla \wedge)^a_b = \nabla^a \nabla_b - \delta^a_b \nabla^2. \tag{4.37}$$

This gives in components

$$\epsilon_{abc} \nabla^b \mathbf{B}^c(x) = \frac{\mu_0}{4\pi} \int_v d^3x' \mathbf{J}_b(x') (\nabla^b \nabla_a - \delta^b_a \nabla^2) \left(\frac{1}{|x - x'|} \right). \tag{4.38}$$

The second term in the above is straightforward to evaluate, since it just gives a delta function using Equation (2.46), so

$$\epsilon_{abc}\nabla^b\mathbf{B}^c(x)=\mu_0\mathbf{J}_a(x)+\frac{\mu_0}{4\pi}\int_{\mathcal{V}}d^3x'\mathbf{J}_b(x')(\nabla^b\nabla_a)\left(\frac{1}{|x-x'|}\right). \quad (4.39)$$

The last term can be further simplified to

$$\epsilon_{abc}\nabla^b\mathbf{B}^c(x)=\mu_0\mathbf{J}_a(x)+\frac{\mu_0}{4\pi}\nabla_a\int_{\mathcal{V}}d^3x'\nabla'_b\mathbf{J}_b(x')\left(\frac{1}{|x-x'|}\right), \quad (4.40)$$

where we have used Equation (2.46) and have performed an integration by parts (we have also assumed that the boundary term vanishes by the current vanishing on $\partial\mathcal{V}$). In the case of magneto-statics where we have steady currents,

$$\nabla_b\mathbf{J}^b(x)=-\frac{\partial\rho}{\partial t}=\text{constant}=\psi. \quad (4.41)$$

We can therefore operate on the integrand and again make use of Equation (2.46)

$$\epsilon_{abc}\nabla^b\mathbf{B}^c(x)=\mu_0\mathbf{J}_a(x)-\frac{\mu_0}{4\pi}\psi\int_{\mathcal{V}}d^3x'\nabla'_a\left(\frac{1}{|x-x'|}\right), \quad (4.42)$$

which can be readily converted to a surface integral and dropped, since in the limit the boundary tends to infinity, the integrand vanishes. Thus we find that

$$\nabla\wedge\mathbf{B}(x)=\mu_0\mathbf{J}(x). \quad (4.43)$$

This can be readily converted to an expression involving the current by integrating the right hand side over an area

$$\int_{\mathcal{A}}d^2\Sigma_a(\nabla\wedge\mathbf{B}(x))^a=\mu_0\int_{\mathcal{A}}d^2\Sigma_a\mathbf{J}^a(x)=\mu_0 I, \quad (4.44)$$

while the left hand side, by Stoke's theorem, can be converted into a line integral

$$\int_{\partial\mathcal{A}}dx^a\mathbf{B}_a(x)=\mu_0 I. \quad (4.45)$$

This equation is known as Ampères' circuit law and its practical merit is that it can be used to calculate the magnetic field when the system has some manifest symmetries in it [8].

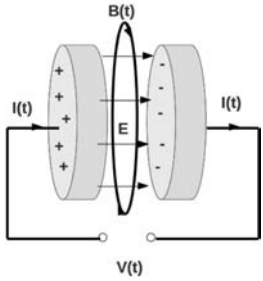
As it stands, Ampère's circuit law is valid when the current is in a steady state. This is not, however, the whole story. If we think of the parallel plate capacitor we encountered in the last chapter, it was assumed we built up the charge by some static transfer process. By pushing a current into it, the plates can similarly build up an amount of charge in well controlled manner, rather than the lumpiness in static case. Also, casting a glance at Equation (4.40) tells us that if the charge density is a general function of time

$$\epsilon_{abc}\nabla^b \mathbf{B}^c(x) = \mu_0 \mathbf{J}_a(x) - \frac{\mu_0}{4\pi}\nabla_a \int_{\mathcal{V}} d^3x' \partial_t \rho(t, x') \left(\frac{1}{|x - x'|}\right), \quad (4.46)$$

so that the last term in general does not vanish. To aid us in better understanding this equation, let us do another experiment.

Experiment 10 *Suppose we apply a time varying voltage across a parallel plate capacitor and consider the time period when it is charging up. What happens to the magnetic field?*

Observation 10 *The magnetic field surrounding the current in the wires is continuous in the separation region of the capacitor plates, as detected by a galvanometer.*



The parallel plate capacitor again. A voltage is applied across the plates so that a current flows and charge accumulates on the disks until it saturates. This is a time dependent process.

Note that no current travels between the two capacitor plates. Since charge is building up on the plates, the electric field in between will be varying as a function of time. If we take the

divergence of Equation (4.43), we find that it is inconsistent with conservation of charge

$$\begin{aligned}\nabla\cdot(\nabla\wedge\mathbf{B}) &= \mu_0\nabla_a\mathbf{J}^a \\ &= -\mu_0\,\partial_t\rho \\ &= -\mu_0\epsilon_0\,\partial_t\nabla\cdot\mathbf{E}.\end{aligned} \tag{4.47}$$

The left hand side is identically zero due to the vector identity Equation (2.28). Maxwell gave the solution to this problem by introducing the displacement current $\tilde{\mathbf{J}}$ (see for instance [6, 8]) to account for the conservation of charge. It is defined as

$$\tilde{\mathbf{J}}^a = \mu_0\epsilon_0\,\partial_t\mathbf{E}^a. \tag{4.48}$$

If we add this to the current in Equation (4.47) then we restore the continuity equation. With this then Ampère's law becomes

$$\nabla\wedge\mathbf{B}(x) - \mu_0\epsilon_0\,\partial_t\mathbf{E}(x) = \mu_0\mathbf{J}(x). \tag{4.49}$$

Note that we have modified this field equation at the local level. It is interesting to see how this arises at the global level (in terms of the basic Ampèrean integrals) as follows. Going back to our observation that the magnetic field is continuous between the metal plates, we know there is no physical movement of charge between the plates. But $\partial_t\rho(x)$ is nonzero on the plates. Therefore $\partial_t\mathbf{E}_i(x)$ is nonzero between the plates and up to a constant factor and a ∇, must be the required displacement current. We can see this is exactly what was happening in Equation (4.46) if we now use **Maxwell I**

$$\epsilon_{abc}\nabla^b\mathbf{B}^c(x) = \mu_0\mathbf{J}_a(x) - \mu_0\epsilon_0\nabla_a\int_{\mathcal{V}} d^3x'\,\partial_t\nabla'_b\mathbf{E}^b(t,x')\left(\frac{1}{|x-x'|}\right). \tag{4.50}$$

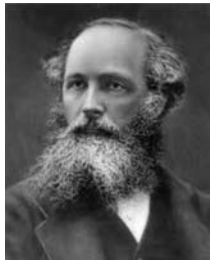
To get rid of the integral as before we pull out the time derivative and perform an integration by parts. This gives

$$\epsilon_{abc}\nabla^b \mathbf{B}^c(x) = \mu_0 \mathbf{J}_a(x) - \mu_0\epsilon_0\,\partial_t \int_{\mathcal{V}} d^3x' \mathbf{E}^b(t, x')\nabla_a \nabla_b \left(\frac{1}{|x - x'|}\right). \tag{4.51}$$

This integral is more complicated to evaluate because of the tensorial nature of the two derivatives, but it must reduce to a Dirac delta function in some fashion. We shall defer its evaluation until the next chapter when we understand the concept of gauge potentials and gauge invariance. Accepting this then for now, the final form of Ampère's original law after the addition of the displacement current which must coincide with the local version previously found is

$$\nabla \wedge \mathbf{B} - \mu_0\epsilon_0\,\partial_t \mathbf{E} = \mu_0 \mathbf{J}. \tag{4.52}$$

This is the last of Maxwell's equations referred to as **Maxwell IV**.



James Clerk Maxwell (1831–1879). The fact that Einstein had pictures of both Isaac Newton and Maxwell hanging on his wall gives a fair idea of Maxwell's contribution.

We can now return to the issue of the numerical value of the permittivity of free space that was encountered in the previous chapter. In Equation (4.52) the displacement current has the permittivity of free space ϵ_0 multiplying the electric field term. This equation also links the electric and magnetic field in the same way that Faraday's law does except that now we have a matter content entering. Because the matter content is explicit, the permittivity has to enter into it. So a numerical value for the permittivity can now be found because we know how to measure currents and electric/magnetic fields. This required in the first instance the arbitrary definition of μ_0, otherwise this would not be possible. The numerical value of the permittivity of free space can now be determined by again

performing simple types of force experiments (for example on a parallel plate capacitor configuration) or just the capacitance. Originally Weber and Kohlrausch [6] made this measurement by discharging a Leyden jar though a galvanometer; the Leyden jar's charge was measured before and after with an electrometer. The key aspect is to ensure that charge is now measured in Coulombs, since currents are given in terms of Coulombs per second. It is found to be $\epsilon_0 = 8.85\times10^{-12}\mathrm{C}^2\mathrm{N}^{-1}\mathrm{m}^{-2}$. It is now possible to define both the unit of voltage and the unit of electric field, since we have a definition of the Coulomb unit. The unit of voltage is the volt where $1\mathrm{V} = 1\mathrm{JC}^{-1}$, while the unit of electric field is the Vm^{-1}.

(⇐ Left) Wilhelm Weber (1804–1891) and Rudolf Kohlrausch (1809–1858) (right ⇒) succeeded in 1856 in measuring the ratio of electrostatic to magnetostatic forces. This number has the dimension of the inverse of velocity squared

4.5 SUMMARY

We collect here Maxwell's equation for completeness

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}, \quad \textbf{(Maxwell I)} \tag{4.53}$$

$$\nabla\cdot\mathbf{B} = 0, \quad \textbf{(Maxwell II)} \tag{4.54}$$

$$\nabla\wedge\mathbf{E} + \partial_t\mathbf{B} = 0, \quad \textbf{(Maxwell III)} \tag{4.55}$$

$$\nabla\wedge\mathbf{B} - \mu_0\epsilon_0\,\partial_t\mathbf{E} = \mu_0\mathbf{J}. \quad \textbf{(Maxwell IV)} \tag{4.56}$$

These field equations tell us how charges and currents are sources for electric and magnetic fields and how electric and magnetic fields are coupled together. Going back to how the electric

and magnetic fields were deduced, in addition to the four Maxwell equations we have the continuity equation and the Lorentz force law

$$\partial_t \rho + \nabla \cdot \mathbf{J} = 0, \tag{4.57}$$

$$\mathbf{F} = \int_v d^3x \rho \mathbf{E} + \mathbf{J} \wedge \mathbf{B}, \tag{4.58}$$

which tells us how individual charges move in the presence of electric and magnetic fields.

A few more remarks are in order here about the theory we have so far. There is the mathematical problem of how to solve these equations and the physical problem of choosing boundary conditions and the sources. In searching for the most common solutions (not the vulgar type), perhaps of practical interest and the simplest mathematically, it is often the case that the fields and the sources are decoupled in a very particular way. An inspection of the above equations would suggest that if we specify the sources then we can try and calculate the corresponding fields. In the inverse scenario they also suggest that if we specify the fields then we can calculate the trajectories of charged particles or bodies. But of course this division is one that we have made for convenience. As soon as the one quantity is specified, the dynamical variable we are trying to calculate will have a return effect (or back-reaction) on the system. This is a difficult problem to get to grips with and the interested reader can consult [8] for further discussion. Standard references for this chapter are [8] and [7].

NOTES

[1] Question: how do you *know* the test charge is at rest?

[2] In fact, as one can see by dimensional analysis, the magnetic flux density transformation will have a factor involving the relative velocity and the square of the speed of light, leading to a much smaller contribution for small speeds.

CHAPTER 5

PHYSICAL DEGREES OF FREEDOM

Perhaps the most startling of all the consequences to emerge from Maxwell's Equations is that they admit wave solutions which in the vacuum propagate with the speed of light - they are light! By decoupling the field equations, that is, separating the electric and magnetic fields from one another by performing a few mathematical operations, we find fields that satisfy wave equations. However, the picture is complicated by the fact that there is some redundancy in the system (there are two physical polarization states of light that the reader probably already knows about). It will therefore be necessary to be careful and to ensure that we do not have any spurious degrees of freedom running around in the system.

This chapter is largely mathematical. Having found the equations that govern electromagnetic fields and simple charged matter, it is obviously necessary to investigate them to begin to understand the field concept in detail. It is now that one wants to bring the full machinery of vector calculus (and more generally differential geometry) into the arena in order to manipulate the mathematical structures so far found.

5.1 THE WAVE EQUATION AND SOLUTIONS

Our starting point is the Maxwell Equations (4.53)–(4.56). Suppose that in some region of space there are no currents or charge

sources, what we call the vacuum[1]. What are the type of electric and magnetic fields that can exist here? In free space Equations (4.53)–(4.56) are simply

$$\nabla \cdot \mathbf{E} = 0, \tag{5.1}$$
$$\nabla \cdot \mathbf{B} = 0, \tag{5.2}$$
$$\nabla \wedge \mathbf{E} + \partial_t \mathbf{B} = 0, \tag{5.3}$$
$$\nabla \wedge \mathbf{B} - \mu_0 \epsilon_0 \, \partial_t \mathbf{E} = 0. \tag{5.4}$$

If we now take the curl of Equations (5.3) and (5.4) we obtain

$$\nabla \wedge \nabla \wedge \mathbf{E} + \nabla \wedge \partial_t \mathbf{B} = 0, \tag{5.5}$$
$$\nabla \wedge \nabla \wedge \mathbf{B} - \mu_0 \epsilon_0 \nabla \wedge \partial_t \mathbf{E} = 0. \tag{5.6}$$

The partial derivatives with respect to time and space commute, i.e. $[\partial_t, \nabla] = 0$, so that

$$\nabla \wedge \nabla \wedge \mathbf{E} + \partial_t \nabla \wedge \mathbf{B} = 0, \tag{5.7}$$
$$\nabla \wedge \nabla \wedge \mathbf{B} - \mu_0 \epsilon_0 \, \partial_t \nabla \wedge \mathbf{E} = 0. \tag{5.8}$$

By using Equations (5.3) and (5.4) a further time we can achieve the decoupling of the electric and magnetic fields, so that

$$\nabla \wedge \nabla \wedge \mathbf{E}(x) + \mu_0 \epsilon_0 \, \partial_t^2 \mathbf{E} = 0, \tag{5.9}$$
$$\nabla \wedge \nabla \wedge \mathbf{B}(x) + \mu_0 \epsilon_0 \, \partial_t^2 \mathbf{B}_i(x) = 0. \tag{5.10}$$

In this way the fields have decoupled into two separate equations of motion. To simplify the two wedge products we have to use the vector identity Equation (2.28), where

$$\nabla \wedge \nabla \wedge \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}, \tag{5.11}$$
$$\nabla \wedge \nabla \wedge \mathbf{B} = \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B}. \tag{5.12}$$

With this, the source free Maxwell Equations finally reduce to

$$-\nabla^2\mathbf{E}(t,x)+\mu_0\epsilon_0\,\partial_t^2\mathbf{E}(t,x)=0, \tag{5.13}$$

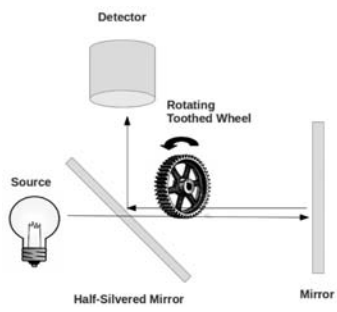
$$=\Box\mathbf{E}(t,x)=0, \tag{5.14}$$

$$-\nabla^2\mathbf{B}(t,x)+\mu_0\epsilon_0\,\partial_t^2\mathbf{B}(t,x)=0. \tag{5.15}$$

$$=\Box\mathbf{B}(t,x)=0, \tag{5.16}$$

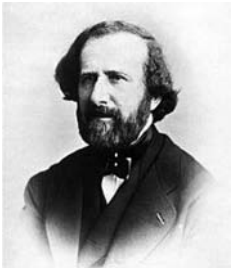
$$\Box:=-\nabla^2+\mu_0\epsilon_0\,\partial_t^2, \tag{5.17}$$

where we have introduced the second order differential operator $\Box$, which is known as the D'Alembertian operator. Its significance is that it is a wave operator and what we have just written down are the three dimensional wave equations. The reader should already be familiar with the wave equation in one spatial dimension, where the solutions in general look like $f(x\pm vt)$, with v the speed of propagation along the x direction. So electric and magnetic fields that are wave like in nature are admitted as solutions and describe propagation in free space (the vacuum) at a speed $c=1/\sqrt{\mu_0\epsilon_0}$. In the previous chapter we saw that μ_0 was a defined quantity and that ϵ_0 was measured relative to this. Historically it was found that by taking the values stated there then numerically the value of c was found to be close to the speed of light $c\approx 3.0\times10^8\,\mathrm{ms}^{-1}$.



The Fizeau-Foucault experiment was a way of measuring the speed of light by mechanical means. It was devised such that the time parameter is linked to a rotational speed. Light is shone on a mirror at a known distance from the source. In between is a toothed wheel that rotates with a known angular velocity. The beam of light is thus converted into a series of pulses. Making observations close to the source and knowing the number of teeth on the wheel allows us measure the transit time of the beam of light. Fizeau arrived at a value of $c\approx 3.1\times10^8\,ms^{-1}$.

The great insight of Maxwell was to realize that light is exactly this type of electromagnetic disturbance and that it is one part of infinite spectrum of different types of an electromagnetic radiation [17, 6]. Measuring the speed of light of course had been performed in an independent fashion previously. The key parameter to measure is a transit time for light to propagate from one point to another, given that we know the separation accurately.



(Left ⇐) Armand Hippolyte Louis Fizeau (1819–1896), in addition to the measurement of the speed of light also predicted red shifting of electromagnetic waves and the discovery of the Doppler Effect. His name is inscribed on the Eiffel Tower.

(Right ⇒) Jean Bernard Léon Foucault (1819–1868) also invented a rather special pendulum that shows that the Earth is rotating, as his name suggests.

5.2 BASIC WAVE SOLUTIONS - SCALAR FIELD THEORY

What do the fields look like that satisfy the wave equation? Looking at Equations (5.13) and (5.15) one sees that we are solving the same basic partial differential equation, but that the two vector fields are not independent. We will need to ensure that the solutions we obtain are consistent. To this end, we will start off just looking at the electric field solutions and then check later on to see what the consistency conditions demand.

To obtain a solution we must consider the coordinate system. So far we have not specified any at all, but to obtain an analytical expression it is obvious that a choice has to be made. In general, the choice of a coordinate system should be based on any geometrical symmetries in the problem, if indeed any exist at all. For example, it would be unwise to choose oblate spheroidal coordinates to solve a problem involving a single long cylinder. We will develop solutions in the three simplest and perhaps most useful coordinate systems: (i) Cartesian coordinates, (ii) spherical coordinates, and (iii) cylindrical coordinates. Our aim here is to arrive at the solutions without complete mathematical rigor, but rather to have an appreciation of the type of solutions and what they mean with respect to one another.

Since we are choosing to consider wave like solutions the time dependency will clearly be oscillatory. Regardless of the spatial coordinates used, we can write down the functional dependence on time. It is simply

$$\mathbf{E}_a(t, x) = \mathbf{E}_a(x)e^{i\omega t} + c.c. \tag{5.18}$$

for a single frequency mode or

$$\mathbf{E}_a(t, x) = \int_{-\infty}^{\infty} d\omega e^{i\omega t}\mathbf{E}_a(\omega, x) + c.c. \tag{5.19}$$

when there is continuous spectrum of frequencies (or modes). Note here the observation about this part of the solution, that we have assumed a Fourier transform of the solution; we are therefore turning the operator in time (that is the second partial derivative in time) into an algebraic expression in frequency space. This is quite the common theme in searching for solutions to the wave equation. For a continuous space (such as the time coordinate) the strategy is to look for a Fourier transform solution, while for a periodic (compact) space the strategy is to look for a *Fourier series* type solution.

Substitution of Equation (5.18) into the Equation (5.13) gives

$$\nabla^2\mathbf{E}_a(x) + \frac{\omega^2}{c^2}\mathbf{E}_a = 0. \tag{5.20}$$

This equation is known as the scalar *Helmholtz equation*, and for the time being we postpone the complication that the electric

field is a constrained vector field and treat it as a scalar quantity with an index.

Hermann Ludwig Ferdinand von Helmholtz (1821–1894). Both a physicist and a physician, he worked on a number of areas ranging from color vision, visual perception, thermodynamics, electrodynamics, plus a good smattering of philosophy.

To solve this equation we will use the method of separation of variables to obtain analytic expressions of the coordinates used. In fact we have done this implicitly in Equation (5.18); write the solution as $\mathbf{E}_a(t,x) = \mathbf{E}_a(x)T(t)$ and then rearrange into separate spatial and time pieces, so that

$$\nabla^2\mathbf{E}_a(x) - \frac{1}{c^2}\mathbf{E}_a(x)\left[\frac{1}{T(t)}\frac{\partial^2 T(t)}{\partial t^2}\right] = 0. \tag{5.21}$$

The term in square brackets must be a constant, $-\omega^2$, so that the spatial part is independent of time. The solutions are exponentials in time. If this constant is negative, we have oscillating exponentials while if it positive the solutions exponentially grow or decay. In general it could be a complex number, $\omega^2 \in \mathbb{C}$. The choice made is dictated by what type of boundary conditions one is putting on the problem and also what type of source they are originating from. Of immediate interest is of some type of oscillating source so that we choose the oscillating solutions from

$$\left[\frac{1}{T(t)}\frac{d^2 T(t)}{dt^2}\right] = -\omega^2. \tag{5.22}$$

This leads directly to the Helmholtz equation Equation (5.20).

The simplest set of coordinates to work in are Cartesian coordinates wherein the Helmholtz equation is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\mathbf{E}_a + \frac{\omega^2}{c^2}\mathbf{E}_a = 0. \tag{5.23}$$

This has the well known plane wave set of solutions which can be obtained by separation of variables if desired. Each coordinate spans the entire real line $\mathbb{R}$, so it will be a Fourier transform solution that we seek. Note also that there is no difference between an upstairs and a downstairs index in these coordinates because the metric is simply the Kronecker Delta tensor. One finds

$$\mathbf{E}_a(x) = E_a e^{\pm ik\cdot x}, \tag{5.24}$$

$$k^a k_a = k_a k_a = k_1 k_1 + k_2 k_2 + k_2 k_2 = k^2, \tag{5.25}$$

$$k^2 := \omega^2 / c^2. \tag{5.26}$$

The constants E_a are just the complex amplitudes of the electric field, which are the arbitrary integration constants. That they are complex should not worry the reader, as it is only the full electric field that needs to be real. With this in mind the full electric field solution reads

$$\mathbf{E}_a(t, x) = E_a e^{i(k\cdot x - \omega t)} + E_a^* e^{-i(k\cdot x - \omega t)}. \tag{5.27}$$

(Exercise: verify this is a solution)

The plane wave solutions are also *eigenfunctions* of the different differential operators that make up the Helmholtz operator, with the respective wavevectors being the *eigenvalues*. This is of course intimately linked with the separation of variables method which we can illustrate thus. Recall that for an operator (or a matrix) $\hat{M}$ with eigenvalues λ_n and eigenfunctions (or vectors) ψ_n one has

$$\hat{M}\psi_n = \lambda_n \psi_n. \tag{5.28}$$

Apply this to Equation (5.23); if we take $\hat{M} = \partial / \partial x^2$ then the eigenfunction is $\psi_n = e^{ik_1 x}$ with the eigenvalue $\lambda_n = -k_1 k_1$. Equation (5.23) then becomes

$$\left(-(k_1)^2 + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{E}_a(y, z) + k^2 \mathbf{E}_a(y, z) = 0, \tag{5.29}$$

where the electric field has the eigenfunction multiplying the remaining part of the solution (separation of variables). In this manner we

then build up the remaining parts of the solution. In other coordinate systems the same strategy can be adopted.

One last thing that we can see here is that in general there may be many wave vectors contributing to the electric field in which case we should sum over them with an appropriate weight for each mode

$$\mathbf{E}_a(t,x) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}[E_a(\mathbf{k})e^{i(k\cdot x-\omega t)} + E_a^*(\mathbf{k})e^{i(-k\cdot x+\omega t)}]. \qquad (5.30)$$

This is exactly a Fourier transform in three dimensional space, as to be expected because we are solving a wave equation. In addition to this, all of the spatial dimensions are infinite (non-compact). The Helmholtz equation then serves to put a constraint on the wave vectors in the form of Equation (5.25).

5.2.1 Solutions in Spherical Coordinates

Turning now to spherical coordinates (r,θ,ϕ), one can try to build up a general solution using spherical waves, the equivalent of plane waves in the previous discussion. There is a subtlety now that since in spherical coordinates the metric is non-trivial, we would need to ensure that we are using the correct form of the Laplacian operator to act of the electric vector field. The Laplacian operator given in Equation (2.66) is for a scalar field. We will assume for now that a scalar field will be sufficient but it is worth checking under what conditions this is valid or is a well defined approximation.

The Helmholtz operator in spherical coordinates acting on a scalar field $\varepsilon(x)$ is given by

$$\left(\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right)\varepsilon(x) + k^2\varepsilon(x) = 0. \qquad (5.31)$$

Assuming again the separation of variables method so that the solution in spherical coordinates take the form $\varepsilon(x) = R(kr)\Theta(\theta)\Phi(\phi)$, one can start the reduction off again (we have included a factor of k in the radial solution so that its argument is dimensionless). The simplest term in this Laplacian is the ϕ partial derivative and so this is the place to start with

$$\frac{1}{\Phi(\phi)}\frac{d^2}{d\phi^2}\Phi(\phi) = -m^2 = \text{constant}. \tag{5.32}$$

In spherical coordinates, any solution must be invariant under a full rotation of the ϕ coordinate so that $\phi \to \phi + 2\pi$ leaves any solution invariant. This forces the two conditions that $m^2 > 0$, so that the solutions are not exponentially growing or decaying, and that $m = 0, \pm 1, \pm 2 \cdots$, so that the solutions are periodic. It is a straightforward differential equation solve, with the two solutions

$$\Phi(\phi) = e^{\pm im\phi}. \tag{5.33}$$

Using this to simplify Equation (5.31) results in

$$\frac{1}{R(kr)r^2}\frac{d}{dr}\left(r^2\frac{dR(kr)}{dr}\right) + \frac{1}{r^2\Theta(\theta)\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta(\theta)}{d\theta}\right) - \frac{m^2}{r^2\sin^2\theta} + k^2 = 0. \tag{5.34}$$

Since each term involve factors of r^2, one needs to ensure that there is a resulting differential equation in r has no other factors involving the other two coordinates in it. So the two terms that have the θ factors in should be set equal to some constant

$$\frac{1}{R(kr)r^2}\frac{d}{dr}\left(r^2\frac{dR(kr)}{dr}\right) + \left(-\frac{l(l+1)}{r^2} + k^2\right) = 0, \tag{5.35}$$

where the constant has been chosen to be the value $-l(l+1)$ for later convenience. Indeed, given the earlier discussion about eigenvalues, one may well be tempted to suspect that the l given here is something to do with this - and it certainly is. It is therefore helpful to rewrite Equations (5.34) and (5.35) in the following form

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR(kr)}{dr}\right) + \left(-\frac{l(l+1)}{r^2} + k^2\right)R(kr) = 0, \tag{5.36}$$

$$\left[\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta(\theta)}{d\theta}\right)-\frac{m^2}{\sin^2\theta}\right]\Theta(\theta)+l(l+1)\Theta(\theta)=0. \quad (5.37)$$

(Exercise: verify this is true.)

The problem has now been reduced to finding solutions for two separate second order differential equations. These two equations occur so often in theoretical physics that they have been named. Equation (5.36) is known as the *spherical Bessel equation*, while Equation (5.37) is known as the *associated Legendre equation*. A standard piece of mathematical analysis to solve these equations is to make a power series substitution into the differential equations and then deduce recurrence relations amongst the coefficients. Any standard textbook on mathematical methods describe this procedure, so we shall not pursue this avenue further (see for example [15]). The radial part of the solution then consists of the *spherical* Bessel functions $j_l(kr)$ (which are regular at the origin) and the *spherical Neumann* functions $n_l(kr)$ (which diverge at the origin). The angular part of the solution consists of the *associated Legendre polynomials* $P_{l,m}(\cos\theta)$. The general solution then looks like

$$\varepsilon(r,\theta,\phi)=\sum_{l,m}\left(V^{(l,m)}j_l(kr)+W^{(l,m)}n_l(kr)\right)P_{l,m}(\cos\theta)e^{im\phi}, \quad (5.38)$$

where the summation runs over all l, $(l=0,1,\cdots)$ and the m summation is bounded by $-l\le m\le l$. In addition to the constrained summation, it is necessary to impose a reality condition on the solutions, so that the physical field as a function of time (rather than frequency) is real valued. This is because both the solutions are complex valued. Note that summations over the discrete angular momentum numbers have replaced the continuous integrations over plane wave vectors. The fact that these are eigenfunctions with eigenvalues is perhaps a little less clear here (after all we haven't explained it) but it is there and much more interesting in this case.

To finish off this subsection, it is useful to write down recursion relations for generating the respective solutions from simpler well known elementary functions (rather than just a power series expansion).

$$j_l(x) = (-x)^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \left(\frac{\sin x}{x}\right), \tag{5.39}$$

$$n_l(x) = -(-x)^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \left(\frac{\cos x}{x}\right), \tag{5.40}$$

$$Y_{l,m}(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_{l,m}(\cos\theta) e^{im\phi}, \tag{5.41}$$

$$P_{l,m}(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m}\left(\frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l\right). \tag{5.42}$$

The $Y_{l,m}(\theta,\phi)$ introduced above are called the *spherical harmonics* and play a fundamental role in spherical coordinates. They are the eigenfunctions on the two dimensional sphere. Note also the different nature of the dispersion relation here compared with the plane wave solutions. In spherical coordinates, via the spherical Bessel functions, the constant k defines the length scale over which the radial solutions vary. The angular form, however, is unconstrained by it and only coupled the by the l eigenvalue.

5.2.2 Solutions in Cylindrical Coordinates

The other standard set of coordinates that every undergraduate has to play with are the cylindrical coordinates (r,θ,z). The method of solution proceeds analogously with spherical coordinates. Note that this is the halfway house between Cartesian and spherical; one dimension is compact (θ), one is semi-infinite (r) and the other is infinite (z). The first thing to do is write down the Helmholtz equation in cylindrical coordinates

$$\left(\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}\right)\varepsilon(x) + k^2\varepsilon(x) = 0. \tag{5.43}$$

Both the θ and z variables do not present much difficulty now in solving as we have already come across their types of solution earlier. They are both complex exponentials with either a continuous or discrete separation constant (think eigenvalue!) that labels the solution. They look like

$$\Theta(\theta) = e^{im\theta},\ m \in \mathbb{Z}, \tag{5.44}$$

$$Z(z) = e^{ipz},\ p \in \mathbb{R}. \tag{5.45}$$

With these solutions the Helmholtz equation reduces to a simple second order differential equation given by

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \left(k^2 - p^2 - \frac{m^2}{r^2}\right)R = 0. \tag{5.46}$$

This is just the standard *Bessel* equation with solutions given by the Bessel functions $J_m(x)$ and $N_m(x)$. They are related simply to previous spherical Bessel and Neumann functions defined by

$$J_L(x) = \sqrt{\frac{2x}{\pi}}\, j_{L-1/2}(x), \tag{5.47}$$

$$N_L(x) = \sqrt{\frac{2x}{\pi}}\, n_{L-1/2}(x). \tag{5.48}$$

The full solution then will require a combination of a sum over the discrete index and an integration over the continuous one. Thus,

$$\varepsilon(r,\theta,z) = \sum_m \int_{-\infty}^{\infty} dp e^{im\phi} e^{ipz}\left[V_a^{(m)} J_m(r\sqrt{k^2 - p^2}) + W_a^{(m)} N_m(r\sqrt{k^2 - p^2})\right]. \tag{5.49}$$

As with the other solutions found for the Helmholtz equation, it is necessary impose a reality constraint so that the physical electric fields are real valued.

5.2.3 Comments

A few remarks are in order here about the scalar field solutions found in the three coordinate systems. The first comment is that it is only in spherical coordinates that the solutions can be square normalizable (see Equation (2.48)). This is intimately related to the fact that *two* of the coordinates are compact (the θ and ϕ) and the *type* of source that is ultimately producing the electric field. The implication

here is that the source has a finite spatial extent in all directions of space. One can therefore put a closed surface around it. This can only be done in spherical polar coordinates where this closed surface would be a sphere of suitable size radius. A second remark following on from this is that the energy associated with these waves will then be finite. While we have not defined so far the energy contained in a fluctuating field, it will turn out the energy density looks like the product of two fields together. In the other two coordinate systems this is not the case. A necessary condition is to maintain the geometry of the problem. In Cartesian coordinates we would be considering an infinite plane (or surface) source giving rise to the plane wave type solutions. Similarly in cylindrical coordinates it would be an infinite line (or curve) source that would produce cylindrical wave type solution. A final remark is that for distances much larger than the corresponding wavelength (given by c/ω) one can often work, at least locally, in Cartesian coordinates as an approximation to one of the others. This is because locally a curved surface can be approximated by a flat plane. One must remember then not to calculate outside its region of applicability. Useful discussions on related matters can be found in both [8] and [7].

5.3 DEGREES OF FREEDOM - POLARIZATION

We have just seen that the electric and magnetic fields satisfy separately wave equations, and one might conclude hastily that there are six degrees of freedom (the three electric E_a and three magnetic B_a undetermined amplitudes). That this is not the case will be seen by being a bit more explicit with the type of wave solutions we encounter and a closer look at how they are satisfying Maxwell's equations. It should be said that this is a difficult topic - we are trying to find consistent solutions to the full vector field equations. This is necessary however to understand what polarization really is.

5.3.1 Polarization in Cartesian Coordinates

Let us consider again the plane wave solutions in Cartesian coordinates for a a particular wavevector. We now also include a similar

solution for the magnetic field and ask the question how they are connected. The general vector solutions will look like

$$\mathbf{E}_a(t,x) = \Re(E_a \exp(i(k_a x_a \pm \omega t)), \tag{5.50}$$

$$\mathbf{B}_a(t,x) = \Re(B_a \exp(i(k_a x_a \pm \omega t)), \tag{5.51}$$

where the wavevector k_a gives the direction of propagation (and wavelength) and both amplitudes are in general constant complex numbers. We must remember now that to derive the two wave equations we had to perform certain vector calculus operations on the first order Maxwell's equations to disentangle the electric and magnetic fields. So we should really substitute the solutions into the original equations. Doing so yields

$$k_a E_a = 0, \tag{5.52}$$

$$k_a B_a = 0, \tag{5.53}$$

$$(k \wedge E)_a \pm \omega B_a = 0, \tag{5.54}$$

$$(k \wedge B)_a \mp \frac{1}{c^2}\omega E_a = 0. \tag{5.55}$$

The first two equations say that one component of the electric and magnetic field are redundant. For a given k_a, one of the physical field vector components can be found in terms of the other two. So we lose a degree of freedom for the electric and magnetic fields. Similarly, for the next two equations, it can be seen that we lose another degree of freedom from each. We have therefore lost four degrees of freedom, bringing down the physical degrees of freedom from six to two.

This now takes us neatly on to the *polarization* of an electromagnetic plane wave. Equation (5.54) says that the magnetic field is normal to the wavevector and the electric field, while Equation (5.55) says that the electric field is normal to the wavevector and the magnetic field. So the three form a mutually orthogonal set of vectors. Without any loss of generality, imagine the wave is propagating in the $+z$ direction. The electric field will then have in general the components (E_x, E_y). The question we now ask is what is the allowed solution space? Let us return to our wave solution Equation (5.50) and write it out in components

$$\mathbf{E}_x(t,x) = \Re[E_1 \exp(i(k_z z - \omega t)], \tag{5.56}$$
$$\mathbf{E}_y(t,x) = \Re[E_2 e^{i\epsilon} \exp(i(k_z z - \omega t)]. \tag{5.57}$$

We have allowed here for a phase difference ϵ between the two components, with $E_1, E_2 \in \mathbb{R}$ and positive. The strategy now is to eliminate the time dependence so that find the required solution space. Using the double angle formula in the above we have

$$\mathbf{E}_x(t,x)/E_1 = \cos(k_z z - \omega t), \tag{5.58}$$
$$\mathbf{E}_y(t,x)/E_2 = \cos(k_z z - \omega t)\cos\epsilon - \sin(k_z z - \omega t)\sin\epsilon. \tag{5.59}$$

The time dependence can now be eliminated from Equation (5.59)

$$\frac{\mathbf{E}_y(t,x)}{E_2} = \frac{\mathbf{E}_x(t,x)}{E_1}\cos\epsilon - \left[1 - \left(\frac{\mathbf{E}_x(t,x)}{E_1}\right)^2\right]^{1/2}\sin\epsilon. \tag{5.60}$$

This leads to the equation of an ellipse

$$\left(\frac{\mathbf{E}_x(t,x)}{E_1}\right)^2 + \left(\frac{\mathbf{E}_y(t,x)}{E_2}\right)^2 - 2\left(\frac{\mathbf{E}_x(t,x)\mathbf{E}_y(t,x)}{E_1 E_2}\right)\cos\epsilon = \sin^2\epsilon, \tag{5.61}$$

where the ellipse is rotated by some angle with respect to the x, y coordinates. It easy to work out what this angle is - just perform a rotation on the x, y electric field components.

(Exercise: calculate the angle α the ellipse is rotated by in the above.)

Following on from this general state of polarization, it is necessary to catalogue the different states of polarization that all descend from the elliptical case, since these are often used in practice.

- **Linear polarization** - the two components are either completely in phase or out of phase with one another by $\epsilon = n\pi$, with $n \in \mathbb{Z}$. The electric field oscillates in a straight tilted line that is determined by the two amplitudes.
- **Circular polarization** - here the amplitudes are equal, $E_1 = E_2$, but now the phase relationship is given by $\epsilon = \pm 2n\pi$, with $n \in \mathbb{Z}$. The positive sign is defined to be left-circularly

polarized, while the negative sign is defined to be right-circularly polarized.

- **Elliptical polarization** - This is when the amplitudes and phase take on general values as already described. The angle of tilt α of the ellipse with respect to the coordinate system of E_x, E_y is given by $\tan\ \alpha = 2E_1E_2\cos\epsilon/(E_1^2 - E_2^2)$.

What about polarization in the other coordinate systems? We know that either the waves are outgoing or incoming with respect to their coordinate origin. They are made up of eigenfunctions that respect the symmetries of the coordinate system about their coordinate origin. The math of it is more complicated (though quite interesting) so only a sketch of the solutions are given. The interested reader can pursue this further in [8] and [13].

5.3.2 Polarization in Spherical Coordinates - TE and TM Modes

In spherical coordinates the full vector Helmholtz equation needs to be solved. These need to be disentangled into two polarization states known as the transverse electric (TE) and transverse magnetic (TM) modes. Heavy use is made of the angular momentum operator $\mathbf{L} = x \wedge \nabla$. We give here just a brief statement of the form they take. The general solution takes the form

$$\mathbf{B} = \sum_{l,m} a^E(l,m)\mathbf{L}\Psi_{l,m}(r,\theta,\phi) - \frac{i}{k}a^M(l,m)\nabla \wedge \mathbf{L}\Pi_{l,m}(r,\theta,\phi), \quad (5.62)$$

$$\mathbf{E} = \sum_{l,m} a^M(l,m)\mathbf{L}\Pi_{l,m}(r,\theta,\phi) + \frac{i}{k}a^E(l,m)\nabla \wedge \mathbf{L}\Psi_{l,m}(r,\theta,\phi), \quad (5.63)$$

where

$$\Psi_{l,m}(r,\theta,\phi) := [c_l j_l(kr) + d_l n_l(kr)]\frac{1}{\sqrt{l(l+1)}}Y_{l,m}(\theta,\phi), \quad (5.64)$$

$$\Pi_{l,m}(r,\theta,\phi) := [c_l' j_l(kr) + d_l' n_l(kr)]\frac{1}{\sqrt{l(l+1)}}Y_{l,m}(\theta,\phi), \quad (5.65)$$

$$c_l, d_l, c_l', d_l' \in \mathbb{C}. \quad (5.66)$$

What this shows is that the solutions consist of solutions to the scalar Helmholtz equation which are then acted on by suitable operators to generate the full vector field solutions. From the boundary conditions and the source conditions all of the integration constants can be determined in as a set of multipole fields. It is helpful to just write down the equivalent polarization conditions for the spherical case. The polarization state given by just the TE modes satisfy

$$x \cdot \mathbf{E} = 0, \tag{5.67}$$

$$x \cdot \mathbf{B} = \text{satisfies scalar Helmholtz equation}, \tag{5.68}$$

while the polarization state of just the TM modes satisfy

$$x \cdot \mathbf{B} = 0, \tag{5.69}$$

$$x \cdot \mathbf{E} = \text{satisfies scalar Helmholtz equation}. \tag{5.70}$$

Multipole fields, while being somewhat more fiddly in a mathematical sense, are extremely useful. For example, if you want to consider light scattering from a dielectric or metallic sphere, these are the solutions typically required to evaluate physical observables. Comparing this with the more standard Cartesian coordinates polarization where we had an ellipse of possibilities, we see something simpler here. The reader is referred to [8] for a fuller discussion. In a similar fashion polarization in cylindrical coordinates can be constructed, and is quite relevant for structures such as waveguides [8].

5.4 GAUGE INVARIANCE AND POTENTIALS

Earlier we encountered a scalar potential that was used to describe the static electric field. By taking the gradient of this we could recover the electric field. A natural question to ask is what is the corresponding potential for the magnetic field? Recall that **Maxwell II** is a statement that there are no magnetic monopoles

$$\nabla_a \mathbf{B}^a = 0. \tag{5.71}$$

Just as we found earlier that a scalar potential solves the curl free condition on the electric field ($\nabla \wedge \mathbf{E} = 0$), our question now is what is the corresponding potential for the above. It is a mathematical result (see for instance [2]) that a vector potential $\mathbf{A}_a$ solves the divergence free condition on the vector field $\mathbf{B}^a$ such that

$$\mathbf{B}^a = \epsilon^{abc} \nabla_b \mathbf{A}_c. \tag{5.72}$$

(Exercise: verify this solves Maxwell II.)

If we now put this into **Maxwell III** we find a new statement for the electric field, that is

$$\begin{aligned} \epsilon_{abc} \nabla^b \mathbf{E}^c + \partial_t \epsilon_{abc} \nabla^b \mathbf{A}^c = 0, \\ \epsilon_{abc} \nabla^b (\mathbf{E}^c + \partial_t \mathbf{A}^c) = 0. \end{aligned} \tag{5.73}$$

Equation (3.7) has been modified to include the time dependent effects of **Maxwell III**. However, it is once again in the form of the curl of some new vector equalling zero, so we know that it can be written as the gradient of some scalar function viz,

$$\begin{aligned} \mathbf{E}_a + \partial_t \mathbf{A}_a = -\nabla_a \phi \Rightarrow \\ \mathbf{E}_a = -\nabla_a \phi - \partial_t \mathbf{A}_a. \end{aligned} \tag{5.74}$$

Just as the magnetic field has been written in terms of a potential, so has the electric field (in terms of the same vector potential), but also an additional scalar potential. One can not help but notice the lack of symmetry here between the electric and magnetic fields. We also seem to have created a problem for ourselves in terms of the number of degrees of freedom we now have in the system. It was found previously that there were only two physical polarization states, yet now we seem to have a total of four (scalar plus vector) degrees of freedom. Why the mismatch?

What we have just touched upon is the concept of *gauge invariance*, first expounded by Hermann Weyl in connection with coordinate transformations in general relativity. The astute reader will have noticed that there has always been something to do with symmetry lurking in the background. The basic point that we will see is that

the potentials we have just discovered are not unique - we can add arbitrary functions (in fact two) to them and the field equations will be unchanged because the physical fields themselves are invariant under this transformation. This type of deep symmetry is the basis for much of modern particle physics and the Standard Model. For an historical perspective on classical gauge theory see [9].

Hermann Klaus Hugo Weyl (1885–1955), mathematician, theoretical physicist and philosopher, he worked on a large number of areas including number theory, group theory, and the geometry associated with combining electromagnetism with general relativity.

To this end, we consider substituting Equations (5.72) and (5.74) into the field Equations (4.53)–(4.56). One finds that

$$-\nabla^2\phi(x)+\partial_t(\nabla_a\mathbf{A}^a(x))=\frac{\rho(x)}{\epsilon_0}, \tag{5.75}$$

$$\epsilon^{abc}\nabla_a\nabla_b\mathbf{A}_c\equiv 0, \tag{5.76}$$

$$\nabla\wedge(-\nabla\phi+\partial_t\mathbf{A})+\partial_t\nabla\wedge\mathbf{A}=0, \tag{5.77}$$

$$(\nabla\wedge\nabla\wedge\mathbf{A})+\mu_0\epsilon_0\,\partial_t(\nabla\phi-\partial_t\mathbf{A})=\mu_0\mathbf{J}. \tag{5.78}$$

Immediately it is clear that these equations will simplify. Equation (5.76) is identically zero, while in Equation (5.77) a simple cancellation of the vector potential occurs together with the vector identity that the curl of a divergence is zero. This leaves the two remaining equations as

$$-\nabla^2\phi(x)+\partial_t\nabla_b\mathbf{A}^b(x)=\frac{\rho(x)}{\epsilon_0}, \tag{5.79}$$

$$+\nabla^2\mathbf{A}_a-\mu_0\epsilon_0\,\partial_t^2\mathbf{A}_a+\nabla_a(\mu_0\epsilon_0\,\partial_t\phi-\nabla_b\mathbf{A}^b)=\mu_0\mathbf{J}_a. \tag{5.80}$$

Here we have used Equation (2.24) together with a small amount of rearrangement. Now it is time to use the gauge invariance. A short

calculation reveals that if we use instead the gauge potentials $\phi', \mathbf{A}'_a$ that are related to the original ones in Equation (5.79) and Equation (5.80) by

$$\phi'(t,x) = \phi(t,x) + \partial_t \Omega(t,x), \tag{5.81}$$

$$\mathbf{A}'_a(t,x) = \mathbf{A}_a(t,x) + \nabla_a \Omega(t,x), \tag{5.82}$$

where $\Omega(t,x)$ is an arbitrary function, then the electric and magnetic fields remain unchanged. Therefore, the field equations for the potentials Equation (5.79) and Equation (5.80) will also remain invariant with this choice of potentials. In technical language they are said to be gauge invariant. This means that we can *choose* the arbitrary function so that ϕ and $\mathbf{A}_a$ satisfy some condition, thereby *choosing* a gauge. Looking at Equation (5.80) one can see that it is nearly a wave equation, we just need to get rid of the last term. We can do this by choosing the gauge

$$\mu_0 \epsilon_0 \partial_t \phi - \nabla \cdot \mathbf{A} = 0. \tag{5.83}$$

The above gauge condition is in fact known as the *Lorentz gauge*. By making this choice one of the four potentials becomes a function of the other three. For example, perform an integration over time such that

$$\phi(t,x) = \frac{1}{\mu_0 \epsilon_0} \int_{-\infty}^{t} dT \nabla \cdot \mathbf{A}(T,x). \tag{5.84}$$

Here the scalar potential becomes a functional of the vector potential.

With this choice of constraint on the gauge potential the two field equations reduce to

$$-\nabla^2 \phi(t,x) + \mu_0 \epsilon_0 \partial_t^2 \phi(t,x) = \frac{\rho(t,x)}{\epsilon_0}, \tag{5.85}$$

$$+\nabla^2 \mathbf{A}(t,x) - \mu_0 \epsilon_0 \partial_t^2 \mathbf{A}(t,x) = \mu_0 \mathbf{J}(t,x). \tag{5.86}$$

It should be apparent that we have not quite finished as we still appear to have three degrees of freedom in the system. We need to

impose a further constraint to be sure that we have only two physical degrees of freedom, which we know to be true from the two states of polarization previously found. Where does this come from?

Let us re-examine the Lorentz gauge condition Equation (5.83) that we imposed. Clearly we chose it because the field equations for the scalar and vector potentials would disentangle from one another rather neatly, but apart from that it looked quite arbitrary. This is the point about gauge transformations that they are arbitrary. Once we choose a gauge we need to ensure two important properties; one is that we have removed all spurious degrees of freedom from the system; and a second is that all the physical observables and properties remain gauge invariant. Let us pin down the first point. What happens to the Lorentz gauge if we if we do another gauge transformation? Obviously the electric and magnetic fields are the same, whereas the Lorentz gauge condition becomes

$$\mu_0\epsilon_0\,\partial_t(\phi+\partial_t\chi)-\nabla_b(\mathbf{A}^b+\nabla^b\chi)=0, \tag{5.87}$$

where $\chi(t,x)$ is another arbitrary function that specifies the gauge transformation. The Lorentz gauge is preserved in the above provided the new gauge transformation parameter satisfies

$$\mu_0\epsilon_0\,\partial_t\partial_t\chi(t,x)-\nabla^2\chi(t,x))=0, \tag{5.88}$$

that is, it must satisfy a scalar wave equation. So this arbitrary scalar can then be chosen to set one of the components of the vector potential to zero. For example in Cartesian coordinates if we have a wave propagating in the z-direction, we can choose $A_z(t,x,y,z)=0$ by a suitable choice of $\chi(t,x,y,z)$ satisfying the wave equation. At this point we arrive at two components of the vector potential representing the two physical degrees of freedom of the electromagnetic field. Note another feature of this process. By performing a gauge transformation on the gauge condition itself, we were able to eliminate another spurious degree of freedom. One might ask if it is possible to do the same thing again and finish with only one degree of freedom. The answer is an emphatic "no", because any subsequent gauge transformation will have a gauge function that *also* satisfies the scalar wave equation and (by the superposition of solution for

linear equation) this second gauge function can be absorbed into the first transformation.

Another commonly encountered gauge choice is the *Coulomb gauge* which is given by

$$\nabla_a \mathbf{A}^a(t,x) = 0. \tag{5.89}$$

With this gauge condition the field equations become

$$-\nabla^2\phi = \frac{\rho}{\epsilon_0}, \tag{5.90}$$

$$+\nabla^2 \mathbf{A}_a - \mu_0\epsilon_0\,\partial_t^2 \mathbf{A}_a - \mu_0\epsilon_0 \nabla_a\,\partial_t\phi = \mu_0 \mathbf{J}_a. \tag{5.91}$$

For the connections between the Lorentz and Coulomb gauges, see [10].

Given now that we have written the field equations in terms of potentials, what are the immediate quantities of interest? Certainly, it would be useful to have the gauge potential as a functional of the current, as we can then simply derive the magnetic field. In addition to this, the Lorentz force law Equation (4.15) when expressed in terms of gauge potentials is also an observable of interest. As a final point, we should return to displacement current encountered in the global setting and try to evaluate Equation (4.51).

Firstly, in the case of magnetostatics, where the displacement current does not contribute, we can rewrite Ampère's law in terms of the vector potential as

$$\begin{aligned}
\nabla \wedge \nabla \wedge \mathbf{A}(x) &= \mu_0 \mathbf{J}(x), \\
\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}(x), \\
-\nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}(x), \\
\mathbf{A}(x) &= \frac{\mu_0}{4\pi} \int_{\mathcal{V}} d^3x' \frac{\mathbf{J}(x')}{|x - x'|},
\end{aligned} \tag{5.92}$$

where we have taken the vector potential to be in the Coulomb gauge and have just inverted the Laplacian operator in exactly the

same way as in the Poisson equation. Moving on to the Lorentz force equation one sees that

$$\begin{aligned} m\frac{d^2x^a(t)}{dt^2} &= -q(\nabla^a\phi + \partial_t \mathbf{A}^a) + q\epsilon^{abc}\frac{dx_b(t)}{dt}\epsilon_{cmn}\nabla^m\mathbf{A}^n \\ &= -q\nabla^a\phi - q\,\partial_t\mathbf{A}^a - q(\delta^b_m\delta^a_n - \delta^b_n\delta^a_m)\frac{dx_b(t)}{dt}\nabla^m\mathbf{A}^n \\ &= -q\nabla^a\phi - q\,\partial_t\mathbf{A}^a - q\frac{dx^b(t)}{dt}\nabla_b\mathbf{A}^a + q\frac{dx^b(t)}{dt}\nabla^a\mathbf{A}_b \\ &= -q\frac{d}{dt}\mathbf{A}^a - q\nabla^a\left(\phi + \frac{dx^b(t)}{dt}\mathbf{A}_b\right), \end{aligned} \tag{5.93}$$

which rearranges to

$$\frac{d}{dt}\left(m\frac{dx^a(t)}{dt} + q\mathbf{A}^a\right) = -q\nabla^a\left(\phi + \frac{dx^b(t)}{dt}\mathbf{A}_b\right). \tag{5.94}$$

The interesting observation here is that the charged particle receives a contribution of $q\mathbf{A}^a$ to its momentum (in classical mechanics, notably when one works in terms of hamiltonians, one has to be careful to distinguish between velocity and conjugate momenta). It looks like that the momenta $m\frac{dx^a(t)}{dt} + q\mathbf{A}^a$ is a gauge dependent quantity.

(Exercise: Check that the Equation (5.94) written in terms of the gauge potential remains gauge invariant.)

Finally we reconsider the displacement current using gauge potentials. If we consider evaluating Equation (4.51) by writing the electric field in terms of the gauge potentials one finds

$$\begin{aligned} \epsilon_{abc}\nabla^b\mathbf{B}^c(x) = \mu_0\mathbf{J}_a(x) - \mu_0\epsilon_0\,\partial_t\int_{\mathcal{V}} d^3x'(-\nabla^b\phi(t,x') \\ -\partial_t\mathbf{A}(t,x'))\nabla_a\nabla_b\left(\frac{1}{|x-x'|}\right). \end{aligned} \tag{5.95}$$

If we choose the Coulomb gauge then the ϕ term integrates simply, while the $\mathbf{A}$ term vanishes after an integration by parts. Therefore

$$\epsilon_{abc}\nabla^b \mathbf{B}^c(x) = \mu_0 \mathbf{J}_a(x) - \mu_0 \epsilon_0 \, \partial_t \nabla_a \phi(t, x). \tag{5.96}$$

The final step is to realize that the conservation of charge results from taking the divergence of the above equation; therefore, we can simply add the Coloumb gauge condition to the divergence of the above. As a field equation it must be gauge invariant so we make the replacement

$$\epsilon_{abc}\nabla^b \mathbf{B}^c(x) = \mu_0 \mathbf{J}_a(x) + \mu_0 \epsilon_0 \, \partial_t (-\nabla_a \phi(t, x) - \partial_t \mathbf{A}(t, x)), \tag{5.97}$$

which *is* now a gauge invariant equation.

The discussion of physical observables remaining gauge invariant is simple provided we are working with the fields and not the gauge potentials. So far we have built up the field theory implicitly assuming some symmetries in the system. Consider the coordinates where the field is sat at say $\mathbf{E}_a(t, x)$. Here we have taken time t to be just a parameter, but the spatial coordinates x have been implicitly assume to transform under the rotation group in three dimensions. If we have chosen a gauge for example where $A_z(t, x) = 0,$ then one can see that this symmetry group has now been *broken*. One needs to ensure then that any measurable will not suffer from this issue (indeed this goes into making sure that any quantum theory of electrodynamics remains internally consistent).

5.5 SUMMARY

We have found in this chapter some of the basic physical properties that emerge by looking at the mathematical details of the the field equations. The equations admit wave like solutions in free space, which we have focused on as the physical solutions of the field equations. Good discussions can be found in [8] and [15]. It is complicated by the fact that the physical propagating modes (two degrees of freedom) are fewer in number than the number of components that constitute the electric and magnetic fields. These are *hidden* by the original fields and the corresponding potentials from which they are derived. This is the central feature of gauge invariance, and

is the key mathematical structure around which electromagnetism and more general field theories revolve. A fuller discussion of the historical roots of gauge invariance and how they feature in classical electrodynamics are given in [9] and [10].

NOTE

[1] The vacuum is a subtle concept. In quantum theories of electromagnetic fields, for example, forces can still arise due to fluctuations in the vacuum. This is a rich but somewhat advanced topic.

CHAPTER 6

PHYSICAL OBSERVABLES

The physical observables we have encountered thus far have been forces that we could measure mechanically. Electric forces were turned into voltage measurements by using an electrometer. Having understood the connection between currents and magnetic fields, galvanometers have been constructed that provide a measure for a current. Each time a mechanical force or torque has been traded for a voltage or current. These are then what we would regard as macroscopic physical observables. They can be used to quantify measurements of different electromagnetic structures when a test body is introduced as a local probe of the system.

Another feature that is apparent is that the field theory has been investigated at the linear level. For example, when we considered the force on a charge or current it has always been proportional to the field in a linear way. This is not surprising because of how the fields are defined in the first place. But we can now go a bit further and ask, since this is a dynamical system, what are the observable properties like energy or momentum of the fields themselves. To answer this type of question, it is necessary to go to the next order and consider *bilinear* field terms. This is where much of the dynamics and observables reside. A detector will only register a small piece of information about the field in this context.

Before getting on to this, however, we first make a short consolidation of the physical observables that are associated not with the fields themselves, but with the movement of charge. A field is

applied and the charges move. This movement under different conditions produces distinct effects. We shall therefore consider the students favourite electrical circuit as a means to understand some of the intricacies involved.

6.1 THE RLC CIRCUIT

Every student of physics knows and loves (or hates) this circuit. It is the Resistor-Inductor-Capacitor circuit and it is quite ubiquitous in its dynamical form. Thinking of it mechanically, it describes an oscillating body that is subjected to a frictional force. The basic observable in this circuit is the current and an applied voltage is the source that drives the circuit. Let us start off analyzing this circuit as a tube of metal (the wire) which lies in the z-direction. The ends have a cross sectional area $\mathcal{A}$ across which we apply a time dependent voltage $V(t)$. The current across the plane $z = 0$ is given by

$$I(t, z = 0) = \int_{\mathcal{A}} d^2 \sum_a \mathbf{J}^a(x, y, z = 0). \tag{6.1}$$

We will assume that at any other cross sectional plane to the tube the current is the same. In principle $\mathbf{J}^a$ is a function of position along the tube. The essential constraint is obviously conservation of charge. The three components we can immediately insert are the parallel plate capacitor (given in Section 3.4), the resistor (given in Section 4.1), and the inductor (given is Section 4.3). In terms of the tube geometry, a cut is made in the tube and a small section removed. Then the faces are made oversize so that the diameter of the face is much larger than the diameter of the tube - this is the insertion of a parallel plate disk capacitor. We make a similar cut for the resistor and insert a different piece of material, that neither conducts too much, nor too little. For the final component (the inductor) we take a section of tube, squash it and then coil it up so that we have a helical structure with an axis and a radius about it. We now take the tube to be small in comparison with the components, that is $\mathcal{A} \to 0$. To finish off the circuit we assume that its shape is rectangular with four corners as shown in Figure (6.1). The question now to ask is what is the dynamical equation that governs this system?

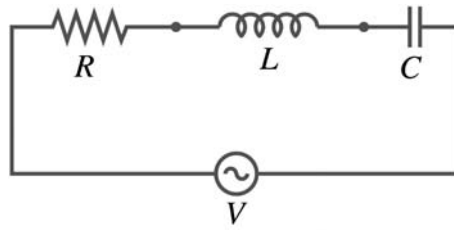


FIGURE 6.1: **The RLC circuit** - the basic form of this equation appears in large number of different areas. A pendulum with friction, a material with absorption and a decaying particle state all share this common type of structure.

The two constraints that determine the equation of motion are the conservation of current, together with that the potential differences across each component must sum to the applied voltage $V(t)$. It is given by the well known equation

$$\frac{1}{C}Q + R\frac{dQ}{dt} + L\frac{d^2Q}{dt^2} = V(t). \tag{6.2}$$

Since we know $V(t)$, we can differentiate it to obtain an expression involving the current, viz

$$L\frac{d^2I(t)}{dt^2} + R\frac{dI(t)}{dt} + \frac{1}{C}I(t) = \frac{dV(t)}{dt}. \tag{6.3}$$

So what have we learned? By applying what is often called a *test* function $V(t)$, we can drive this circuit so that it responds. The response is the movement of charge and the physical observable is the current as a function of time (as measured by a galvanometer). From a simple knowledge of differential equations, we recognize two solutions to the above - the transient and the particular integral. If the voltage $V(t)$ is switched off then the solution reverts to the transient solution where the constants of integration are determined from its behavior at the switching off time. The transient solution looks like exponential functions which have a real and imaginary component. They are

$$I(t) = c_1 e^{\alpha t} + c_2 e^{\beta t}, \quad \text{for} \quad V(t) = 0, \tag{6.4}$$

$$\alpha = R/2L + \sqrt{(R/2L)^2 - 1/(LC)}, \tag{6.5}$$

$$\beta = R/2L - \sqrt{(R/2L)^2 - 1/(LC)}, \tag{6.6}$$

where $c_1, c_2 \in \mathbb{C}$ are constants of integration. There is always an exponentially decaying component to these solutions but these may well be oscillating as well. When the test function is on, the solution is the particular integral plus the transient solution that is determined from earlier behavior. Note also that this is perhaps the simplest example of the use of Green functions that one may encounter, and serves as a basis for more complicated problems. To solve using a Green function we simply write the current as

$$I(t) = \int dt' G(t-t') \frac{dV(t')}{dt'}. \tag{6.7}$$

We are making here an explicit causal link where a source of oscillation at an earlier time propagates and has its resulting effect at a later time. With this Equation (6.3) is transformed to

$$L\frac{d^2G(t-t')}{dt^2} + R\frac{dG(t-t')}{dt} + \frac{1}{C}G(t-t') = \delta(t-t'). \tag{6.8}$$

It may not be obvious, but it is expedient to perform a Fourier transform on this equation so that we can trade time derivatives for multiples of frequency ω. One finds

$$-\omega^2 LG(\omega) + Ri\omega G(\omega) + \frac{1}{C}G(\omega) = 1. \tag{6.9}$$

The time dependent Green function then is obtained by transforming back into the time coordinate

$$G(t-t') = \int \frac{d\omega}{2\pi} e^{i\omega(t-t')} G(\omega) \tag{6.10}$$

$$= \int \frac{d\omega}{2\pi} e^{i\omega(t-t')} (-\omega^2 L + Ri\omega + 1/C)^{-1}. \tag{6.11}$$

With this Green function we have the system response (the current) as a functional of the input test function (the voltage) viz

$$I(t) = \int \frac{d\omega}{2\pi} e^{i\omega(t-t')} (-\omega^2 L + Ri\omega + 1/C)^{-1} \frac{dV(t')}{dt'}. \tag{6.12}$$

In addition to the movement of charge through each component given by Equation (6.3), there is also their energy/power behavior as a function of time. For each component there is a potential difference (and a corresponding force) across it that does work on the charge to move it between the two points p_1 and p_2. This movement of charge is given exactly by the current. Therefore, the instantaneous rate of doing work W is for each of the three components

$$\begin{aligned} W &= \frac{d}{dt}\int_{p_1}^{p_2} dx^a \mathbf{F}_a \\ &= \int_{p_1}^{p_2} dx^a \frac{dq}{dt}\nabla_a \phi \\ &= \frac{dq}{dt}[\phi(p_2)-\phi(p_1)] \\ &= I(t)V(t). \end{aligned} \tag{6.13}$$

With this result, we can consider each component in turn. The resistor is particularly simple - just use Ohm's law for the voltage in terms of the current to reach

$$W_R = RI^2(t). \tag{6.14}$$

This is the energy lost in the resistor which is converted into heating in the component which would then be given off as thermal radiation (another topic in itself). Similarly, for the inductor using Equation (4.33) we have

$$W_L = LI(t)\frac{dI(t)}{dt} = \frac{d}{dt}\left(\frac{L}{2}I^2(t)\right). \tag{6.15}$$

This expression represents how the inductor releases and stores its energy. For the capacitor, we have to work at the level of charge rather than current such that

$$W_C = \frac{1}{C}Q(t)\frac{dQ(t)}{dt} = \frac{d}{dt}\left(\frac{1}{2C}Q^2(t)\right). \tag{6.16}$$

Again, this gives how the capacitor stores and releases energy into the circuit. For the most part this is standard electrical theory

and the reader can consult any standard text (for example [7]) for more discussion.

Of course, in order to derive these macroscopic observables, information has been thrown away to reach these results. We have traded the local current density for a current and coarse grained over the geometry of the circuit. Its virtue lies in its applicability and utility. All of electronics is built up from similar considerations. This is perhaps one the simplest examples of physical observables - it sits on the matter side and is linear. We now turn to similar considerations for the fields themselves and their bilinear nature.

6.2 FIELD ENERGY, MOMENTUM, AND ANGULAR MOMENTUM

The fields that we originally encountered were due to the presence of charged bodies or steady currents. Observed static forces on similar charges or currents were our reason to introduce the fields and necessarily in a linear way. We also have found that electromagnetic fields can propagate in the vacuum. The natural question to then ask is if such propagating fields can carry energy with them and can produce physical forces on other bodies. To answer this, we will have to work at the next level of complication and consider quantities that are bilinear in the fields.

A simple illustration of this can be made by considering the Lorentz force law. Suppose we ask what the force on a charge q is due to a passing electromagnetic wave that is linearly polarized. For an oscillating electric field, the Lorentz force law looks like

$$\mathbf{F} = q\mathbf{E}_0 \cos(\omega t). \tag{6.17}$$

Taking as a reference point an optical frequency in the above with an order of magnitude value of $\omega/2\pi \sim 10^{15}\,\mathrm{Hz}$, one sees that the charge will be forced to oscillate extremely rapidly. If we can agree that this fluctuation (that is the displacement caused by the force) is too fast to be measured, then it becomes necessary to employ an

averaging procedure. In such cases we take the time average over one period T of oscillation with the result that

$$\langle \mathbf{E}_i(t) \rangle_T = 0, \tag{6.18}$$

$$\langle \mathbf{B}_i(t) \rangle_T = 0. \tag{6.19}$$

So physical quantities that are linear in the fields when they are oscillating average to zero in this scenario. It is then necessary to work with higher order polynomial field quantities to obtain non-vanishing physical measurables.

6.2.1 The Stress Tensor and Conservation of Momentum

Firstly, let us formulate the conservation of momentum by starting with the Lorentz force law applied to a continuum charge density enclosed in a volume $\mathcal{V}$.

$$\mathbf{F}_a(x)\,|_{\mathcal{V}} = \int_{\mathcal{V}} d^3x \rho(x) \mathbf{E}_a(x) + \epsilon_{abc} \mathbf{J}^b(x) \mathbf{B}^c(x). \tag{6.20}$$

Conservation laws in physics are ultimately the consequence of symmetries. From classical mechanics, one of the key results is that the equations of motion do not change under a symmetry transformation and this leads to a constant of motion (this can be expressed most elegantly in the Lagrangian formalism where the concept of a Noether current is introduced). In this way, invariance under time translations leads to conservation of energy; invariance under spatial translations leads to conservation of momentum; and invariance under rotations leads to conservation of angular momentum. We will begin with a consideration of the field equations. Symmetries also give rise to a conserved "charge", where the idea of charge is used in a very general sense. Historically it arose from electric charge where the underlying gauge symmetry of the field equations leads to this. It is also worth remarking on the tensor nature of the encountered charges. If one recalls Equation (4.3) for the conservation of electric charge we see a scalar and a vector quantity present. This is because we are looking at the transport of a scalar quantity, the

electric charge. If we now wish to think about the conservation of momentum, it should not be a surprise to see the introduction of a different tensorial object.

Our strategy then is to eliminate the matter fields by the use Maxwell's Equations (4.53) and (4.56). Then we will have an expression that is bilinear in the fields. Since we have a bounding area $\partial\mathcal{V}$ of the volume, we are thus looking to find a stress on this area. It is therefore necessary to convert the volume integral into a surface integral.

$$\mathbf{F}_a|_{\mathcal{V}} = \int_{\mathcal{V}} d^3x \epsilon_0 (\nabla_c \mathbf{E}^c)\mathbf{E}_a + \frac{1}{\mu_0}(\epsilon^{bcd}\nabla_b \mathbf{B}_c - \frac{1}{c^2}\partial_t \mathbf{E}^d)\epsilon_{dea}\mathbf{B}^e. \quad (6.21)$$

Noting that

$$\partial_t(\mathbf{E}\wedge\mathbf{B}) = \partial_t \mathbf{E}\wedge\mathbf{B} + \mathbf{E}\wedge\partial_t \mathbf{B}, \quad (6.22)$$

and using this in the above allows the force to be written as

$$\mathbf{F}_a|_{\mathcal{V}} = \int_{\mathcal{V}} d^3x \epsilon_0 (\nabla_b \mathbf{E}^b)\mathbf{E}_a + \frac{1}{\mu_0}\epsilon^{bcm}\epsilon_{mna}(\nabla_b \mathbf{B}_c)\mathbf{B}^n$$
$$-\frac{1}{c^2}\partial_t(\epsilon_{abc}\mathbf{E}^b\mathbf{B}^c) + \epsilon_0(\epsilon_{abc}\mathbf{E}^b\partial_t \mathbf{B}^c). \quad (6.23)$$

Before making further simplifications let us understand what we have so far. The left hand side is clearly a mechanical quantity. It is the sum of all the forces on the charge distribution and so therefore can be written using Newton's second law as

$$\mathbf{F}_a|_{\mathcal{V}} = \frac{d}{dt}\mathbf{P}_a^{mech.}. \quad (6.24)$$

On the right hand side we can also see a total derivative with respect to time. The fact that it is cross product between the electric and magnetic field means that for wave like solutions, it is pointing in the direction of propagation. It also has the units of momentum which suggests that we identify it with field momentum, viz

$$\mathbf{P}_a^{field} = \epsilon_0 \int_{\mathcal{V}} d^3x \epsilon_{abc} \mathbf{E}^b \mathbf{B}^c. \tag{6.25}$$

The remaining terms are bilinear in the fields and also have a derivative. If we collect the two types of momentum together then a small rearrangement using Equation (4.55) gives

$$\frac{d}{dt}\left(\mathbf{P}_a^{mech.} + \mathbf{P}_a^{field}\right) = \int_{\mathcal{V}} d^3x [\epsilon_0 \epsilon^{bcm} \epsilon_{mna} (\nabla_b \mathbf{E}_c) \mathbf{E}^n + \frac{1}{\mu_0} \epsilon^{bcm} \epsilon_{mna} (\nabla_b \mathbf{B}_c) \mathbf{B}^n] + \epsilon_0 [(\nabla_b \mathbf{E}^b) \mathbf{E}_a]. \tag{6.26}$$

In this form we see that the time derivative of the total momentum must equal the force acting upon it. So the right hand side must be the force of the physical fields. We can also make the right hand side symmetrical in the fields by adding into it **Maxwell II** (Equation (4.54)) so that

$$\frac{d}{dt}\left(\mathbf{P}_a^{mech.} + \mathbf{P}_a^{field}\right) = \int_{\mathcal{V}} d^3x \epsilon_0 [(\nabla_b \mathbf{E}^b) \mathbf{E}_a + \epsilon^{bcm} \epsilon_{mna} \mathbf{E}^n \nabla_b \mathbf{E}_c] + \frac{1}{\mu_0} [(\nabla_b \mathbf{B}^b) \mathbf{B}_a + \epsilon^{bcm} \epsilon_{mna} \mathbf{B}^n \nabla_b \mathbf{B}_c]. \tag{6.27}$$

A final step in the simplification is to use the tensor identity

$$\epsilon^{abc} \epsilon_{cmn} = \delta_m^a \delta_n^b - \delta_n^a \delta_m^b. \tag{6.28}$$

Using this result one finds

$$\frac{d}{dt}\left(\mathbf{P}_a^{mech.} + \mathbf{P}_a^{field}\right) = \int_{\mathcal{V}} d^3x \nabla_b [\epsilon_0 (\mathbf{E}_a \mathbf{E}^b - \frac{1}{2} \delta_a^b \mathbf{E}_c \mathbf{E}^c) + \frac{1}{\mu_0} (\mathbf{B}_a \mathbf{B}^b - \frac{1}{2} \delta_a^b \mathbf{B}_c \mathbf{B}^c)]. \tag{6.29}$$

(Exercise: verify this.)

From the right hand side we see that we are taking the divergence of a symmetric two-index tensor. It is in fact the stress tensor and from the above we define it to be

$$T_{ab} := \epsilon_0(\mathbf{E}_a\mathbf{E}_b - \frac{1}{2}g_{ab}\mathbf{E}_c\mathbf{E}^c) + \frac{1}{\mu_0}(\mathbf{B}_a\mathbf{B}_b - \frac{1}{2}g_{ab}\mathbf{B}_c\mathbf{B}^c). \tag{6.30}$$

To check what it means we just rewrite the above conservation of momentum law and make use of Gauss's theorem

$$\begin{aligned}\frac{d}{dt}\left(\mathbf{P}_a^{mech.} + \mathbf{P}_a^{field}\right) &= \int_{\mathcal{V}} d^3x \nabla_b T_a^b \\ &= \int_{\partial\mathcal{V}} d^2\textstyle\sum_b T_a^b.\end{aligned} \tag{6.31}$$

If we integrate the stress tensor over an area we find the total force on that area. By measuring the force on this boundary surface, we are then indeed measuring the stress tensor T_{ab} and so it represents a physical observable.

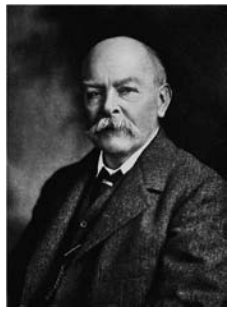
Note also here that it is really the total momentum that is the meaningful quantity and that in general it will not be possible to make a sharp split between the two. Consider again Equation (6.29); the right hand describes the flow of field momentum across the bounding surface of the volume. The left had side has both a changing field and mechanical part. How the momentum is distributed between these two inside the volume is not specified. This is not so much of an issue here since it is the stress tensor evaluated on the bounding surface that can be measured (the total change in momentum must be the applied force).

One such measurement of the electromagnetic stress is the radiation pressure on a surface. This is the experimental observable that paved the way to throwing the corpuscular theory of light into some confusion (see for instance [12]).

6.2.2 The Poynting Vector and Energy Conservation

Carrying on from the previous matter and fields system, where forces are acting on the matter, we can also ask if they do work on them. If this is true then because this is a time varying system, we can calculate the rate of doing work on them. It is only the electric field that can do work on the assembled charges; they must have a velocity and from the Lorentz force law Equation (4.58), the magnetic field doesn't contribute since it is orthogonal. If the fields are doing work on the charges then energy must be lost from the field

and an accompanying flow of energy must result. This is the basic statement of Poynting's theorem.



John Henry Poynting (1852–1914) worked with Maxwell in the late 1870s and formulated his conservation law in 1874.

The rate of doing work on matter by the physical fields is given by

$$W = \int_{\mathcal{V}} d^3x \mathbf{E}_a \mathbf{J}^a, \tag{6.32}$$

and not surprisingly the algebraic manipulations are similar to before. We use **Maxwell IV** (Equation (4.56)) to express the current density in terms of the electric and magnetic fields so that

$$\begin{aligned} W &= \int_{\mathcal{V}} d^3x \mathbf{E}_a \frac{1}{\mu_0}(\epsilon^{abc}\nabla_b \mathbf{B}_c - \frac{1}{c^2}\partial_t \mathbf{E}^a) \\ &= \int_{\mathcal{V}} d^3x \frac{1}{\mu_0}(\nabla_b(\epsilon^{abc}\mathbf{E}_a\mathbf{B}_c) - (\epsilon^{abc}\nabla_b \mathbf{E}_a)\mathbf{B}_c - \frac{1}{c^2}\mathbf{E}_a\partial_t \mathbf{E}^a). \end{aligned} \tag{6.33}$$

For the last step, we just have to use **Maxwell III** (Equation (4.55)) for the final simplification

$$\begin{aligned} W &= \int_{\mathcal{V}} d^3x \frac{1}{\mu_0}(-\nabla_b(\epsilon^{abc}\mathbf{E}_a\mathbf{B}_c) + (-\partial_t \mathbf{B}^a)\mathbf{B}_a - \frac{1}{c^2}\mathbf{E}_a\partial_t \mathbf{E}_a) \\ &= -\int_{\mathcal{V}} d^3x \left[\frac{1}{\mu_0}\nabla_b(\epsilon^{abc}\mathbf{E}_a\mathbf{B}_c) + \frac{1}{c^2}\mathbf{E}_a\partial_t \mathbf{E}^a + \frac{1}{\mu_0}\mathbf{B}_a\partial_t \mathbf{B}^a\right] \\ &= -\int_{\mathcal{V}} d^3x \nabla_a \mathbf{S}^a + \partial_t U \qquad (6.34) \\ &= -\int_{\partial\mathcal{V}} d^2\textstyle\sum_a \mathbf{S}^a + \int_{\mathcal{V}} d^3x\, \partial_t U, \qquad (6.35) \end{aligned}$$

where

$$\mathbf{S}^a := \frac{1}{\mu_0}\epsilon^{abc}\mathbf{E}_b\mathbf{B}_c, \tag{6.36}$$

$$U := \frac{1}{2}\left(\epsilon_0\mathbf{E}_a\mathbf{E}^a + \frac{1}{\mu_0}\mathbf{B}_a\mathbf{B}^a\right). \tag{6.37}$$

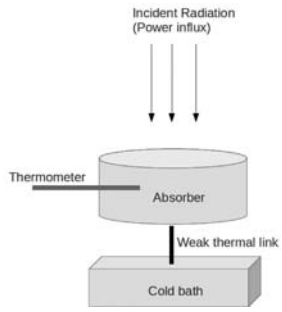
We see then that U represents the energy density of the field while $\mathbf{S}^a$ is the energy flux and is called the *Poynting vector*. This is the statement of conservation of energy; if fields do work on charges then the fields energy density must change together with a flux out of/into the enclosing volume. In the case where no work is done on the internal charges we see conservation of energy just for the field. The Poynting vector when integrated over some closed surface represents the total power of the field that has flowed into or out of that bounded volume.

Samuel Pierpont Langley (1834–1906) first invented the bolometer in 1878 in connection with his work in astronomy.

The type of measurement required for the power is a bit different from the previous conservation of momentum observables. In that case we were measuring a force, which could be achieved by a mechanical type balancing procedure. For a power measurement we need to employ a different approach. Essentially we need the incoming radiation to be deposited completely in the matter itself so that the fields do work on the charges. One such a device is a bolometer (first invented by Samuel Langley). It measures incoming electromagnetic power by turning the deposited power in the detector into heat. This in turn can be converted phenomenologically into a change of of temperature, which in turn can be converted to an incident power.

Note that as with measuring the stress tensor, the observable has had to be traded (or balanced) with a similar quantity (in this

case a thermal energy). There is a second more subtle point connected with the process of depositing energy and Equation (6.35). How can one be sure that all of the incident power flux is converted into mechanical power and thereby heat? The left hand side of Equation (6.35) has both a mechanical and a field part, both of which are scalar quantities. In principle the incident flux that enters the volume can be distributed between the two contributions completely arbitrarily. The other point is that we are having to invoke some outside phenomenology in order to make the measurement. Therefore it is necessary to use thermodynamics and coarse grained material properties to try to capture the field information. The reader might like to think over these issues at a later date.



A bolometer being used to measure the magnitude of the Poynting vector. If all of the incident flux is deposited into the absorbing material and this turned directly into heat, then a measurement of the deposited power can be made. A measured temperature difference can be then be turned into a power, as the weak thermal link to the cold bath is designed to ensure a steady state configuration.

6.2.3 Conservation of Angular Momentum

We have seen two conservation laws at work in basic electromagnetic field theory, the conservation of energy and the conservation of momentum. As any man on the street will tell you, it is only natural to consider the conservation of angular momentum as well. Just as in the case of the conservation of momentum, where a force couples to a displacement of charge, we need to consider how a *torque* couples to an angular change. If we consider Equation (6.21) again but drop the integral sign we have at the point x that

$$\delta \mathbf{F}_a(x) = \epsilon_0(\nabla_c \mathbf{E}^c)\mathbf{E}_a + \frac{1}{\mu_0}(\epsilon^{bcd}\nabla_b \mathbf{B}_c - \frac{1}{c^2}\partial_t \mathbf{E}^d)\epsilon_{dea}\mathbf{B}^e. \qquad (6.38)$$

Let us suppose for the time being we are working in standard Cartesian coordinates. Then x is also the position vector from the coordinate system origin - this means that we can use it to form the torque $\delta\mathbf{\Gamma}_a(x)$ at the point x by the standard wedge product

$$\begin{aligned}\delta\mathbf{\Gamma}_a(x) &= \epsilon_{abc}x^b\delta\mathbf{F}^c(x)\\ &= \epsilon_0(\nabla_m\mathbf{E}^m)\epsilon_{abc}x^b\mathbf{E}^c + \frac{1}{\mu_0}\epsilon_{abc}x^b(\epsilon^{mnd}\nabla_m\mathbf{B}_n\\ &\quad -\frac{1}{c^2}\partial_t\mathbf{E}^d)\epsilon^c_{de}\mathbf{B}^e. \end{aligned} \tag{6.39}$$

It is possible to save oneself some work here by recalling the calculation of the stress tensor Equations (6.29) and (6.30). Basically we can form the simple product of x^a with the stress tensor and other components because of two reasons. Firstly the partial time derivative doesn't act on the x^a so it can be moved around freely. Secondly, the spatial derivative operators will also act on x^a in a simple way. Summing up all the contributions to the torque by reinstating the integral, we find the equivalent of Equation (6.27) is

$$\begin{aligned}\mathbf{\Gamma}_a^{Mech.} &= -\epsilon_0\partial_t\int_{\mathcal{V}}d^3x\epsilon_{abc}x^b\mathbf{E}_d\epsilon^{cde}\mathbf{B}_e\\ &\quad \epsilon_0\int_{\mathcal{V}}d^3x\epsilon_{abc}x^b[(\nabla_m\mathbf{E}^m)\mathbf{E}^c + \epsilon^{pqm}\epsilon^c_{mn}\mathbf{E}^n\nabla_p\mathbf{E}_q]\\ &\quad + \frac{1}{\mu_0}\int_{\mathcal{V}}d^3x\epsilon_{abc}x^b[(\nabla_m\mathbf{B}^m)\mathbf{B}^c + \epsilon^{pqm}\epsilon^c_{mn}\mathbf{B}^n\nabla_p\mathbf{B}_q]. \end{aligned} \tag{6.40}$$

What we see here is exactly the angular equivalent of the conservation of momentum. The first point to note in Equation (6.40) is that from simple mechanics the applied torque on a body is equal to the rate of change of angular momentum, viz

$$\mathbf{\Gamma}_a^{mech.} = \frac{d}{dt}\mathbf{L}_a^{mech.}. \tag{6.41}$$

The first term on the right was originally the field momentum, but now it has the extra factor of x^a involved. This must be the corresponding field angular momentum, so that

$$\mathbf{L}_a^{field} = \epsilon_0\int_{\mathcal{V}}d^3x\epsilon_{abc}x^b\mathbf{E}_d\epsilon^{cde}\mathbf{B}_e. \tag{6.42}$$

This can simplified using the tensor identity Equation (6.28) to

$$\mathbf{L}_a^{field} = \epsilon_0 \int_{\mathcal{V}} d^3x \mathbf{E}_a (x^b \mathbf{B}_b) - \mathbf{B}_a (x^b \mathbf{E}_b). \tag{6.43}$$

The remaining terms in Equation (6.40) can be written again in terms of the stress tensor so that

$$\begin{aligned} \frac{d}{dt}(\mathbf{L}_a^{mech.} + \mathbf{L}_a^{field}) &= \int_{\mathcal{V}} d^3x \epsilon_{abc} x^b \nabla_m T^{mc} \\ &= \int_{\mathcal{V}} d^3x \nabla_m \epsilon_{abc} x^b T^{mc}. \end{aligned} \tag{6.44}$$

Since the stress tensor is symmetric we have been able to put the x^a inside the brackets as it it differentiates to give a Kronecker delta. This in turn means the stress tensor indices contract with the ϵ_{abc} giving a null result. Since it is a total derivative we can again form a surface integral so that

$$\frac{d}{dt}(\mathbf{L}_a^{mech.} + \mathbf{L}_a^{field}) = \int_{\partial \mathcal{V}} d^2 \sum_m \epsilon_{abc} x^b T^{mc}. \tag{6.45}$$

We have arrived at another physical measurable and conservation law. The above is the statement of the conservation of angular momentum; the left hand side is the rate of change of the total (field plus mechanical) angular momentum and that this must equal the total torque due to the fields which is the measurable. The tensor quantity on the right is the angular equivalent of the stress tensor.

Can we measure the angular momentum of the field? It turns out to be possible but involves an extra degree of sophistication. Essentially (again) it requires transferring angular momentum to matter. The interested reader can consult [4] for more details of this advanced topic.

6.3 POLARIZATION (AGAIN) COURTESY OF STOKES

In Chapter 5, we encountered the polarization of the electromagnetic field - the fact that there are two physical degrees of freedom that have to be extracted from the electric and magnetic fields.

In that case, where we considered plane wave type solutions, it was enough to know which components of the electric and magnetic field were nonzero with respect to the direction of propagation given by the wave vector. We return to this topic now from the point of view of the physical measurables that this chapter has been focused on.

Experiment 11 (Étienne-Louis Malus) *Place a special material in front of a light source called a polarizer and a second similar polarizer (sometimes called an analyzer) after the first polarizer. Now rotate one with respect to the other. What happens?*

Observation 11 *The intensity of observed light (as seen by the eye for example) depends on the relative angle between the polarizer and the analyzer. Its functional form is* $I(\theta) = I(0)\cos^2(\theta)$. *This is known as Malus' law.*

Étienne-Louis Malus (1775–1812), formerly a military engineer in the army of Napoleon, first published his law in 1809. He also has his name on the Eiffel Tower.

As we have already mentioned, the electric or magnetic field by itself is not measurable for rapidly oscillating fields simply because it is too fast for a detector to respond to it. What we want to do for the state of polarization is trade up for a bilinear object, similar to the stress tensor or the Poynting vector that has already been found. Then we can apply the same averaging procedure to obtain a physical measurable. So consider the following object

$$O_{ab}(t,x) := \mathbf{E}_a(t,x)\mathbf{E}_b(t,x). \tag{6.46}$$

For the plane-wave like solutions we know already that the time average will be nonzero for components that contain the directions of polarization and zero otherwise. What we didn't see

earlier are two important aspects, firstly that there is a parameter space that describes the state of polarization even for rapidly oscillating fields. In addition to this, due to a lack of information about the source that sometimes occur (e.g. natural light), the polarization state may have a random element to it leading to an incomplete specification.

We now define the Stokes parameters for a plane wave propagating in the z-direction (in Cartesian coordinates). Four parameters are introduced to describe the state of polarization such that

$$S_0 := \langle \mathbf{E}_x(t,x)\mathbf{E}_x(t,x)\rangle + \langle \mathbf{E}_y(t,x)\mathbf{E}_y(t,x)\rangle, \tag{6.47}$$

$$S_1 := \langle \mathbf{E}_x(t,x)\mathbf{E}_x(t,x)\rangle - \langle \mathbf{E}_y(t,x)\mathbf{E}_y(t,x)\rangle, \tag{6.48}$$

$$S_2 := 2\langle \mathbf{E}_x(t,x)\mathbf{E}_y(t,x)\cos\epsilon(t)\rangle, \tag{6.49}$$

$$S_3 := 2\langle \mathbf{E}_x(t,x)\mathbf{E}_y(t,x)\sin\epsilon(t)\rangle. \tag{6.50}$$

Recall that the variable $\epsilon(t)$ is the phase difference between the x and y electric field components as given in Equation (5.59), but we have now allowed it to have a general time dependence. One can immediately verify that with these definitions the Stokes parameters satisfy

$$S_0^2 = S_1^2 + S_1^2 + S_1^2, \tag{6.51}$$

for fully polarized light i.e. where there is definite known amount of each component such that $\epsilon(t)$ is a constant.

For totally unpolarized light, by definition

$$\langle \mathbf{E}_x(t,x)\mathbf{E}_x(t,x)\rangle = \langle \mathbf{E}_y(t,x)\mathbf{E}_y(t,x)\rangle. \tag{6.52}$$

This implies that the Stokes parameters then take the form

$$S_0^2 \neq 0, S_1^2 = S_1^2 = S_1^2 = 0. \tag{6.53}$$

George Gabriel Stokes (1819–1903), mathematician, physicist, theologian, and politician. The parameters which bear his name were published in 1852 and then were forgotten for some time. He also held the Lucasian Professor of Mathematics chair at Cambridge.

There is a nice way of thinking about the state of polarization geometrically. The degree of polarization is given by a point that lies on or in a three dimensional ball. If the point lies on the surface, called the Poincaré sphere, the state is completely polarized. Inside the ball we have in general a partially polarized state - a mixture of unpolarized and polarized. At the center the state is completely unpolarized. Collectively then any degree of polarization can be represented by the inequality

$$S_0^2 \geq S_1^2 + S_1^2 + S_1^2. \tag{6.54}$$

These can be related to the polarization states associated with plane waves (linear, circular, and elliptical) encountered in Chapter 5. To aid in the presentation we group the four Stokes parameters into the object $\mathcal{S} = (S_0,\ S_1,\ S_2,\ S_3)$, with the following statements:

- **Linear polarization** - this is when $\mathcal{S} = (S_0 \neq 0,\ S_1 = \pm S_0,\ S_2 = 0,\ S_3 = 0)$. The positive sign indicates linear polarization in the x direction, while the minus sign corresponds to y direction polarization.
- **Circular polarization** - here we require that $\mathcal{S} = (S_0 \neq 0,\ S_1 = 0,\ S_2 = 0,\ S_3 = \pm s_0)$. The positive is left circularly polarized, the negative is right circularly polarized.
- **Elliptical polarization** - this is the most general state and any particular point on the sphere will correspond to an elliptically polarized state.

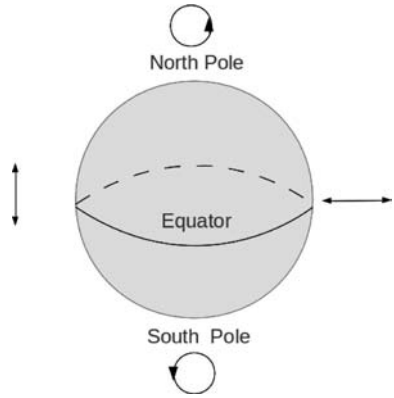


FIGURE 6.2: **The Poincaré sphere** - The different states of polarization are given by a point on the bounding sphere while point in the interior are only partially polarized.

In Figure 6.2, one can see graphically how the different polarization states sit on the sphere. There is the further geometrically interesting property that rotations on the Poincaré sphere correspond to transformations of the state of polarization. Thus we can, with the appropriate definitions, start to build this into a vector space and consider other operations such as additions. Basic group theory can be used to describe manipulations of polarized beams. See [14] for some useful discussion.

Jules Henri Poincaré (1854–1912), another polymath in the history of science. He excelled in all areas, contributing to topology, chaos, and mathematical physics. The Poincaré group is the symmetry which is built into all relativistic field theories.

6.4 SUMMARY

The aim of this chapter was to establish the basic field theory measurables and the principles of their measurement. In the first instance voltages and currents that we have applied externally were

used to understand the coarse grained dynamics of simple circuits. Then we have seen that to observe the field properties (energy, momentum, angular momentum and its polarization) it is necessary to consider bilinear field quantities as the observables. These characterize the electromagnetic field as a dynamical quantity very much like what one encounter in classical mechanics. Starting point for topics discussed in this chapter can be found in [8] and [7].

CHAPTER 7

DISTRIBUTIONS OF CHARGE, MACROSCOPIC MATTER, AND BOUNDARY CONDITIONS

In this chapter, we will consider the implications of how real matter interacts with the physical electromagnetic field. There are many subtle problems associated in trying to understand the basic mechanisms. We want to be able to think about this without resorting to any sort of "fundamental" microscopic theory. The principle difficulty one encounters is in applying electromagnetism to account for structure of matter. The question is that if Coulomb law holds at the microscopic level then what is the underlying stability of matter - why doesn't everything just collapse in on itself or blow apart. As the reader is no doubt aware, it is in the arena of quantum mechanics that this has to be sorted out. However, this book is on the classical theory of electromagnetic fields. One must be careful not to tread in other pastures, lest we not treat the topic consistently.

Thinking about this then somewhat further, one can ask how the fields change when they enter into or are confined in normal solid

matter. In such circumstances one needs to know what happens at the separating intermediate surface between two such media. If we were to regard Coulomb's law as something very basic that would continue to be true at microscopic distances (so that it accounts for the structure of matter), then as we look at smaller and smaller length scales, how does the distribution of charge appear? What happens to this charge distribution as we change length scale? Necessarily, this will have to be an in part phenomenological approach because we do not know the microscopic theory. So the basic physical observables and measurements should not lose sight of the phenomenological assumptions.

As in previous chapters, the setup and formalism is such that the electric and magnetic cases are treated in an analogous mirror like fashion. The astute reader has probably become aware that the magnetic cases tend to be a little bit more complicated than the equivalent electric one. The basic reason for this, in a general sense, is that for the magnetic field we are always finding a rotational angular momentum type variable that is intertwined with the magnetic field. In the electric case, it is typically a linear distance or velocity that has the equivalent connection.

7.1 MULTIPOLES

Thus far, we have developed the theory and measurement of electromagnetic fields by themselves. Of course the measurement process requires that these fields interact with our measuring devices, but that they themselves are not really disturbed. However, they are of course made from real matter so the question now turns to how fields interact with real matter. Let us revisit the parallel plate capacitor à la Faraday [6] and consider the following experiment.

Experiment 12 (Faraday, 1836) *Take the parallel plate capacitor, apply a voltage across the plates and measure its capacitance with an electrometer. What happens when we insert an insulator in between the plates?*

Observation 12 *The capacitance is found to increase as the the measured charge on the plates increases.*

The experimental apparatus Faraday used in 1836 to quantitatively study the effect of dielectrics in between capacitor plates. In this version, the spherical capacitor plates geometry removed any doubts about the particular geometric configurations (e.g. edge effects).

From this experiment we establish the fact that the charge density has changed on the plates. So how should we interpret this? If the charge density on the plates has changed then the charge density in the insulator must also have changed to produce this effect. But it can't be the simple addition or subtraction of charge because this would lead to a violation of the conservation of charge. It would give an observable net Coulomb like force outside the capacitor. There are obviously charges inside the insulator, but they must be in the form of *bound* states - groups of charge that form a stable unit. When the insulator is placed between the plates, the electric field due to the potential difference must cause the charges in the insulator to be redistributed. In terms of the bound state they are displaced with respect to one another, but relax back to their original configuration when they are taken out from the capacitor plates. This then is the picture we have of the mechanism by which the interaction takes place.

Let us try and capture this in terms of the field equations [8]. Consider Equation (4.53) (**Maxwell I**) where a charge density is a source for an electric field. We can actually think about separating the charge density into different contributions, depending on a length scale and whether it is free charge or is part of a bound state. If we do this then

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0}\left(\rho^{Free} + \rho^{Bound}\right). \tag{7.1}$$

By free charge we mean charge that can run around in the material subject only to resistive forces (collisions, but what really

are collisions?) that will eventually bring it to a stop. For a bound state charge, a phenomenological model could be a pair of equal and opposite charges that have an harmonic potential binding them together (a spring). This is perhaps the simplest example of a bound state charge distribution. Note there that the spring constant would have to be introduced by hand - it is a phenomenological parameter that isn't calculated but would have to be measured. In general, however, the distribution of charge in a bound state will be more complicated and so we now address this point.

Consider the following model of a collection of charges. To start off with let us assume they are point like so that the charge density can be expressed in terms of a number of Dirac delta functions and a set of center of mass coordinates. As before we make a distinction between mobile charge and bound charge. Let us assume that the material we are considering has both mobile and bound charges present, where for the time being we focus on the bound charges. Next, let us divide the charges up into groups (or systems of charge that we shall colorfully refer to as "molecules") that we index by N. Around each of these we can surround them with a small volume and regard it as an object in its own right (this is exactly how the molecular structure of matter is developed). So this grouping looks like

$$\rho^{Bound}(t,x) = \sum_N \rho_N(t,x) = \sum_N \rho(t,x-x_N), \tag{7.2}$$

where x_N is the center of mass coordinate for the molecule N. Next, we use the seemingly trivial equation

$$\rho(t,x) = \int d^3y \delta^3(y)\rho(t,x-y), \tag{7.3}$$

$$\rho_N(t,x) = \int d^3y \delta^3(y)\rho_N(t,x-y). \tag{7.4}$$

Since we are thinking about a large number of systems, that is an ensemble of distinct charge distributions, this system starts to take on a statistical character. To that end it is sensible to replace the above equations with spatially averaged versions. This means replacing the delta function distribution in the above with a statistical test function that localizes the function in a similar way but as applied to a large number of charge groups. We define a spatial averaging, $\langle \cdots \rangle$, to be

realized by a test function $\pi(x)$ (a smooth and finite version of a Dirac delta function) such that

$$\langle f(t,x)\rangle := \int d^3y \pi(y) f(t,x-y). \tag{7.5}$$

It has the effect of localizing any function about some preferred point. As it is a statistical function, the length scale which the averaging is done over should be much larger than the molecular dimension. Next we consider the charge density about the individual center of mass coordinates using the previous test function. Then Equation (7.4) becomes

$$\langle \rho_N(t,x)\rangle = \int d^3y \pi(y) \rho_N(t,x-y) \tag{7.6}$$

$$= \int d^3y \pi(y) \sum_i q_{(i,N)} \delta^3(x-y-x_N-x_{(iN)}) \tag{7.7}$$

$$= \sum_i q_{(i,N)} \pi(x-x_N-x_{(iN)}). \tag{7.8}$$

The spatially averaged charge density has been decomposed into a sum of contributions about each groups center of mass indicated by the index i. Since the spatial averaging occurs on a length scale much bigger than $x_{(i,N)}$ (that is $\pi(x - x_N)$ is slowly varying on the length scale of the interchange separations), it is a simple matter to perform a Taylor series expansion in powers of $x_{(i,N)}$

$$\langle \rho_N(t,x)\rangle = \sum_i q_{(i,N)} [\pi(x-x_N) - x^a_{(iN)} \nabla_a \pi(x-x_N) + \frac{1}{2} x^a_{(i,N)} x^b_{(i,N)} \nabla_a \nabla_b \pi(x-x_N)] + \cdots, \tag{7.9}$$

from which we define the multipole moments of the N-th molecule as

$$q_N = \sum_i q_{(i,N)} = \text{molecular charge}, \tag{7.10}$$

$$d^a_N = \sum_i q_{(i,N)} x^a_{(i,N)} = \text{molecular dipole moment}, \tag{7.11}$$

$$Q^{ab}_N = 3 \sum_i q_{(i,N)} x^a_{(i,N)} x^b_{(i,N)} = \text{molecular quadrupole moment}. \tag{7.12}$$

The factor of three in the quadrupole moment is a convenient definition for normalization only and nothing fundamental. Given the definition of the test function, together with the delta function nature of the charge distributions, this can be recast as

$$\langle \rho_N(t,x)\rangle = \langle q_N \delta^3(x-x_N)\rangle - \nabla_a \langle d_N^a \delta^3(x-x_N)\rangle + \frac{1}{6}\nabla_a \nabla_b \langle Q_N^{ab}\delta(x-x_N)\rangle + \cdots \tag{7.13}$$

It is now straightforward to sum over all the N molecules to obtain the macroscopic polarization. Equation (7.1) has to have the charge density sources replaced with the spatially averaged versions and can be written as

$$\nabla_a \epsilon_0 \mathbf{E}^a(t,x) = \langle \rho_{free}(t,x)\rangle + \sum_N \langle \rho_N(t,x)\rangle \tag{7.14}$$

$$= \langle \rho_{free}(t,x) + \sum_N q_N \delta^3(x-x_N)\rangle - \nabla_a \langle d_N^a \delta^3(x-x_N)\rangle + \nabla_a \nabla_b \langle Q_N^{ab}\delta(x-x_N)\rangle + \cdots \tag{7.15}$$

Now we make the following definitions

$$\rho(t,x) := \langle \rho_{free}(t,x) + \sum_N q_N \delta^3(x-x_N)\rangle \tag{7.16}$$

$$= \text{macroscopic charge density},$$

$$\mathbf{P}^a(t,x) := \sum_N \langle d_N^a \delta^3(x-x_N)\rangle \tag{7.17}$$

$$= \text{macroscopic polarization density},$$

$$\mathbf{Q}^{ab}(t,x) := \sum_N \langle Q_N^{ab}\delta^3(x-x_N)\rangle \tag{7.18}$$

$$= \text{macroscopic quadrupole density}. \tag{7.19}$$

With these definitions, the resulting field equation is

$$\nabla_a(\epsilon_0 \mathbf{E}^a + \mathbf{P}^a - \nabla_b \mathbf{Q}^{ab} + \cdots) = \rho \tag{7.20}$$

$$\nabla_a \mathbf{D}^a = \rho, \tag{7.21}$$

where it is customary to denote the combination of the bare electric field and the multipole additions as the *displacement* field, defined as

$$\mathbf{D}^a := \epsilon_0 \mathbf{E}^a + \mathbf{P}^a - \nabla_b \mathbf{Q}^{ab} + \cdots \tag{7.22}$$

In relation to Equations (7.21) and (7.22) it is clear what this means physically if the displacement field is linear in the electric field: the measured charge is greater than without the dielectric, so the electric field must be modified inside the dielectric. If there are only point sources which can attract or repel then they must be in the form of bound states, that is stable collections of spatially distributed charge, and these are exactly the multipole moments we have already worked out. A further point to mention is that we have been deliberately vague about the length scales involved. If we assume the atomic theory of matter and molecular theory we can indeed put in numbers here. Instead, what we have shown here is how one can go about establishing such a theory. By using a macroscopic theory plus some phenomenology, one can start building up such a model.

In an exactly similar fashion, the analogous situation for the magnetic field can be carried out. Equation (4.56) (**Maxwell IV**) has to be modified for both electric and magnetic fields. The equivalent of Equation (7.1) for bound state currents is

$$\nabla \wedge \mathbf{B} - \mu_0 \epsilon_0 \, \partial_t \mathbf{E} = \mu_0 \left(\mathbf{J}^{Free} + \mathbf{J}^{Bound} \right). \tag{7.23}$$

Similarly to Equation (7.2), the bound state current density is

$$\langle \mathbf{J}^{Bound}(t,x) \rangle = \sum_N \langle \mathbf{J}_N(t,x) \rangle. \tag{7.24}$$

The starting point to turn this into the equivalent of (7.21) is to reconsider the current density in the light of Equation (7.7) such that

$$\langle \mathbf{J}_N(t,x) \rangle = \int d^3 y \pi(y) \mathbf{J}_N(t, x-y). \tag{7.25}$$

The expression for $\mathbf{J}_N$ takes the same form as for charge density (i.e., the argument is centered around the center of mass coordinate

and the charge to center separation), but in addition to this we should not lose sight of the velocities of the collection of charges that collectively constitute the basic current. One has

$$\mathbf{J}_N(t,x-y)=\sum_i q_{(i,N)}(\mathbf{v}_{(i,N)}+\mathbf{v}_N)\delta^3(x-y-x_N-x_{(iN)}), \tag{7.26}$$

$$\langle\mathbf{J}_N(t,x)\rangle=\sum_i q_{(i,N)}(\mathbf{v}_{(i,N)}+\mathbf{v}_N)\pi(x-x_N-x_{(iN)}). \tag{7.27}$$

As before, it is necessary to expand the spatial averaging function in a power series in $x_{(iN)}$. To simplify matters we shall perform the expansion only to linear order in $x_{(iN)}$. Doing this one finds

$$\begin{aligned}\langle\mathbf{J}_N(t,x)\rangle=&\sum_i q_{(i,N)}(\mathbf{v}_{(i,N)}+\mathbf{v}_N)\pi(x-x_N)\\&-\sum_i q_{(i,N)}x^a_{(i,N)}(\mathbf{v}_{(i,N)}+\mathbf{v}_N)\nabla_a\pi(x-x_N)+\cdots\\&=\sum_i q_{(i,N)}\mathbf{v}_N\pi(x-x_N)\\&+\sum_i q_{(i,N)}\mathbf{v}_{(i,N)}\pi(x-x_N)\\&-\sum_i q_{(i,N)}x^a_{(i,N)}\mathbf{v}_N\nabla_a\pi(x-x_N)\\&-\sum_i q_{(i,N)}x^a_{(i,N)}\mathbf{v}_{(i,N)}\nabla_a\pi(x-x_N)+\cdots.\end{aligned} \tag{7.28}$$

The reason for writing the above in this format is that each line has a separate interpretation and can therefore rewritten in simpler terms. The first line in Equation (7.28) is the average current density of molecule N

$$\sum_i q_{(i,N)}\mathbf{v}_N\pi(x-x_N)=\langle q_N\mathbf{v}_N\delta^3(x-x_N)\rangle. \tag{7.29}$$

The second line in Equation (7.28) gives zero since on average there is no current density associated with the relative velocities, so that

$$\sum_i q_{(i,N)}\mathbf{v}_{(i,N)}\pi(x-x_N)=0. \tag{7.30}$$

Looking now at Equation (7.9), the partial time derivative of the dipole term is just the third line above

$$\partial_t \sum_i q_{(i,N)} x^a_{(i,N)} \pi(x - x_N) = -\sum_i q_{(i,N)} x^a_{(i,N)} \mathbf{v}_N \cdot \nabla \pi(x - x_N), \quad (7.31)$$

so that the third line in Equation (7.9) is

$$\partial_t \langle d^a_N \delta(x - x_N) \rangle = -\sum_i q_{(i,N)} x^a_{(i,N)} \mathbf{v}_N \cdot \nabla \pi(x - x_N). \quad (7.32)$$

The last line that involves both relative separations and velocities requires further work once we have simplified the current density. Returning to Equation (7.23) one can now substitute in the the molecular density as follows

$$\begin{aligned} \nabla \wedge \mathbf{B} - \mu_0 \epsilon_0 \partial_t \mathbf{E} &= \mu_0 (\langle \mathbf{J}^{free} + \sum_N \mathbf{J}_N(t,x) \rangle) \\ &= \mu_0 (\langle \mathbf{J}^{free} + \sum_N q_N \mathbf{v}_N \delta^3(x - x_N) \rangle) \\ &\quad + \mu_0 \partial_t \langle d_N \delta^3(x - x_N) \rangle \\ &\quad - \mu_0 \sum_N \sum_i q_{(i,N)} x^a_{(i,N)} \mathbf{v}_{(i,N)} \nabla_a \pi(x - x_N) \\ &\quad + \cdots . \end{aligned} \quad (7.33)$$

The above immediately shows that there is a clear contribution to the current density in the first line. In the second line, the partial time derivative of the dipole moment becomes the partial time derivative of the polarization density which we have already encountered. This can be put together with the electric field contribution to form the displacement field. The last term presents the genuinely new term. We want to factor this in with the $\mathbf{B}^a$ field, but to do this requires the Levi-Civita tensor. It is therefore necessary to use the identity Equation (6.28) again so that

$$\epsilon^{dab} \epsilon_{dmn} x^m_{(i,N)} \mathbf{v}_{b(i,N)} = x^a_{(i,N)} \mathbf{v}_{n(i,N)} - \delta^a_n (x_{(i,N)} \cdot \mathbf{v}_{(i,N)}). \quad (7.34)$$

However, the molecular separations and velocities are orthogonal so the last term above vanishes. We now have two wedge products,

one of which can be used in conjunction with the magnetic field. Using this in Equation (7.33) one finds

$$\begin{aligned}
\nabla \wedge \mathbf{B} - \mu_0 \sum_N \sum_i q_{(i,N)} x^a_{(i,N)} \mathbf{v}_{(i,N)} \nabla_a \pi(x - x_N) \\
-\mu_0 \epsilon_0 \, \partial_t \mathbf{E} - \mu_0 \, \partial_t \langle d_N \delta(x - x_N) \rangle \\
= \mu_0 \langle \mathbf{J}^{free} + \sum_N q_N \mathbf{v}_N \delta(x - x_N) \rangle \\
+ \cdots \\
\nabla \wedge \mathbf{B} - \mu_0 \sum_N \sum_i q_{(i,N)} x_{(i,N)} \wedge \mathbf{v}_{(i,N)} \wedge \nabla \pi(x - x_N) \\
-\mu_0 \epsilon_0 \, \partial_t (\mathbf{E} + \langle d_N \delta(x - x_N) \rangle) \\
= \mu_0 \langle \mathbf{J} \rangle + \cdots . \qquad (7.35)
\end{aligned}$$

The magnetization and the macroscopic current are thus defined as

$$\begin{aligned}
\mathbf{J}(t,x) &:= \langle \mathbf{J}^{free} + \sum_N q_N \mathbf{v}_N \delta(x - x_N) \rangle \qquad (7.36) \\
&= \text{macroscopic current density,} \\
\mathbf{M}^a(t,x) &:= \sum_N \langle m^a_N \delta^3(x - x_N) \rangle \\
&= \text{macroscopic magnetization density,} \qquad (7.37) \\
m^a_N &:= \sum_i q_{(i,N)} x_{(i,N)} \wedge \mathbf{v}_{(i,N)} \\
&= \text{molecular magnetic moment.} \qquad (7.38)
\end{aligned}$$

Therefore, Equation (7.35) can be recast as

$$\nabla \wedge \mathbf{H} - \partial_t \mathbf{D} = \mu_0 \mathbf{J}, \qquad (7.39)$$

$$\mathbf{H}^a := \frac{1}{\mu_0} \mathbf{B}^a - \mathbf{M}^a + \cdots , \qquad (7.40)$$

where we have only included the macroscopic magnetization in the magnetic field strength $\mathbf{H}^a$.

From the definitions of the two macroscopic fields we can now define the two important material properties. These are the permittivity tensor ϵ^a_b and the permeability tensor μ^a_b and are given by

$$\epsilon_b^a := \frac{\delta \mathbf{D}^a}{\delta \mathbf{E}^b} = \delta_b^a + \frac{\delta \mathbf{P}^a}{\delta \mathbf{E}^b} + \cdots, \tag{7.41}$$

$$\mu_b^a := \mu_0 \frac{\delta \mathbf{H}^a}{\delta \mathbf{B}^b} = \delta_b^a + \frac{\delta \mathbf{M}^a}{\delta \mathbf{B}^b} + \cdots. \tag{7.42}$$

The above relations constitute the manner in which the physical fields interact with phenomenological models of matter. As such, they serve as background fields which determine how the electromagnetic fields themselves behave in matter.

7.1.1 Interlude on Electric and Magnetic Dipoles

The multipole expressions just encountered are quite general. It is often the case, however, that the dipole terms represent the most interesting part of the total expressions. This is so because they represent the largest contribution to the series. The next set of observables one would want to consider are the couplings of the electric and magnetic dipoles to external electric and magnetic fields. The natural set of observables to consider are obviously the observed forces, but in addition, since a distance scale is involved (the spatial separation of the charge or the radius of the current loop), a torque is also now possible.

To derive expressions for the forces and torques for a simple (molecular) electric and magnetic dipole one must appreciate that this is two particle problem. Considering Equations (7.11) and (7.38) for a two particle system reveals the following. The electric dipole consists of two equal and opposite charges $\pm q$ with some separation between them (we consider here the special case of two equal masses to make the bound states nice symmetric objects and simplify the calculations), while the magnetic dipole is two equal and opposite charges circulating around a common center of mass. We can therefore write down two equations of motion (The Lorentz force law) that governs each charge in the presence of spatially varying electric and magnetic fields. Firstly, consider the electric dipole

$$\mathbf{F}_a(x_{(1)}) = q\mathbf{E}_a(x_{(1)}), \tag{7.43}$$

$$\mathbf{F}_a(x_{(2)}) = -q\mathbf{E}_a(x_{(2)}). \tag{7.44}$$

We now expand each position vector of the charges about the center of mass coordinate X in a power series in their separation vector x, which are defined by

$$X := (x_{(1)} + x_{(2)})/2, \tag{7.45}$$

$$x := x_{(1)} - x_{(2)}. \tag{7.46}$$

Replacing the arguments of the external fields with the coordinates (X, x) gives a power series in x

$$\mathbf{F}_a(x_{(1)}) = q\mathbf{E}_a(X + x/2) \tag{7.47}$$

$$= q\mathbf{E}_a(X) + \frac{q}{2}x^a\nabla_a\mathbf{E}_a(X) + \mathcal{O}(x^2), \tag{7.48}$$

$$\mathbf{F}_a(x_{(2)}) = -q\mathbf{E}_a(X - x/2) \tag{7.49}$$

$$= -q\mathbf{E}_a(X) + \frac{q}{2}x^a\nabla_a\mathbf{E}_a(X) - \mathcal{O}(x^2). \tag{7.50}$$

Therefore, adding together the two forces on the dipole results in

$$\mathbf{F}_a(X) = \mathbf{F}_a(x_{(1)}) + \mathbf{F}_a(x_{(2)}) \tag{7.51}$$

$$= qx^b\nabla_b\mathbf{E}_a(X) + \mathcal{O}(x^2) \tag{7.52}$$

$$= d^b\nabla_b\mathbf{E}_a(X) + \mathcal{O}(x^2). \tag{7.53}$$

One can see that in a uniform electric field, the dipolar force vanishes to a first order approximation.

The torque is simply evaluated now about the center mass coordinate to be

$$\mathbf{\Gamma}_a(X) = \frac{1}{2}\epsilon_{abc}x^b\mathbf{F}^c(x_{(1)}) + \frac{1}{2}\epsilon_{abc}x^b\mathbf{F}^c(x_{(2)}) \tag{7.54}$$

$$= q\epsilon_{abc}x^b\mathbf{E}^c(X) \tag{7.55}$$

$$= \epsilon_{abc}d^b\mathbf{E}^c(X). \tag{7.56}$$

So for a constant electric field, a dipole will be acted upon to produce a torque until the vector d^a is aligned with the field.

Turning our attention now to the magnetic dipole we start again with the Lorentz Force law equations where a spatially varying magnetic field is present

$$\mathbf{F}(x_{(1)}) = q\mathbf{v}_{(1)}(x_{(1)}) \wedge \mathbf{B}(x_{(1)}), \tag{7.57}$$

$$\mathbf{F}(x_{(2)}) = -q\mathbf{v}_{(2)}(x_{(2)}) \wedge \mathbf{B}(x_{(2)}). \tag{7.58}$$

As before we expand the magnetic field about the center of mass coordinates so that

$$\mathbf{F}(x_{(1)}) = q\mathbf{v}_{(1)}(X + x/2) \wedge \mathbf{B}(X + x/2) \tag{7.59}$$

$$= q\mathbf{v}_{(1)}(X) \wedge \mathbf{B}(X) + \frac{q}{2}\mathbf{v}_{(1)}(X) \wedge x^a \nabla_a \mathbf{B}(X) + \frac{q}{2} x^a \nabla_a \mathbf{v}_{(1)}(X) \wedge \mathbf{B}(X) + \mathcal{O}(x^2), \tag{7.60}$$

$$\mathbf{F}(x_{(2)}) = -q\mathbf{v}_{(2)}(X - x/2) \wedge \mathbf{B}(X - x/2) \tag{7.61}$$

$$= -q\mathbf{v}_{(2)}(X) \wedge \mathbf{B}(X) + \frac{q}{2}\mathbf{v}_{(2)}(X) \wedge x^a \nabla_a \mathbf{B}(X) + \frac{q}{2} x^a \nabla_a \mathbf{v}_{(2)}(X) \wedge \mathbf{B}(X) + \mathcal{O}(x^2). \tag{7.62}$$

For the symmetrical pair of circulating charges we are considering, $\mathbf{v}_{(1)}(X) = \mathbf{v}_{(2)}(X) = \mathbf{v}$. The total force on the magnetic dipole is just the sum of the above two

$$\mathbf{F}(X) = \mathbf{F}(x_{(1)}) + \mathbf{F}(x_{(2)}) \tag{7.63}$$

$$= q\mathbf{v} \wedge x^a \nabla_a \mathbf{B}(X) + q x^a \nabla_a \mathbf{v} \wedge \mathbf{B}(X) + \mathcal{O}(x^2). \tag{7.64}$$

$$= q x^a \nabla_a (\mathbf{v} \wedge \mathbf{B}(X)) + \mathcal{O}(x^2) \tag{7.65}$$

We know that to form the magnetic dipole moment we require at least an extra Levi-Civita tensor so we use Equation (6.28) to

introduce the necessary wedge product. This leads to, after some calculation

$$\mathbf{F}(X) = q(x \wedge \mathbf{v}) \wedge \nabla \wedge \mathbf{B}(X) + q\nabla_a(\mathbf{v}^a x \wedge \mathbf{B}(X)) \tag{7.66}$$

$$= m \wedge \nabla \wedge \mathbf{B}(X) + q\nabla_a(\mathbf{v}^a x \wedge \mathbf{B}(X)). \tag{7.67}$$

The last term is zero for the special case of the circular symmetric dipole. This can be seen by taking $X = 0$ and working in cylindrical polar coordinates. In this case the velocity only has a θ component, while $\mathbf{B}(0)$ and x are independent of θ. The force on the magnetic dipole then further simplifies to

$$\mathbf{F}(X) = m \wedge \nabla \wedge \mathbf{B}(X) \tag{7.68}$$

$$= \nabla(m_a \mathbf{B}^a(X)) - m(\nabla_a \mathbf{B}^a(X)) \tag{7.69}$$

$$= \nabla(m_a \mathbf{B}^a(X)). \tag{7.70}$$

As one can see, the magnetic case is considerably more awkward to deal with. However, the torque on a magnetic dipole is nice and simple. Evaluating this in the center of mass coordinates and using Equation (7.54) one finds

$$\mathbf{\Gamma}_a(X) = qx \wedge \mathbf{v} \wedge \mathbf{B}(X) \tag{7.71}$$

$$= m \wedge \mathbf{B}(X). \tag{7.72}$$

If we were to work at the next order of approximation i.e., quadrupoles, the type and number of manipulations will be far greater. For other derivations see [8] and [5].

7.2 BOUNDARY CONDITIONS

As the field equations stand, we could in principle try to find solutions for the electric and magnetic fields when they are in a material region as described by the permittivity and permeability tensors. So far, we have really been thinking about a *continuous*

medium - a region of space where the material properties are described by smoothly varying functions. But what happens if we come across a sharp boundary, for example a glass-air interface? We know that the material properties will change *discontinuously* across such a space. What is the right way of handling this?

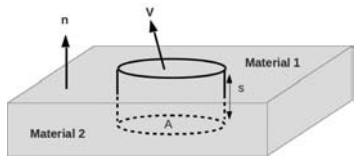
Considering the four Maxwell's equations, we are asking for solutions of the two bare fields $\mathbf{E}^a$ and $\mathbf{B}^a$ and the two dressed fields $\mathbf{D}^a$ and $\mathbf{H}^a$, in the two different regions. We should certainly be able to find separate solutions for the fields in the two separate regions. Since the field equations involve derivatives of fields, we can see that we will potentially run into trouble using them because of the discontinuities. However, if we recast them in their integral form we are more likely to have success as a number of the spatial derivatives will vanish.

To that end, let us start with the scalar equations, **Maxwell II** and the modified **Maxwell I**, Equation (7.21), and write them in their integral form

$$\int_{\partial\mathcal{V}} d^2\Sigma_a \mathbf{B}^a = 0, \tag{7.73}$$

$$\int_{\partial\mathcal{V}} d^2\Sigma_a \mathbf{D}^a = Q. \tag{7.74}$$

Suppose we consider a small volume that is located on the boundary of two distinct material regions that intersects both regions. If the box is "thin" then the faces that are parallel to the the separating region space will have the bigger areas. We can approximate the integral above by just a sum over the areas of the top and bottom faces.



The continuity conditions are obtained by considering a cylinder with area A (with normal vector n) upper and lower surfaces and thickness s. It is a thin cylinder which means $s << \sqrt{A}$. The vector $\mathbf{V}$ has a negligible scalar product with the side area normal.

Of course, we still have to form the scalar product with the surface normal which leads to the result

$$(\mathbf{B}_{(1)} - \mathbf{B}_{(2)})^a n_a = 0, \tag{7.75}$$

where n^a is the surface normal to the interface region. By the exact same construction we can derive an analogous result for the displacement field

$$(\mathbf{D}_{(1)} - \mathbf{D}_{(2)})^a n_a = \sigma, \tag{7.76}$$

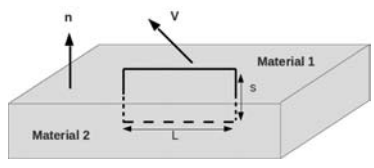
where σ is the surface charge residing on the interface region. Because we have taken the limit that the volume is thin, the charge density is then distributed in a volume that sits on the interface. So the volume charge density becomes localized to a surface charge density.

If we now consider the vector field equations we can adopt the same procedure as before and write them in their integral form

$$\int_{\partial \mathcal{A}} dx^a \mathbf{E}_a + \int_{\mathcal{A}} d^2\Sigma^a \, \partial_t \mathbf{B}_a = 0, \tag{7.77}$$

$$\int_{\partial \mathcal{A}} dx^a \mathbf{H}_a - \int_{\mathcal{A}} d^2\Sigma^a \, \partial_t \mathbf{D}_a = I. \tag{7.78}$$

It is necessary now to cut the interface region in a different way. Since we have a line integral and an area, we can see that we need to form a small loop that cuts both regions and that forms a small area. The small sides are normal to the interface surface and the long sides parallel to it.



The continuity conditions are obtained by considering a rectangular loop with length L and thickness s that intersects both regions. It is a small loop which means $s << L$. The vector $\mathbf{V}$ has a negligible scalar product with the side element vectors.

As the small sides tend to zero $(s \to 0)$, the area also tends to zero so that the area integrals featuring the time derivative of the **B** and **D** fields vanish. The loop integral can be split into four pieces. The small side elements which are in the direction of the normal to the interface surface vanish i.e., the scalar product projection with the normal. This means it is only the orthogonal projection of the normal given by the wedge product that is nonzero. Therefore

$$(\mathbf{E}_{(1)} - \mathbf{E}_{(2)}) \wedge n = 0. \tag{7.79}$$

To obtain the final continuity equation we need the analogous situation of the surface charge previously encountered. One must include a surface current density **w** so that

$$(\mathbf{H}_{(1)} - \mathbf{H}_{(2)}) \wedge n = \mathbf{w}. \tag{7.80}$$

With these equations one can now glue solutions in two distinct material regions at their common boundary. Good discussions can be found in [8] and [5].

7.3 ELECTROMAGNETIC INTERACTIONS WITH MATTER

A fundamental topic to address is the nature of how and the effect electromagnetic fields have on real physical matter. Indirectly this topic is implicit in all of the discussions so far, as the control of charge and current has been performed by us using physical matter. We now wish to go further and ask in what way the matter *reacts* in the presence of physical fields, be they oscillating, static, large, or small. We will assume that a mixture of macroscopic and microscopic modelling of bulk real matter will be sufficient for this purpose. In reality, we would really need to think about the quantum nature of matter in order to get a consistent calculation scheme in which to talk about atoms and their coupling to electromagnetic fields. Turning this around, however, we can try to push these models as far as

they can go as a way to better understand the necessity of any quantum scheme. You have to start somewhere!

Our first macroscopic model for bulk matter then will be to assume that it consists of some type of lattice of bound state charges, that overall has no net electrical charge. Suppose we consider one such pair (i.e. we look at the local charge distribution at some point) and ask what is the effect of applying an electric field to it. The charges must be displaced by some small distance, where we are assuming that the bounding potential can admit a new equilibrium position. Indeed, thinking about the simple molecular models encountered earlier in the chapter, we can imagine that they are flexible in the presence of applied fields. Alternatively, think of the sample as being overall charge neutral but that we know it has positive and negative charge degrees of freedom inside it. Suppose they are sitting on top of one another. By applying an external field to the sample we force such pairs to separate.

The electric polarization $\delta P^a(t, x)$ for a collection of a bound state of charges (the $\pm q$ pair of the electric dipole) is then defined to be

$$\delta P^a(t,x) \coloneqq qn(t,x)\langle \delta x^a(t,x)\rangle, \tag{7.81}$$

where $n(t, x)$ is number density distribution for the collection of bound states (for the time being we set $\epsilon_0 = 1$ and reinstate it at the end) and $\delta x^a(t, x)$ is the increase in separation from equilibrium of the charge pair. The brackets indicate that since the bound state is a microscopic structure and we are applying the field on a macroscopic sample, we should think of it in some statistical sense (if we were using quantum mechanics here, this would simply be the expectation value between ground states). In complete generality this happens due to the application of some electric field at an earlier time and some other point in space $\delta \mathbf{E}_b(t', x')$. There is then a polynomial relationship between the two

$$\begin{aligned}\delta P^a(t,x) = &\int dt' d^3x' \chi^{ab}(t-t', x-x')\delta \mathbf{E}_b(t',x') \\ &+ \int dt' d^3x' \int dt'' d^3x'' \xi^{abc}(t,t',t'',x,x',x'') \\ &\times \delta \mathbf{E}_b(t',x')\delta \mathbf{E}_c(t'',x'') + \mathcal{O}(\delta \mathbf{E}^3).\end{aligned} \tag{7.82}$$

The first term defines the linear theory, while the second and higher orders define the non-linear theory. The quantities χ^{ab} and ξ^{abc} are called in general *response functions*. They are clearly non-local in time and in space and tell us how strongly a system reacts to given input (or test) field. Therefore, they also represent physical observables. See [1] for a nice discussion.

We will now make two simplifying assumptions which will allow us to derive useful relations and that also have physical relevance. The first is that we shall work at the linear level. Not surprisingly, this makes many calculations tractable. Secondly, we will consider response functions that do not depend on space so that Equation (7.82) can be written as

$$\delta P^a(t,x) = \int dt' \chi^{ab}(t-t',x)\delta \mathbf{E}_b(t',x). \tag{7.83}$$

Clearly in this form the right hand side is a convolution integral. It therefore affords simple multiplicative form when Fourier transformed, by the use of the convolution theorem. In frequency space it becomes

$$\delta P^a(\omega,x) = \chi^{ab}(\omega,x)\delta \mathbf{E}_b(\omega,x), \tag{7.84}$$

and thus

$$\chi^{ab}(\omega,x) = \frac{\delta P^a(\omega,x)}{\delta \mathbf{E}_b(\omega,x)}. \tag{7.85}$$

If we can calculate the the polarization as a function of electric field, we can extract the response function. This particular response function is known as the electric susceptibility. What may not be obvious is that it is basically the Green's function of the model. If we know the equation of motion of the charge then we can invert this and obtain the Green function. This is exactly the same as what we encountered for the RLC circuit and so carries over straightforwardly. We will do this now for a simple model of a damped harmonic oscillator.

Recall the Lorentz force law, Equation (4.15). For an harmonic oscillator that is damped, the equation of motion of the charge q with displacement vector $\delta x^a(t)$ takes the form

$$m\langle\delta\ddot{x}^a(t)\rangle+\gamma\langle\delta\dot{x}^a(t)\rangle+m\omega_0^2\langle\delta x^a(t)\rangle=\hat{O}\langle\delta x^a(t)\rangle=q\delta\mathbf{E}^a(t). \quad (7.86)$$

Here again on the left hand side we are considering the separation of charge to be a statistical quantity. So we just have invert the operator $\hat{O}$ to obtain the susceptibility. A short cut is possible here by just substituting in harmonic oscillator type solutions, i.e.

$$\delta\mathbf{E}^a(t)=\delta\mathbf{E}^a(\omega)e^{i\omega t}, \quad (7.87)$$

$$\langle\delta x^a(t)\rangle=\delta x^a(\omega)e^{i\omega t}. \quad (7.88)$$

Substituting these into Equation (7.86) and assuming a constant number density $n(t, x) = N$ gives

$$\delta x^a(\omega)=\frac{q}{m(\omega_0^2-\omega^2)+i\gamma\omega}\delta\mathbf{E}^a(\omega), \quad (7.89)$$

$$\delta P^a(\omega)=\frac{q^2N}{m(\omega_0^2-\omega^2)+i\gamma\omega}\delta\mathbf{E}^a(\omega). \quad (7.90)$$

We therefore find the response function to be

$$\chi_b^a(\omega)=\frac{q^2N\delta_b^a}{m(\omega_0^2-\omega^2)+i\gamma\omega}. \quad (7.91)$$

Our model is thus describing the following behavior. A macroscopic sample has an oscillating electric field applied to it. The field can see the local charge difference of the bound states that constitute the model and drives them to oscillate. This will then ultimately depend on the nature of the bound state forces as to how strong or weak their subsequent oscillations will be. The response function represents another physical observable, this time about the macroscopic sample being probed. The local charge oscillations cause the input field to be deviated (or scattered) in some fashion and we will observe this as an altered output field. This is basis for explaining the laws of refraction in terms of a refractive index by linking it to the electric susceptibility.

A brief comment is also in order about the equivalent magnetic system, where a fluctuating test magnetic field gives rise to

a fluctuating current loop. In the electric case, the applied test electric field caused two equal and opposite charges to separate some distance. Upon application of the magnetic field, the charges that are separated start rotating about their center of mass. Just as the electric field will be replaced by the magnetic field, the perturbed separation will be replaced by the perturbed angular momentum. In this fashion, we can calculate a response function for the magnetic case.

7.4 ELECTRICAL CONDUCTIVITY

It is also possible to obtain a response function for how well a material conducts electricity. Ohm's law that one typically learns at school is the first term in the series of Equation (4.4), that is

$$I = V / R. \tag{7.92}$$

This is a coarse grained phenomenological model. We can also, however, turn this into a continuum version. The current needs to be replaced by the current density $I \rightarrow \mathbf{J}^a$, and the potential difference turns into the the electric field $V \rightarrow \mathbf{E}^a$. This should not be too alarming because the original form of Ohm's law had lost all the details about the geometry and microscopic behavior. Both of the these are vector quantities so in general they will be related by a two index tensor

$$\mathbf{J}^a(t,x) = \int dt' \sigma_b^a(t-t',x)\mathbf{E}^b(t',x), \tag{7.93}$$

where σ_b^a is defined as the conductivity tensor. This is yet another constitutive relation that is necessary to specify when solving the full field equations in certain circumstances. It is a response function and so exactly parallels the structure encountered earlier for the polarization. Since in a conductor the charge is not bound, a simple alternative to Equation (7.86) is to drop the binding potential term, whence

$$m\langle \delta\ddot{x}^a(t)\rangle + \gamma\langle \delta\dot{x}^a(t)\rangle = \hat{O}\langle \delta x^a(t)\rangle = q\delta\mathbf{E}^a(t). \tag{7.94}$$

If we perform the same substitution for the electric field and the displacement vector, the above in frequency space becomes

$$\delta P^a(\omega) = -\frac{q^2}{m\omega^2 - i\gamma\omega}\delta \mathbf{E}^a(\omega), \tag{7.95}$$

$$\chi_b^a(\omega) = -\frac{q^2 N\delta_b^a}{m\omega^2 - i\gamma\omega}, \tag{7.96}$$

which is known as the *Drude model*, and for frequencies that are so large that damping is negligible

$$\delta P^a(\omega) = -\frac{q^2}{m\omega^2}\delta \mathbf{E}^a(\omega), \tag{7.97}$$

$$\chi_b^a(\omega) = -\frac{q^2 N\delta_b^a}{m\omega^2}, \tag{7.98}$$

which is known as the Plasma model.

Paul Karl Ludwig Drude (1863–1906), made advances in connecting optical phenomena with electromagnetism and the thermal properties of solids.

We can relate the conductivity to the susceptibility in the following way (to simplify the notation we will take them to be scalars and promote them to tensors at the end). Firstly, write down Equation (7.39),

$$\nabla \wedge \mathbf{H}(x,\omega) - i\omega \mathbf{D}(x,\omega) = \mu_0 \mathbf{J}(x,\omega). \tag{7.99}$$

Now substitute in the the frequency space susceptibility and conductivity

$$\nabla \wedge \mathbf{H}(x,\omega) - i\omega(1+\chi)\mathbf{E}(x,\omega) = \mu_0 \sigma \mathbf{E}(x,\omega)$$

$$\nabla \wedge \mathbf{H}(x,\omega) = i\omega(1+\chi-\frac{i\sigma}{\omega})\mathbf{E}(x,\omega). \quad (7.100)$$

If we consider setting $\chi = 0$, the system then has some conductivity. Now set $\sigma = 0$ and substitute the Drude model in for the susceptibility. The imaginary part again has appeared to give us a conductivity. Therefore, we can calculate the conductivity from susceptibility by

$$\sigma^{ab}(\omega) = -i\omega\chi^{ab}(\omega). \quad (7.101)$$

What one can see then is that for $\omega \neq 0$, there is no real difference between a dielectric and a conductor, because ultimately they are both describing the movement of charge subjected to binding potentials.

7.5 FIELD EQUATIONS WITH GENERAL MATTER

Given that we have simple models for macroscopic matter in terms their underlying micro structure, it is instructive to make the corresponding substitutions into the field equations. A first obvious point is that the field equations in general are no longer local in time, since the response function is non-local in time (see Equation (7.82)). The assumption with this type of interaction is that the field arrives where the dipole (or some microstructure) is located, it then disturbs the system, and finally the system tries to return to its equilibrium state, thereby giving up the energy just acquired. This interaction must take place over a a certain amount of time that allows this disturbance and then relaxation to take place. What this means for the field equations is that it is simpler to work with the frequency components of the fields and the matter; if it is necessary, one can Fourier transform to the time difference coordinates at the end.

Let us consider then a macroscopic medium with a permittivity $\epsilon_b^a(\omega,x)$ a permeability $\mu_b^a(\omega,x)$ and a conductivity $\sigma_b^a(\omega,x)$. Then the constitutive relations are given by

$$\mathbf{D}^a(\omega,x) = \epsilon^a_b(\omega,x)\mathbf{E}^b(\omega,x), \tag{7.102}$$

$$\mathbf{H}^a(\omega,x) = \mu^a_b(\omega,x)\mathbf{B}^b(\omega,x), \tag{7.103}$$

$$\mathbf{J}^a(\omega,x) = \sigma^a_b(\omega,x)\mathbf{E}^b(\omega,x). \tag{7.104}$$

The field equations in frequency space are

$$\nabla\cdot\mathbf{D}(\omega,x) = \rho(\omega,x), \tag{7.105}$$

$$\nabla\cdot\mathbf{B}(\omega,x) = 0, \tag{7.106}$$

$$\nabla\wedge\mathbf{E}(\omega,x) + i\omega\mathbf{B}(\omega,x) = 0, \tag{7.107}$$

$$\nabla\wedge\mathbf{H}(\omega,x) - i\omega\mathbf{D}(\omega,x) = \mu_0\mathbf{J}(\omega,x). \tag{7.108}$$

To simplify, we multiply Equation (7.107) by the inverse of the permeability tensor and take the curl, so that

$$\nabla\wedge\mu^{-1}\nabla\wedge\mathbf{E}(\omega,x) + i\omega\nabla\wedge\mathbf{H}(\omega,x) = 0. \tag{7.109}$$

We now use Equation (7.108) to simplify the second term

$$\nabla\wedge\mu^{-1}\nabla\wedge\mathbf{E}(\omega,x) - \omega^2\mathbf{D}(\omega,x) + i\omega\mu_0\mathbf{J}(\omega,x) = 0. \tag{7.110}$$

Finally, using the constitutive relations Equations (7.102) and (7.104) one obtains

$$\nabla\wedge\mu^{-1}\nabla\wedge\mathbf{E}(\omega,x) - \omega^2\epsilon(\omega,x)\mathbf{E}(\omega,x) + i\omega\mu_0\sigma(\omega,x)\mathbf{E}(\omega,x) = 0, \tag{7.111}$$

$$\nabla\wedge\mu^{-1}\nabla\wedge\mathbf{E}(\omega,x) - (\omega^2\epsilon(\omega,x) - i\omega\mu_0\sigma(\omega,x))\mathbf{E}(\omega,x) = 0. \tag{7.112}$$

This is the most general vector Helmholtz type equation that one is likely to encounter. There is a similar equation for the magnetic field $\mathbf{H}^a$.

(Exercise: What is the corresponding vector Helmholtz equation that $\mathbf{H}^a$ must satisfy?)

It is worth commenting on some aspects of the form of these equations. The first point is due to the tensorial nature of material properties, in general wave propagation will be different in all directions. Next it will also be inhomogeneous as the tensors vary in position. They will depend on the frequency of the fields and due to the real and imaginary parts we will find both wave like solutions combined with growing or decaying modes. These will be fixed by the particular boundary conditions imposed.

7.6 SUMMARY

This chapter has focused on how real matter interacts with the electromagnetic field. Maxwell's field equations have been generalized to material media and the basic conditions have been laid down that are necessary for the fields to obey. Response functions have been derived that simple material models give, as well as the idea of small structures (multipoles) that cause the bare fields to become dressed. From here one can start to consider the many applications electromagnetism has to everything around us. Discussions of real matter can be found in [8] and [7].

CHAPTER 8

CONCLUSIONS AND OMISSIONS

Alas, we have reached journey's end. This book has been intended for the reader to gain some insight into the foundations of electromagnetism and field theory. In this sense it is a complimentary textbook and should hopefully serve to throw up questions for the reader to further pursue.

In any such work (and on a finite and short time scale) it is not possible to be complete. I have not covered many topics which are also necessary to have a full understanding (whatever that might mean). Listed now are a host of topics that may turn out to be a second volume to this book, or can be pursued by the reader at their leisure. Most of the topics I have not covered are either firmly in the applications arena, or else sufficiently advanced that I thought it would be an unnecessary detour from the objective of the book.

The radiation of moving charges, such as found in antennae or in accelerating beams of charges has been omitted. This is an important topic and the reader can do no better than consulting [7] to make a start here. In particular the Liénard-Wiechert potentials would be an appropriate place to start together with simple dipole radiation.

Special relativity and the formulation of electromagnetism as a relativistic field theory has been completely omitted. While this is a topic necessary to understand advanced field theory (for example, to get to grips with modern quantum field theories), I felt it would be

too distracting to enter in the current volume. In particular, one must be quite comfortable with Lagrangian and Hamiltonian mechanics to appreciate its full significance, and these topics in themselves can quite happily fill a textbook. It is also the case here that one is treading a more mathematical path. For example the nature of how fields interact with atomic matter can be quite mathematically technical [16] and one might also then think about group theory issues and the spinor formulation of electrodynamics [3]. These are indeed interesting advanced topics, but since they have a less well established measurement theory basis, I have not pushed at these doors.

A final couple of topics worth mentioning and worthy of further discussion are the scattering of electromagnetic waves and diffraction, which can be viewed from a unified rather than separate approach. These are certainly basic topics and correlate strongly with the philosophy here of a measurement centric approach, but to do them full justice would require enough material to fatten out a second book. The interested reader can consult [8] or [13] to make a start on this topic.

This book has tried to discuss electromagnetism from the point of view of not taking anything on faith. Coupled together with this has been the constructive bottom up approach to developing the field theory concepts that give the governing dynamics of the system. As a closing set of thoughts, one might just say that it is not a closed book. While electrodynamics is a well developed and tested theory, there do occur conceptual problems as well that do not have universally agreed upon solutions. Radiation reaction, self force, and pre-acceleration are some of the difficulties one encounters when charges interact with the fields they themselves produce. In a similar spirit, it is not always clear which stress tensor to use when considering interactions within materials themselves (for example the stress tensor we have defined or the Minkowski version). These issues indicate that even classically there are consistency issues that may need further thought perhaps with the help of some quantum input. This is a good task for the interested reader to further research and a good place to end.

EXERCISES AND SOLUTIONS

EXERCISES

Exercise 1

Take an average party balloon that is fully inflated and charge it up by rubbing it on your hair. What is a typical voltage that the balloon acquires due to this electrostatic charging process?

Exercise 2

Consider the following electrometer configuration. A disc has an area of 0.02m^2 that is separated from a fixed plate by distance 0.2mm. The potential difference between the two plates is held at 200V. What is the force between the two plates and how does this depend on the potential difference? how sensitive is this to the voltage across the plates? Suppose the upper disc is attached to a spring. If in equilibrium the separation of the two plates is d (when the potential difference is V) or a (when the potential difference is zero), deduce a criterion for stable equilibrium for the separation d configuration.

Exercise 3

Suppose we have four very long metallic plates that form a waveguide type structure, that is, two long strips of a certain width and parallel to one another, and another two long strips of another width and also parallel two one another. These are then glued together to form a rectangular structure that is extended in one direction. Suppose now we choose to hold the different plates at different electrostatic potentials. Two opposite faces are held at zero potential, while

the other two are held at a constant value V. What is the potential inside the wave-guide? The wave-guide has a length L that is much larger than the separations of the faces.

Exercise 4

Consider one of the disk plates that form a parallel plate disk capacitor. Suppose it has a static charge density on it that varies as a quadratic power of the radial coordinate. What is the potential and the electric field of this charge distribution directly above the center of the disk? What does it look like far away and close to the disk?

Exercise 5

Find an expression for the magnetic field of an elliptical current loop at a general point with respect to its center due to a steady current, in the limit that the point is far away from the ellipse. Do this by considering the vector potential first from which the magnetic field can be derived. Evaluate the corresponding expression when the ellipse is circular and nearly circular.

Exercise 6

Consider the gauge transformation that relates the Coloumb gauge to the Lorentz gauge. What equation must the corresponding gauge function satisfy? What does this look like in reciprocal space?

Exercise 7

Consider firstly the Green function that solves the scalar Helmholtz equation. In terms of this, what is the Green tensor that solves the vector Helmholtz equation? What transverse condition does it satisfy? Find expressions for the Green tensor in both wave vector space and position space. As the wavelength of the radiation is varied, which terms in the Green tensor are the most important?

Exercise 8

A perfectly absorbing sheet of metal has 200W of light incident on it normally for 10s. What is the total linear momentum transferred to the sheet? If the magnitude of the Poynting vector at the top of Earth's atmosphere is $\approx 1.4\text{kWm}^{-2}$, what is the average radiation pressure on the above metal sheet? If we wanted to use this as a solar sail, what size of sheet is necessary to capture 1N of force?

Exercise 9

Consider the expression for the energy, momentum, and angular momentum of the electromagnetic field. Suppose into these we substitute plane wave solutions. What are the particle like properties featuring in these expressions?

Exercise 10

Suppose magnetic monopoles **did** exist. What form would Maxwell's equations take? Discuss the symmetry between the electric and magnetic charges. What form would the Lorentz force law now take? How would you reconcile using $\mathbf{B} = \nabla \wedge \mathbf{A}$?

SOLUTIONS

Solution 1

This is one of those orders of magnitude, guesstimate questions based partly on common experience. If this is not common knowledge, you can easily do this simple experiment yourself! To measure the charge on the balloon and so determine its voltage, we need another balloon that is similarly charged. We can then balance the corresponding repulsive force with another mechanical force. It is a simple matter to balance this against gravity. Suspend each balloon from a point by two pieces of thread that are the same length. Because the balloons have about the same amount of net charge of the same kind, they will repel. The angle α the thread will make with the vertical will be in the range $0 < \alpha < \pi/2$. One now just resolves forces at the center of each balloon to balance off gravity and the repulsive forces.

Putting in some numbers, a typical balloon has a mass $m \approx 5$g, while a suspension angle of $\alpha = 30°$ and separation of their centers of 50cm are also typical. The downward force is thus $\approx 5 \times 10^{-2}$N and therefore the horizontal force is $\approx 5 \sin(30) \approx 2.5 \times 10^{-2}$N. We can thus equate this to the repulsive Coulomb force

$$2.5\times10^{-2} = \frac{q^2}{4\pi\epsilon_0(0.5)^2}$$
$$q \approx 8\times10^{-6}\,\mathrm{C}.$$

Assuming now that the charge has been distributed uniformly over the sphere, all we need is the capacitance of a sphere to get the voltage - and we know this. The radius of the balloon is $R \approx 15\text{cm}$ so

$$\begin{aligned} V &= \frac{q}{4\pi\epsilon_0 R}, \\ &\approx 50000\text{V}. \end{aligned}$$

Solution 2

For this problem we can assume the parallel plates are well described by the infinite plane idealization since the separation s is much smaller than the square root of the discs area A. The potential difference and therefore the force across the plates, remembering the lower plate is fixed, is given by (from Chapter 3)

$$\begin{aligned} V &= \frac{\sigma A s}{2\epsilon_0}, \\ E &= \frac{\sigma A}{2\epsilon_0}, \\ F &= \frac{\sigma^2 A^2}{2\epsilon_0} \\ &= \frac{\epsilon_0 V^2 A^2}{2s^2}. \end{aligned}$$

Therefore putting in the numbers, one obtains for the force between the plates that $F \approx 1.77\text{N}$. to get the sensitivity at this voltage, we just need to differentiate with respect to the voltage, resulting in

$$\begin{aligned} \frac{dF}{dV} &= \frac{\epsilon_0 V A^2}{s^2} \\ &\approx 1.77 \times 10^{-5}\text{Nm}^{-1}. \end{aligned}$$

Since we have now attached a spring to the upper disc, the linear restoring force now needs to be included in the forces acting on the disc. Suppose the plates have some separation x, then

$$F = -\frac{\epsilon_0 V A^2}{x^2} + k(x - x_0),$$
$$\frac{dF}{dx} = \frac{2\epsilon_0 V A^2}{x^3} + k.$$

When $V = 0$, $x = a$ so therefore at the second equilibrium $x_0 = a$. At the other equilibrium one has

$$0 = -\frac{\epsilon_0 V A^2}{d^2} + k(x - a),$$
$$k = \frac{\epsilon_0 V A^2}{d^2(d - a)}.$$

For this position to be a stable equilibrium requires that $dF/dx > 0$ and so

$$\begin{aligned} \frac{2\epsilon_0 V A^2}{d^3} + k &> 0, \\ \frac{2\epsilon_0 V A^2}{d^3} + \frac{\epsilon_0 V A^2}{d^2(d - a)} &> 0, \\ a &< 3d/2. \end{aligned}$$

Solution 3

Obviously we need to work in Cartesian coordinates, so let the long direction be z. We define the faces at the potential V to be $x = \pm a$ and the zero potential faces to be $y = 0, b$. This is a boundary problem associated with solving Laplace's equation in Cartesian coordinates

$$\nabla^2 \phi = \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2} \right) \phi(x, y, z) = 0.$$

In the case that we regard L as being very long, we can drop the z-dependence such that the above reduces to the two-dimensional Laplace equation in the x, y coordinates. The separation of variables

for this equation leads to both exponential and sinusoidal solutions. The *boundary* conditions require that the potential vanishes in two planes which can only be fulfilled by the sine or cosine eigenfunction. Writing the solution as

$$\phi(x,y) = X(x)Y(y),$$

one has

$$X(x) = a_1 \sinh(kx) + a_2 \cosh(kx),$$
$$Y(y) = b_1 \sin(ky) + b_2 \cos(ky).$$

The boundary conditions tell us the following; X(x) must be symmetric in x since the potential is the same on the planes $x = \pm a$. This means $a_1 = 0$. Next, since the potential vanishes at $y = 0$, we must have $b_2 = 0$; and finally at $y = b$, the only way for the sine function to vanish there is if $kb = n\pi$ and therefore that

$$k = \frac{n\pi}{b},$$

where $n \in \mathbb{Z}$. So the solution looks so far like

$$\phi(x,y) = \sum_{n=1}^{\infty} c_n \cosh\left(\frac{n\pi x}{b}\right) \sin\left(\frac{n\pi y}{b}\right),$$

the sum being performed over all positive n since this is a linear equation and we must in general superpose all possible solutions. The negative integers do not give any new solutions which is why the n have been restricted to only the positive values. To determine the coefficients we use the last piece of information available, that we know the potential at $x = a$

$$V = \sum_{n=1}^{\infty} c_n \cosh\left(\frac{n\pi a}{b}\right) \sin\left(\frac{n\pi y}{b}\right),$$

This is just a standard Fourier series in y; one uses the orthogonality of the eigenfunctions to invert the above equation

$$\int_0^b dy V \sin\left(\frac{n\pi y}{b}\right) = (c_n/2)\cosh\left(\frac{n\pi a}{b}\right)(b/2).$$

$$c_n = \frac{2V}{b\cosh\left(\frac{n\pi a}{b}\right)}\left[\frac{b}{n\pi}\cos(n\pi) - \frac{b}{n\pi}\right],$$

$$c_n = \frac{2V}{\cosh\left(\frac{n\pi a}{b}\right)}\left[\frac{1}{n\pi}(-1)^n - \frac{1}{n\pi}\right].$$

So the coefficients are only nonzero for odd values of n.

Solution 4

The starting point for this problem is going to be Poisson's equation in its inverted form. We will calculate the potential first from which we can obtain the electric field by differentiation. The potential is given in terms of the charge density by

$$\phi(x) = \frac{1}{4\pi\epsilon_0}\int d^3x' \frac{\rho(x')}{|x - x'|}.$$

Working in cylindrical coordinates, let the radius of the disk be a and the charge density be $\rho(r,\theta,z) = \delta(z)\lambda r^2$. The potential then becomes

$$\phi(r,\theta,z) = \frac{1}{4\pi\epsilon_0}\int_0^{2\pi} d\theta' \int_0^a dr' r' \int_{-\infty}^{\infty} dz' \frac{\lambda r'^2 \delta(z')}{\sqrt{z^2 + r'^2}}$$

$$= \frac{\lambda}{2\epsilon_0}\int_0^a dr' \frac{r'^3}{\sqrt{z^2 + r'^2}}.$$

If we make the substitution $r'^2 = X$ the integral then takes a standard integral form that one can find in a book in integrals (or just do it yourself)

$$\phi(z) = \frac{\lambda}{4\epsilon_0}\int_{X(0)}^{X(a)} dX \frac{X}{\sqrt{z^2 + X}}.$$

This integrates to

$$\begin{aligned}\phi(z) &= \frac{\lambda}{4\epsilon_0}\left[-\frac{2(2z^2-X)}{3}\sqrt{z^2+X}\right]_{X(0)}^{X(a)} \\ &= \frac{\lambda}{4\epsilon_0}\left[-\frac{2(2z^2-a^2)}{3}\sqrt{z^2+a^2}+\frac{2(2z^2)}{3}\sqrt{z^2}\right] \\ &= \frac{\lambda}{4\epsilon_0}\left[\frac{4z^3}{3}-\frac{2(2z^2-a^2)}{3}\sqrt{z^2+a^2}\right].\end{aligned}$$

The electric field then has only a z-component and is

$$\begin{aligned}\mathbf{E}_z &= -\nabla_z\phi(z) \\ &= -\frac{\lambda}{4\epsilon_0}\left[4z^2-\frac{8z}{3}\sqrt{z^2+a^2}-\frac{2z(2z^2-a^2)}{3\sqrt{z^2+a^2}}\right].\end{aligned}$$

Finally we should consider the limits when $z << a$ and also $z >> a$. The potential in these limits reduces to

$$\begin{aligned}\phi(z) &\approx \frac{2a^3}{3}\left(\frac{\lambda}{4\epsilon_0}\right), \text{ when } z << a, \\ \phi(z) &= \left(\frac{\lambda}{4\epsilon_0}\right)\left(\frac{4z^3}{3}-\frac{2(2z^2-a^2)}{3}z\sqrt{1+a^2/z^2}\right) \\ &\approx \frac{a^4}{6z}\left(\frac{\lambda}{4\epsilon_0}\right), \text{ when } z >> a.\end{aligned}$$

From these approximations we see on axis that the electric field vanishes just above the disc, and falls off as $1/z^2$ far away from the disc; this is to be expected since at large distances away from the disc one cannot resolve any structure and it is only the total charge that is meaningful.

Solution 5

To evaluate the magnetic field we will need to use the Biot-Savart law since we have an expression for the current I. In this case it takes the line integral form

$$\mathbf{B}(\mathbf{x}) = \frac{\mu_0 I}{4\pi} \oint \frac{d\mathbf{l}' \wedge (\mathbf{x} - \mathbf{x}')}{|(\mathbf{x} - \mathbf{x}')|^3},$$

where $d\mathbf{l}$ is the line element of the ellipse and $(\mathbf{x} - \mathbf{x}')$ is the vector between the point $\mathbf{x}'$ where the line element is located and $\mathbf{x}$ where we are evaluating the field. It will clearly be necessary to use the equation of the ellipse to be able to perform the integration. Since we are dealing with a magnetostatic situation it is simpler to calculate the vector potential first and from this calculate the field (this way removes an initial wedge product). The vector potential is given by

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0 I}{4\pi} \oint \frac{d\mathbf{l}'}{|(\mathbf{x} - \mathbf{x}')|}.$$

So we can see here that since the vector potential only has a rotational component the field will be out of the plane. We work in spherical polar coordinates with the origin at the center of the ellipse, with the ellipse itself lying in the $\theta = \pi/2$ plane. To start with we will evaluate the field at the general off axis point $\mathbf{x} = (r, \theta, \phi)$ so that we can differentiate it. After this we will restrict it to on axis. The points on the ellipse are given by $\mathbf{x}' = (r', \pi/2, \phi')$ so that

$$|(\mathbf{x} - \mathbf{x}')| = \sqrt{r^2 + r'^2 - 2rr' \sin\theta \cos(\phi - \phi')}.$$

The more involved part is the length vector of the current element itself. It is the variable ϕ' that parametrizes the ellipse, but now r' is a function of ϕ' (for the circle r' is of course a constant). The position vector $\mathbf{x}'$ lies in the x, y plane of the equivalent Cartesian coordinates. In terms of this basis then the position vector is given by $\mathbf{x}' = (r' \cos\phi', r' \sin\phi')$. Therefore the line element in this basis is given by

$$\begin{aligned} d\mathbf{l} &= \frac{\partial \mathbf{x}'}{\partial \phi'} d\phi', \\ &= (-r' \sin\phi' d\phi' + (\partial_{\phi'} r') \cos\phi' d\phi', r' \cos\phi' d\phi' + (\partial_{\phi'} r') \sin\phi' d\phi'). \end{aligned}$$

From this one then has an expression for the vector potential in the Cartesian coordinate basis. One must be careful here to not

loose sight of the different coordinate systems and variables. The vector potential components then are

$$\mathbf{A}_x(r,\theta,\phi)=\frac{\mu_0 I}{4\pi}\oint d\phi'\frac{-r'\sin\phi'+(\partial_{\phi'}r')\cos\phi'}{\sqrt{r^2+r'^2-2rr'\sin\theta\cos(\phi-\phi')}},$$

$$\mathbf{A}_y(r,\theta,\phi)=\frac{\mu_0 I}{4\pi}\oint d\phi'\frac{r'\cos\phi'+(\partial_{\phi'}r')\sin\phi'}{\sqrt{r^2+r'^2-2rr'\sin\theta\cos(\phi-\phi')}}.$$

From these expressions the magnetic field can be calculated in full generality, even though we can evaluate them explicitly in certain limits. We will evaluate them in the two limits, $r >> r'$ and $r << r'$. Firstly for $r >> r'$

$$\mathbf{A}_x(r,\theta,\phi)=\frac{\mu_0 I}{4\pi r}\oint d\phi'[-r'\sin\phi'+(\partial_{\phi'}r')\cos\phi']$$
$$\times[1-r'^2/2r^2+(r'/r)\sin\theta\cos(\phi-\phi')],$$

$$\mathbf{A}_y(r,\theta,\phi)=\frac{\mu_0 I}{4\pi r}\oint d\phi'[r'\cos\phi'+(\partial_{\phi'}r')\sin\phi']$$
$$\times[1-r'^2/r^2-(r'/r)\sin\theta\cos(\phi-\phi')].$$

It is now time to include the equation of the ellipse to evaluate the integral. In standard two dimensional Cartesian coordinates the ellipse is parameterized by for constants a and b

$$x=a\cos\phi',$$
$$y=b\sin\phi',$$
$$\Rightarrow$$
$$r'^2=a^2\cos^2\phi'+b^2\sin^2\phi',$$
$$\frac{dr'}{d\phi'}=\frac{(a^2-b^2)\sin 2\phi'}{2r'}.$$

If the ellipse is nearly circular then $a=b+\epsilon$ where ϵ is small number that will serve as an expansion parameter. Therefore

$$r' \approx b + \epsilon \cos^2 \phi',$$
$$\frac{dr'}{d\phi'} \approx \epsilon \sin 2\phi'.$$

With these expressions we can now evaluate the gauge potential components

$$\begin{aligned}
\mathbf{A}_x(r,\theta,\phi) = \frac{\mu_0 I}{4\pi r} \oint d\phi' [&-b\sin\phi' - \epsilon \sin\phi' \cos^2\phi' + \epsilon \sin 2\phi' \cos\phi'][1 \\
&- b^2/2r^2 - b\epsilon \cos^2\phi'/r^2 - b\sin\theta\cos(\phi-\phi')/r \\
&- \epsilon \sin\theta \cos^3(\phi-\phi')/r], \\
\mathbf{A}_y(r,\theta,\phi) = \frac{\mu_0 I}{4\pi r} \oint d\phi' [&b\cos\phi' + \epsilon \cos^3\phi' + \epsilon \sin 2\phi' \sin\phi'][1 \\
&- b^2/2r^2 - b\epsilon \cos^2\phi'/r^2 - b\sin\theta\cos(\phi-\phi')/r \\
&- \epsilon \sin\theta \cos^3(\phi-\phi')/r] .
\end{aligned}$$

By inspection we can simplify these by remembering the rules for integrating sine and cosine functions. With this in mind they become

$$\begin{aligned}
\mathbf{A}_x(r,\theta,\phi) = \frac{\mu_0 I}{4\pi r^2} \oint d\phi' [&b^2 \sin\theta \sin\phi' \cos(\phi-\phi') \\
&+ b\epsilon \sin\theta \sin\phi' \cos^3(\phi-\phi') \\
&+ b\epsilon \sin\theta \sin\phi' \cos^2\phi' \cos(\phi-\phi') \\
&- b\epsilon \sin\theta \sin 2\phi' \cos\phi' \cos(\phi-\phi')], \\
\mathbf{A}_y(r,\theta,\phi) = \frac{\mu_0 I}{4\pi r^2} \oint d\phi' [&-b^2 \sin\theta \cos\phi' \cos(\phi-\phi') \\
&- b\epsilon \sin\theta \cos\phi' \cos^3(\phi-\phi') \\
&- b\epsilon \sin\theta \cos^3\phi' \cos(\phi-\phi') \\
&- b\epsilon \sin\theta \sin 2\phi' \sin\phi' \cos(\phi-\phi')],
\end{aligned}$$

where ϵ^2 terms have been dropped. The integrals can now be evaluated straightforwardly since they take the form of products of sine

and cosine functions once the double angle has been expanded. With this in mind, they reduce to

$$\begin{aligned}\mathbf{A}_x(r,\theta,\phi) &= \frac{\mu_0 I}{2r^2}[(b^2\sin\theta\sin\phi)/2 + (3b\epsilon\sin\theta\sin\phi)/8 \\ &\quad -(b\epsilon\sin\theta\sin\phi)/8] \\ &= \frac{\mu_0 I\sin\theta\sin\phi}{2r^2}[b^2/2 + b\epsilon/4], \\ \mathbf{A}_y(r,\theta,\phi) &= \frac{\mu_0 I}{2r^2}[-(b^2\sin\theta\cos\phi)/2 - 3(b\epsilon\sin\theta\cos\phi)/8) \\ &\quad -5(b\epsilon\sin\theta\cos\phi)/8], \\ &= \frac{\mu_0 I\sin\theta\cos\phi}{2r^2}[-b^2/2 - b\epsilon].\end{aligned}$$

To finally obtain the magnetic field in the z direction we need to partially differentiate these expressions with respect x and y. So we write the above expressions back into Cartesian coordinates

$$\begin{aligned}\mathbf{A}_x &= \frac{\mu_0 Iy}{2(x^2+y^2+z^2)^{3/2}}[b^2/2 + b\epsilon/4], \\ \mathbf{A}_y &= \frac{\mu_0 Ix}{2(x^2+y^2+z^2)^{3/2}}[-b^2/2 - b\epsilon]\ .\end{aligned}$$

The magnetic field in the z direction is just $\mathbf{B}_z = \partial_x \mathbf{A}_y - \partial_y \mathbf{A}_x$ so

$$\mathbf{B}_z = -\frac{\mu_0 I}{2r^3}[b^2 + 5b\epsilon/4] - \frac{3\mu_0 I}{2r^5}[b^2(x^2+y^2)/2 + b\epsilon(x^2+y^2/4)].$$

The difference then from the circular current loop will be seen at points off the z axis.

Solution 6

Let us assume that we know the potentials in the Coulomb gauge, given by $(\phi, \mathbf{A})$, and so that we will perform a gauge transformation on these to the Lorentz gauge with potentials $(\phi', \mathbf{A}')$. The Lorentz condition is

$$\frac{1}{c^2}\partial_t\phi' - \nabla\cdot\mathbf{A}' = 0.$$

Since these potentials are related to the Coulomb gauge potentials by a gauge function $\Omega(t,x)$, we have

$$\frac{1}{c^2}\partial_t(\phi+\partial_t\Omega) - \nabla\cdot(\mathbf{A}+\nabla\Omega) = 0,$$

$$\frac{1}{c^2}\partial_t\phi + \frac{1}{c^2}\partial_t^2\Omega - \nabla^2\Omega = 0,$$

where the Coulomb gauge condition $\nabla\cdot\mathbf{A} = 0$ has been used. If we now Fourier transform both the gauge function and the scalar potential then an expression in reciprocal space can also be obtained, so that

$$+\frac{1}{c^2}\partial_t^2\Omega(t,x) - \nabla^2\Omega(t,x) = -\frac{1}{c^2}\partial_t\phi(t,x),$$

$$+\frac{1}{c^2}\partial_t^2\Omega(t,\mathbf{k}) + \mathbf{k}^2\Omega(t,\mathbf{k}) = \partial_t\phi(t,\mathbf{k}).$$

The gauge function then depends explicitly on the charge content and will be proportional to the velocity of these charges.

Solution 7

The scalar Green function $\Delta(x,y)$ that satisfies the scalar Helmholtz equation is given by

$$(\nabla^2 + k^2)\Delta(x,y) = \delta^3(x-y).$$

The tensor Green function $\Delta_{ab}(x,y)$ that satisfies the vector Helmholtz equation is similarly given by

$$((\nabla\wedge\nabla)^c_a - k^2\delta^c_a)\Delta_{cb}(x,y) = -\delta_{ab}\delta^3(x-y).$$

The trick here is to try and write the Green tensor in such a way that when it is substituted into the vector Helmholtz equation, it reduces to the scalar equation. To that end we try the following decomposition

$$\Delta_{ab}(x, y) = \alpha \nabla_a \nabla_b \Delta(x, y) + \beta \delta_{ab} \Delta(x, y),$$

where α and β are constants to be determined. The double wedge of the derivative operator can be written as $(\nabla \wedge \nabla)_{ac} = \nabla_a \nabla_c - \delta_{ac} \nabla^2$. This leads directly to the the two conditions

$$\alpha = \beta / k^2, \beta = 1.$$

Therefore the Green tensor is

$$\Delta_{ab}(x, y) = \frac{1}{k^2} \nabla_a \nabla_b \Delta(x, y) + \delta_{ab} \Delta(x, y).$$

It is a simple matter to see that this satisfies a transverse constraint. If we act on the Green tensor with the derivative operator one finds

$$\begin{aligned} \nabla^b \Delta_{ab}(x, y) &= \frac{1}{k^2} \nabla_a \nabla^2 \Delta(x, y) + \nabla_a \Delta(x, y) \\ &= 0. \end{aligned}$$

This result has its root in Gauss's law and is the basic statement that for propagating fields there is a local plane of fluctuations that are normal to the direction of propagation.

It is a simple matter to obtain an expression for the Green function in wave vector space; simply perform a Fourier transform

$$(\nabla^2 + k^2)\Delta(x, y) = \delta^3(x - y)$$

$$\int \frac{d^3p}{(2\pi)^3} e^{ip\cdot(x-y)} (-p^2 + k^2)\Delta(p) = \int \frac{d^3p}{(2\pi)^3} e^{ip\cdot(x-y)}$$

$$\Rightarrow$$

$$\Delta(x, y) = \int \frac{d^3p}{(2\pi)^3} \frac{e^{ip\cdot(x-y)}}{-p^2 + k^2}.$$

Straightforward differentiation now gives Green tensor for the vector Helmholtz equation

$$\Delta_{ab}(x,\, y) = \frac{1}{k^2}\nabla_a\nabla_b\Delta(x,\, y) + \delta_{ab}\Delta(x,\, y)$$
$$= \int \frac{d^3p}{(2\pi)^3}\frac{e^{ip\cdot(x-y)}}{-p^2+k^2}(-p_a p_b)/k^2 + \delta_{ab}).$$

We can also calculate what the Green function looks like in position space by evaluating the above integrals. To do this requires performing an integral in the complex plane. Since this topic lies outside the main drive of this book, we will assume the reader knows how to perform such integrals (i.e. method of residues etc). For the scalar Green function a $\pm i\epsilon$ is added to the denominator that gives a prescription of how to enclose the poles in the complex plane. Having done this the integral can then be evaluated by finding the residues of the integrand. This results in the two expressions

$$\Delta^{\pm}(x,\, y) \;=\; \frac{e^{\pm ik|x-y|}}{4\pi\,|\,x-y\,|}.$$

The + sign prescription gives the retarded solution (where effect follows cause) and the − sign gives the advanced solution (where cause follows effect), amounting to either incoming or outgoing waves. From this we can simply deduce the position space version of the Green tensor, which is for the retarded solution

$$\Delta^{+}_{ab}(x,\, y) = (\frac{1}{k^2}\nabla_a\nabla_b + \delta_{ab})\frac{e^{ik|x-y|}}{4\pi\,|\,x-y\,|}$$
$$= \frac{e^{ik|x-y|}}{4\pi\,|\,x-y\,|}\left(\delta_{ab} - \frac{(x-y)_a(x-y)_b}{|\,x-y\,|^2}\right)$$
$$+ \frac{e^{ik|x-y|}}{4\pi\,|\,x-y\,|}\frac{i}{k\,|\,x-y\,|}\left(\delta_{ab} - 3\frac{(x-y)_a(x-y)_b}{|\,x-y\,|^2}\right)$$
$$+ \frac{e^{ik|x-y|}}{4\pi\,|\,x-y\,|}\frac{1}{(k\,|\,x-y\,|)^2}\left(-\delta_{ab} + 3\frac{(x-y)_a(x-y)_b}{|\,x-y\,|^2}\right)$$

The expression for the Green tensor has therefore organized itself into increasing powers in the denominator of the dimensionless variable $k\,|\,x-y\,|$. The $(k\,|\,x-y\,|)^2$ is called the near field term,

since this dominates when for radiation of wavelength λ one has $|x-y| << \lambda$. The $(k|x-y|)$ term is called the intermediate field when $|x-y| \sim \lambda$ and therefore dominates the expression. Finally the remaining term is called the far field piece which is the most important when $|x-y| >> \lambda$. Interestingly, the intermediate field is exactly $\pi/2$ out of phase with the far field and near field Green tensor contributions.

Solution 8

Assuming that the incident radiation is propagating in the z direction and that the metal sheet spans the x, y plane, the conservation of momentum equation gives that the radiation pressure ρ in the z direction is just the stress tensor component T_{zz}. Using the form of the stress tensor in terms of the fields, this is just $T_{zz} = -(\epsilon_0 \mathbf{E}^2 + \mathbf{B}^2/\mu_0)/2$. In an amount of time Δt, the energy ΔE incident on the sheet is therefore $\Delta E = \int d^2 \sum |T_{zz}| c\Delta t$ which is numerically just the power times the time, $\Delta E = 2000\text{J}$. On the other side, multiplying the radiation pressure by Δt and integrating over the sheet gives us the total change in linear momentum ΔP which is therefore $\Delta P = \Delta E/c \approx 6.6 \times 10^{-6} \text{kgms}^{-1}$.

For a plane wave we take the magnitudes of the electric and magnetic field to be related by $|\mathbf{E}| = c|\mathbf{B}|$ (this can be verified by just making a plane wave substitution into Maxwell's field equations) which allows us to take the radiation pressure to be $\rho = |\mathbf{S}|/c$. So numerically we find that on the metal sheet $\rho \approx 4.7 \times 10^{-6} \text{Nm}^{-2}$. To produce 1N of force on a solar sail then would require an area of $\approx 2.1 \times 10^6 \text{m}^2$. If this was a square sail then the length of a side would be approximately 1km long.

Solution 9

By consider plane wave solutions of the field equations and summing over their corresponding eigenvalues, we are attempting to generate a solution that is typically a fluctuation. To be definite, we will look at field fluctuations in a matter free region of space and start from the scalar and vector potentials. Let us fix the gauge to be the Coloumb gauge so that $\phi = 0$ and $\nabla \cdot \mathbf{A} = 0$. The mode expansion in plane waves of the gauge field then takes the form

$$\mathbf{A}_b(t, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} [\mathbf{a}_b(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} + \mathbf{a}_b^*(\mathbf{k}) e^{i(-\mathbf{k}\cdot\mathbf{x} + \omega t)}].$$

This is easily verified as a solution to the wave equation with the constraint that $\omega^2 = c^2\mathbf{k}\cdot\mathbf{k}$. The gauge condition requires that $\mathbf{k}\cdot\mathbf{a} = 0$. To that end we can introduce two polarization vectors that characterise the modes of vibration. Write them as $\mathbf{e}^{(\lambda)}(\mathbf{k})$, where $\lambda = 1, 2$ are the two polarization state labels. Using these the vector Fourier modes $\mathbf{a}_b(\mathbf{k})$ are then replaced with a scalar quantity that has the polarization mode attached, that is $a^{(\lambda)}(\mathbf{k})$. With these two quantities the gauge potential now looks like

$$\mathbf{A}(t,\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\sum_{\lambda=1}^{2}\mathbf{e}^{(\lambda)}(\mathbf{k})\,[a^{(\lambda)}(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} + (a^{(\lambda)})^*(\mathbf{k})e^{i(-\mathbf{k}\cdot\mathbf{x}+\omega t)}].$$

Obviously, the polarization vectors will be required to satisfy certain constraints that describe the basic choice of the Coloumb gauge. They are

$$\begin{aligned}\mathbf{e}^{(\lambda)}(\mathbf{k})\cdot\mathbf{k} &= 0,\\ \mathbf{e}^{(\lambda)}(\mathbf{k})\cdot\mathbf{e}^{(\lambda')}(\mathbf{k}) &= \delta^{\lambda\lambda'},\end{aligned}$$

the second condition just ensuring that the polarization vectors form a local set of orthonormal basis vectors that, from the first condition, are transverse to the wave vector. This is then the basic field and its decomposition. Let us now turn to evaluating the momentum of the field given by

$$\mathbf{P} = \epsilon_0\int d^3x\,\mathbf{E}\wedge\mathbf{B}.$$

The fields $\mathbf{E}$ and $\mathbf{B}$ can be written in terms of the mode decomposition as

$$\begin{aligned}\mathbf{E}(t,\mathbf{x}) &= -\nabla\phi + \partial_t\mathbf{A}\\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^3}(-i\omega)\sum_{\lambda=1}^{2}\mathbf{e}^{(\lambda)}(\mathbf{k})[a^{(\lambda)}(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\\ &\quad -(a^{(\lambda)})^*(\mathbf{k})e^{i(-\mathbf{k}\cdot\mathbf{x}+\omega t)}],\\ \mathbf{B}(t,\mathbf{x}) &= \nabla\wedge\mathbf{A}\\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^3}(-i)\sum_{\lambda=1}^{2}\mathbf{k}\wedge\mathbf{e}^{(\lambda)}(\mathbf{k})[a^{(\lambda)}(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\\ &\quad -(a^{(\lambda)})^*(\mathbf{k})e^{i(-\mathbf{k}\cdot\mathbf{x}+\omega t)}].\end{aligned}$$

If we now substitute these into the expression for the momentum and perform the integration over the spatial coordinate, we will obtain delta functions in the wave vectors. As an example, let us evaluate the first term in the electric field with the second term in the magnetic field,

$$\mathbf{P}(1,2) = \epsilon_0 \int d^3x \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{d^3\mathbf{k}'}{(2\pi)^3} (-\omega) \sum_{\lambda=1}^{2} \sum_{\lambda'=1}^{2} \mathbf{e}^{(\lambda)}(\mathbf{k}) \wedge \mathbf{k}' \wedge \mathbf{e}^{(\lambda')}(\mathbf{k}') \times a^{(\lambda)}(\mathbf{k})(a^{(\lambda)'})^*(\mathbf{k}') e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}-(\omega-\omega')t}.$$

The argument of momentum in the above $(1, 2)$ denotes the cross term specification. Performing the integral over space gives the delta function $\delta^3(\mathbf{k}-\mathbf{k}')$. Then the integral over $\mathbf{k}'$ can be done simply, also giving $\omega' = \omega$. The two wedge product then simplifies to

$$\begin{aligned} \mathbf{e}^{(\lambda)}(\mathbf{k}) \wedge \mathbf{k} \wedge \mathbf{e}^{(\lambda')}(\mathbf{k}) &= (\mathbf{e}^{(\lambda)}(\mathbf{k}) \cdot \mathbf{k})\mathbf{e}^{(\lambda')}(\mathbf{k}) - \mathbf{k}(\mathbf{e}^{(\lambda)}(\mathbf{k}) \cdot \mathbf{e}^{(\lambda')}(\mathbf{k})) \\ &= -\mathbf{k}\delta^{\lambda\lambda'}. \end{aligned}$$

The above then simplifies to

$$\mathbf{P}(1,2) = \epsilon_0 \sum_{\lambda=1}^{2} \int \frac{d^3\mathbf{k}}{(2\pi)^3} a^{(\lambda)}(\mathbf{k})(a^{(\lambda)}(\mathbf{k}))^*(\omega\mathbf{k}).$$

We obtain a similar result for the other cross term involving an $a^{(\lambda)}(\mathbf{k})$ and its complex conjugate, while the other two terms integrate to zero. This can be seen as follows for the first term in the electric and magnetic field cross product

$$\begin{aligned} \mathbf{P}(1,1) &= \epsilon_0 \sum_{\lambda=1}^{2} \int \frac{d^3\mathbf{k}}{(2\pi)^3} a^{(\lambda)}(\mathbf{k})(a^{(\lambda)}(\mathbf{k}))(\omega\mathbf{k}) \\ &= 0, \end{aligned}$$

since the integrand is an odd function and therefore integrates to zero. The final result were are left with then is

$$\mathbf{P} = \epsilon_0 \sum_{\lambda=1}^{2} \int \frac{d^3\mathbf{k}}{(2\pi)^3} (2\omega) a^{(\lambda)}(\mathbf{k})(a^{(\lambda)}(\mathbf{k}))^*\mathbf{k}.$$

The energy of the field E is given by

$$E = \int d^3x \frac{1}{2}\left(\epsilon_0 \mathbf{E}_a \mathbf{E}^a + \frac{1}{\mu_0}\mathbf{B}_a \mathbf{B}^a\right).$$

With similar manipulations to before this reduces to

$$E = \epsilon_0 \sum_{\lambda=1}^{2} \int \frac{d^3\mathbf{k}}{(2\pi)^3}(2\omega) a^{(\lambda)}(\mathbf{k})(a^{(\lambda)}(\mathbf{k}))^*(\omega).$$

The angular momentum of the field is a little bit more tricky because we have the explicit appearance of the position vector in the integral, that is

$$\mathbf{L} = \epsilon_0 \int d^3x \mathbf{x} \wedge \mathbf{E} \wedge \mathbf{B}.$$

Again we substitute in the field decompositions as before, but in addition, to generate the position vector we need to partially differentiate the exponential by the wave vector, i.e.

$$\mathbf{x} e^{i(-\mathbf{k}\cdot\mathbf{x}+\omega t)} = i\frac{\partial}{\partial \mathbf{k}} e^{i(-\mathbf{k}\cdot\mathbf{x}+\omega t)}.$$

Using this in the simplification results in the field angular momentum

$$\mathbf{L} = \epsilon_0 \sum_{\lambda=1}^{2} \int \frac{d^3\mathbf{k}}{(2\pi)^3}(2\omega) i\frac{\partial}{\partial \mathbf{k}}(a^{(\lambda)}(\mathbf{k})(a^{(\lambda)}(\mathbf{k}))^*) \wedge \mathbf{k}.$$

Looking at the three results we see that the common factor 2ω acts as an overall normalization for the modes and the density of these modes are given by the bilinear $a^{(\lambda)}(\mathbf{k})a^{(\lambda)}(\mathbf{k})^*$. The macroscopic physical observables $(\mathbf{P}, E, \mathbf{L})$ map to the mode parameters $(\mathbf{k}, \omega, -\mathbf{k} \wedge i\partial/\partial\mathbf{k})$ which up to a constant factor are the mode values of the momentum, energy and angular momentum. In fact in the quantum theory this constant factor is Planck's constant and these modes then describe a particle like excitation.

Solution 10

If magnetic monopoles did exist we would have to ascribe to them a new magnetic charge g or magnetic charge density $\rho_M(x)$. This would require **Maxwell II** to take the new form

$$\nabla \cdot \mathbf{B}(x) = \rho_M(x).$$

In addition to this, assuming that it can move in much the same way as an electric charge, one now has a magnetic current. It is therefore necessary to modify Faraday's law (**Maxwell III**) to include this so that

$$\nabla \wedge \mathbf{E}(x) + \partial_t \mathbf{B} = -\mathbf{J}_M(x).$$

For the other two Maxwell equations it is sensible to label the electric charge and current as $\rho_E(x)$ and $\mathbf{J}_E$. If we now form a complex charge given by $\rho + i\rho_M$ and a complex current $\mathbf{J}_E + i\mathbf{J}_M$ then there exists a new symmetry of the field equations. Starting with the charge the new version of **Maxwell I** reads

$$\nabla \cdot (\mathbf{E}(x) + i\mathbf{B}(x)) = \rho_E(x) + i\rho_M(x).$$

It is simple enough to see now that the transformations (in fact a $U(1)$ Abelian group rotation)

$$\rho_E(x) + i\rho_M(x) \to e^{i\theta}(\rho_E(x) + i\rho_M(x)),$$
$$\mathbf{E}(x) + i\mathbf{B}(x) \to e^{i\theta}(\rho_E(x) + i\rho_M(x)),$$

leave the above Gauss law invariant and is therefore a symmetry of this equation. We also need to check the remaining complex field equation which is

$$\nabla \wedge (\mathbf{E}(x) + i\mathbf{B}(x)) - i\,\partial_t(\mathbf{E} + i\mathbf{B}) = i(\mathbf{J}_E(x) + i\mathbf{J}_M(x)).$$

With $(\mathbf{J}_E(x) + i\mathbf{J}_M(x)) \to e^{i\theta}(\mathbf{J}_E(x) + i\mathbf{J}_M(x))$ as before, the generalized field equations are invariant under this symmetry transformation.This is perhaps the simplest example of what is known as a duality relation, and physically it is describing how we can interchange electric and magnetic charges without any effect on the physics - what you mean by electric or magnetic charge depends on how you look at it.

For electric charges the Lorentz force law stays the same, but we now have in addition a magnetic Lorentz force law for the magnetic charge g which is

$$\mathbf{F}^E = q\mathbf{E} + q\mathbf{v} \wedge \mathbf{B},$$
$$\mathbf{F}^M = g\mathbf{B} - g\mathbf{v} \wedge \mathbf{E}.$$

Note at this point there is no duality symmetry here; if there are two types of charges, you should be able to distinguish them (by definition) by applying electric and magnetic fields. However, if a particle has both types of charge (known as a dyon) then the above becomes

$$\mathbf{F} = q\mathbf{E} + g\mathbf{B} + \mathbf{v} \wedge (q\mathbf{B} - g\mathbf{E}),$$

which is invariant under the symmetry transformation. To see this the above needs to be written in terms of the complex charge $Q := q + ig$ and the complex field $\mathbf{G} := \mathbf{E} + i\mathbf{B}$, together with their complex conjugates. One finds

$$\mathbf{F} = \bar{Q}\mathbf{G} + Q\bar{\mathbf{G}} + i\mathbf{v} \wedge (\bar{Q}\mathbf{G} - Q\bar{\mathbf{G}}),$$

where each piece now is separately invariant.

Turning our attention now to the vector potential, we were able to introduce this originally because $\nabla \cdot \mathbf{B}$ was exactly zero and therefore $\mathbf{B} = \nabla \wedge \mathbf{A}$ was a mathematical result. If we now have a monopole source then in spherical coordinates

$$\mathbf{B}_r = \frac{g}{4\pi r^2}.$$

Can we obtain this from a gauge potential? the curl of the gauge field has components

$$(\nabla \wedge \mathbf{A})_r = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\mathbf{A}_\phi \sin\theta) - \frac{\partial \mathbf{A}_\theta}{\partial\phi},$$
$$(\nabla \wedge \mathbf{A})_\theta = \frac{1}{r\sin\theta}\frac{\partial \mathbf{A}_r}{\partial\phi} - \frac{1}{r}\frac{\partial(r\mathbf{A}_\phi)}{\partial r},$$
$$(\nabla \wedge \mathbf{A})_\phi = \frac{1}{r}\frac{\partial(r\mathbf{A}_\theta)}{\partial r} - \frac{1}{r}\frac{\partial \mathbf{A}_r}{\partial\theta}.$$

The simplest solution we can look for is when $\mathbf{A}_r = 0$ and $\mathbf{A}_\theta = 0$. The first equation then tells us that $\mathbf{A}_\phi$ should be inversely

proportional to sin θ, while the second requires that it is inversely proportional to r. The first equation also tells us that we need a cos θ (plus a constant) in the numerator so that when differentiated it will cancel off the remaining sin θ in the denominator. One can see therefore that the gauge potential

$$\mathbf{A}_\phi = \frac{g(1-\cos\theta)}{4\pi r \sin\theta},$$

gives the monopole magnetic field. However, there is a problem because this has a singular behavior at $\theta = \pi$. It is this singular aspect that causes the standard vector calculus identity $\nabla\cdot(\nabla\wedge\mathbf{A}) = 0$ to breakdown, as it needs to if we going to extract a monopole charge. In fact it doesn't matter at all that the gauge field has a singularity provided the magnetic field shows no such behavior. We can side step the singularity in the potential by introducing a second potential. The idea here is that each potential is only to be used in a subspace where it doesn't have a singularity. Together they then can cover the whole space. An example choice would be

$$\mathbf{A}^+_\phi = \frac{g(1-\cos\theta)}{4\pi r \sin\theta}, \quad \text{for } \theta < \pi - \delta,$$
$$\mathbf{A}^-_\phi = -\frac{g(1+\cos\theta)}{4\pi r \sin\theta}, \quad \text{for } \theta > \delta,$$

the plus referring to the upper hemisphere and the minus sign, to the lower hemisphere (the δ is a small angle). Provided one uses these on their correct coordinate patch we have a solution for the magnetic field that is correct everywhere.

As a final point, to see how all this ties together, what happens to the two gauge potentials in the overlap region, for example at $\theta = \pi/2$? They are magically related by a gauge transformation as follows

$$\mathbf{A}^+_\phi(\theta=\pi/2) - \mathbf{A}^-_\phi(\theta=\pi/2) = \frac{g}{2\pi r},$$
$$= \nabla_\phi\left(\frac{g\phi}{2\pi}\right),$$

so the gauge function is $\Lambda = g\phi/2\pi$.

BIBLIOGRAPHY

[1] "Condensed Matter field Theory, A. Altland and B. Simons, Cambridge University Press (2006).

[2] "Mathematics for Physics and Physicist", W. Appel, Princeton University Press (2007).

[3] "Electrodynamics and Classical Theory of Fields and Particles" A.O. Barut, Dover (1981).

[4] "Mechanical Detection and Measurement of the Angular Momentum of Light," R. A. Beth, Phys. Rev. **50** (1936) 115.

[5] "Electricity and Magnetism: Volume I" (Third Edition), B. I. Bleaney and B. Bleaney, Oxford University Press (1989).

[6] "Electrodynamics from Ampère to Einstein", O. Darrigol, Oxford University Press (2003).

[7] "The Feynman Lectures on Physics: Volume II", R. P. Feynman, R. B. Leighton, M. Sands, Addison-Wesley Publishing (1989).

[8] "Classical electrodynamics" (Second Edition), J. D. Jackson, Wiley (1998).

[9] "Historical Roots of Gauge Invariance", J. D. Jackson and L. B. Okun, Rev. Mod. Phys. **73** (2001) 663 [hep-ph/0012061].

[10] "From Lorenz to Coulomb and other explicit gauge transformations", J. D. Jackson, Am. J. Phys. **70** (2002) 917 [physics/0204034].

[11] "The Story of Electricity", G. D. Leon, Dover (1988).

[12]"Contributions of John Henry Poynting to the Understanding of Radiation Pressure", R. Loudon and C. Baxter, Proc. R. Soc. A (2012) **468**, 1825-1838.

[13] "Classical Field Theory: Electromagnetism and Gravitation", F. E. Low, Wiley VCH (1997).

[14] "Classical Optics and its Applications" (Second Edition), M. Mansuripur, Cambridge University Press (2009).

[15] "Mathematical Methods for Physics and Engineering", K. F. Riley, M. P. Hobson and S. J. Bence (Third Edition), Cambridge University Press (2006).

[16] "Dynamics of Charged Particles and Their Radiation Field" H, Spohn, Cambridge University Press (2004).

[17] "A History of the Theories of Aether and Electricity: From the age of Descartes to the Close of the Nineteenth Century (1910)", E. T. Whittaker, Benediction Classics, Oxford (2012).

INDEX