1
A Brief History

Summary: The unification of electricity, magnetism, and light by James
Maxwell in the 1800s was a landmark in human history and has continued
even today to influence technology, physics, and mathematics in profound and
surprising ways. Its history (of which we give a brief overview in this chapter)
has been and continues to be studied by historians of science.

1.1. Pre-1820: The Two Subjects of Electricity and Magnetism

Who knows when our ancestors first became aware of electricity and
magnetism? I imagine a primitive cave person, wrapped up in mastodon fur,
desperately trying to stay warm in the dead of winter, suddenly seeing a spark
of static electricity. Maybe at the same time in our prehistory someone felt
a small piece of iron jump out of their hand toward a lodestone. Certainly
lightning must have been inspiring and frightening, as it still is.

But only recently (meaning in the last four hundred years) have these
phenomena been at all understood. Around 1600, William Gilbert wrote his
infuential *De Magnete*, in which he argued that the earth was just one big
magnet. In the mid-1700s, Benjamin Franklin showed that lightning was
indeed electricity. Also in the 1700s Coulomb's law was discovered, which
states that the force F between two stationary charges is

$$F = \frac{q_1 q_2}{r^2},$$

where q_1 and q_2 are the charges and r is the distance between the charges
(after choosing correct units). Further, in the 1740s, Leyden jars were
invented to store electric charge. Finally, still in the 1700s, Galvani and Volta,
independently, discovered how to generate electric charges, with the invention
of galvanic, or voltaic, cells (batteries).

1

1.2. 1820–1861: The Experimental Glory Days of Electricity and Magnetism

In 1820, possibly during a lecture, Hans Christian Oersted happened to move a compass near a wire that carried a current. He noticed that the compass's needle jumped. People knew that compasses worked via magnetism and at the same time realized that current was flowing electricity. Oersted found solid proof that electricity and magnetism were linked.

For the next forty or so years amazing progress was made finding out how these two forces were related. Most of this work was rooted in experiment. While many scientists threw themselves into this hunt, Faraday stands out as a truly profound experimental scientist. By the end of this era, most of the basic empirical connections between electricity and magnetism had been discovered.

1.3. Maxwell and His Four Equations

In the early 1860s, James Clerk Maxwell wrote down his four equations that linked the electric field with the magnetic field. (The real history is quite a bit more complicated.) These equations contain within them the prediction that there are electromagnetic waves, traveling at some speed c. Maxwell observed that this speed c was close to the observed speed of light. This led him to make the spectacular conjecture that light is an electromagnetic wave. Suddenly, light, electricity, and magnetism were all part of the same fundamental phenomenon.

Within twenty years, Hertz had experimentally shown that light was indeed an electromagnetic wave. (As seen in Chapter 6 of [27], the actual history is not quite such a clean story.)

1.4. Einstein and the Special Theory of Relativity

All electromagnetic waves, which after Maxwell were known to include light waves, have a remarkable yet disturbing property: These waves travel at a fixed speed c. This fact was not controversial at all, until it was realized that this speed was independent of any frame of reference.

To make this surprise more concrete, we turn to Einstein's example of shining lights on trains. (No doubt today the example would be framed in terms of airplanes or rocket ships.) Imagine you are on a train traveling at 60 miles per hour. You turn on a flashlight and point it in the same direction as the train is moving. To you, the light moves at a speed of c (you think your

speed is zero miles per hour). To someone on the side of the road, the light
should move at a speed of 60 miles per hour $+c$. But according to Maxwell's
equations, it does not. The observer off the train will actually see the light
move at the same speed c, which is no different from your observation on the
train. This is wacky and suggests that Maxwell's equations must be wrong.

In actual experiments, though, it is our common sense (codified in
Newtonian mechanics) that is wrong. This led Albert Einstein, in 1905, to
propose an entirely new theory of mechanics, the special theory. In large part,
Einstein discovered the special theory because he took Maxwell's equations
seriously as a statement about the fundamental nature of reality.

1.5. Quantum Mechanics and Photons

What is light? For many years scientists debated whether light was made
up of particles or of waves. After Maxwell (and especially after Hertz's
experiments showing that light is indeed a type of electromagnetic wave), it
seemed that the debate had been settled. But in the late nineteenth century,
a weird new phenomenon was observed. When light was shone on certain
metals, electrons were ejected from the metal. Something in light carried
enough energy to forcibly eject electrons from the metal. This phenomenon is
called the *photoelectric effect*. This alone is not shocking, as it was well known
that traditional waves carried energy. (Many of us have been knocked over by
ocean waves at the beach.) In classical physics, though, the energy carried by a
traditional wave is proportional to the wave's amplitude (how high it gets). But
in the photoelectric effect, the energy of the ejected electrons is proportional
not to the amplitude of the light wave but instead to the light's frequency. This
is a decidedly non-classical effect, jeopardizing a wave interpretation for light.

In 1905, in the same year that he developed the Special Theory of Relativity,
Einstein gave an interpretation to light that seemed to explain the photoelectric
effect. Instead of thinking of light as a wave (in which case, the energy would
have to be proportional to the light's amplitude), Einstein assumed that light is
made of particles, each of which has energy proportional to the frequency, and
showed that this assumption leads to the correct experimental predictions.

In the context of other seemingly strange experimental results, people
started to investigate what is now called quantum mechanics, amassing a
number of partial explanations. Suddenly, over the course of a few years in the
mid-1920s, Born, Dirac, Heisenberg, Jordan, Schrödinger, von Neumann, and
others worked out the complete theory, finishing the first quantum revolution.
We will see that this theory indeed leads to the prediction that light must have
properties of both waves and particles.

1.6. Gauge Theories for Physicists:
The Standard Model

At the end of the 1920s, gravity and electromagnetism were the only two known forces. By the end of the 1930s, both the strong force and the weak force had been discovered.

In the nucleus of an atom, protons and neutrons are crammed together. All of the protons have positive charge. The rules of electromagnetism would predict that these protons would want to explode away from each other, but this does not happen. It is the strong force that holds the protons and neutrons together in the nucleus, and it is called such since it must be strong enough to overcome the repelling force of electromagnetism.

The weak force can be seen in the decay of the neutron. If a neutron is just sitting around, after ten or fifteen minutes it will decay into a proton, an electron, and another elementary particle (the electron anti-neutrino, to be precise). This could not be explained by the other forces, leading to the discovery of this new force.

Since both of these forces were basically described in the 1930s, their theories were quantum mechanical. But in the 1960s, a common framework for the weak force and the electromagnetic force was worked out (resulting in Nobel Prizes for Abdus Salam, Sheldon Glashow, and Steven Weinberg in 1979). In fact, this framework can be extended to include the strong force. This common framework goes by the name of the *standard model*. (It does not include gravity.)

Much earlier, in the 1920s, the mathematician Herman Weyl attempted to unite gravity and electromagnetism, by developing what he called a *gauge theory*. While it quickly was shown not to be physically realistic, the underlying idea was sufficiently intriguing that it resurfaced in the early 1950s in the work of Yang and Mills, who were studying the strong force. The underlying mathematics of their work is what led to the unified electro-weak force and the standard model.

Weyl's gauge theory was motivated by symmetry. He used the word "gauge" to suggest different gauges for railroad tracks. His work was motivated by the desire to shift from global symmetries to local symmetries. We will start with global symmetries. Think of the room you are sitting in. Choose a corner and label this the origin. Assume one of the edges is the x-axis, another the y-axis, and the third the z-axis. Put some unit of length on these edges. You can now uniquely label any point in the room by three coordinate values.

Of course, someone else might have chosen a different corner as the origin, different coordinate axes, or different units of length. In fact, any point in the room (or, for that matter, any point in space) could be used as the origin, and so on. There are an amazing number of different choices.

Now imagine a bird flying in the room. With your coordinate system, you could describe the path of the bird's flight by a curve $(x(t), y(t), z(t))$. Someone else, with a different coordinate system, will describe the flight of the bird by three totally different functions. The flight is the same (after all, the bird does not care what coordinate system you are using), but the description is different. By changing coordinates, we can translate from one coordinate system into the other. This is a global change of coordinates. Part of the deep insight of the theory of relativity, as we will see, is that which coordinate changes are allowed has profound effects on the description of reality.

Weyl took this one step further. Instead of choosing one global coordinate system, he proposed that we could choose different coordinate systems at each point of space but that all of these local coordinate systems must be capable of being suitably patched together. Weyl called this patching "choosing a gauge."

1.7. Four-Manifolds

During the 1950s, 1960s, and early 1970s, when physicists were developing what they called gauge theory, leading to the standard model, mathematicians were developing the foundations of differential geometry. (Actually this work on differential geometry went back quite a bit further than the 1950s.) This mainly involved understanding the correct nature of curvature, which, in turn, as we will see, involves understanding the nature of connections. But sometime in the 1960s or 1970s, people must have begun to notice uncanny similarities between the physicists' gauges and the mathematicians' connections. Finally, in 1975, Wu and Yang [69] wrote out the dictionary between the two languages (this is the same Yang who was part of Yang-Mills). This alone was amazing. Here the foundations of much of modern physics were shown to be the same as the foundations of much of differential geometry.

Through most of the twentieth century, when math and physics interacted, overwhelmingly it was the case that math shaped physics:

$$\text{Mathematics} \Rightarrow \text{Physics}$$

Come the early 1980s, the arrow was reversed. Among all possible gauges, physicists pick out those that are Yang-Mills, which are in turn deep generalizations of Maxwell's equations. By the preceding dictionary, connections that satisfy Yang-Mills should be special.

This leads us to the revolutionary work of Simon Donaldson. He was interested in four-dimensional manifolds. On a four-manifold, there is the space of all possible connections. (We are ignoring some significant facts.) This space is infinite dimensional and has little structure. But then Donaldson decided to take physicists seriously. He looked at those connections that were Yang-Mills. (Another common term used is "instantons.") At the time, there was no compelling mathematical reason to do this. Also, his four-manifolds were not physical objects and had no apparent link with physics. Still, he looked at Yang-Mills connections and discovered amazing, deeply surprising structure, such as that these special Yang-Mills connections form a five-dimensional space, which has the original four-manifold as part of its boundary. (Here we are coming close to almost criminal simplification, but the underlying idea that the Yang-Mills connections are linked to a five-manifold that has the four-manifold as part of its boundary is correct.) This work shocked much of the mathematical world and transformed four-manifold theory from a perfectly respectable area of mathematics into one of its hottest branches. In awarding Donaldson a Field's Medal in 1986, Atiyah [1] wrote:

> The surprise produced by Donaldson's result was accentuated by the fact that his methods were completely new and were borrowed from theoretical physics, in the form of Yang-Mills equations. ... Several mathematicians (including myself) worked on instantons and felt very pleased that they were able to assist physics in this way. Donaldson, on the other hand, conceived the daring idea of reversing this process and of using instantons on a general 4-manifold as a new geometrical tool.

Many of the finest mathematicians of the 1980s started working on developing this theory, people such as Atiyah, Bott, Uhlenbeck, Taubes, Yau, Kobayashi, and others.

Not only did this work produce some beautiful mathematics, it changed how math could be done. Now we have

$$\text{Physics} \Rightarrow \text{Mathematics}$$

an approach that should be called *physical mathematics* (a term first coined by Kishore Marathe, according to [70]: This text by Zeidler is an excellent place to begin to see the power behind the idea of physical mathematics).

Physical mathematics involves taking some part of the real world that is physically important (such as Maxwell's equations), identifying the underlying mathematics, and then taking that mathematics seriously, even in contexts far removed from the natural world. This has been a major theme of mathematics

since the 1980s, led primarily by the brilliant work of Edward Witten. When
Witten won his Field's Medal in 1990, Atiyah [2] wrote:

> Although (Witten) is definitely a physicist his command of mathematics is rivaled
> by few mathematicians, and his ability to interpret physical ideas in mathematical
> form is quite unique. Time and again he has surprised the mathematical community
> by a brilliant application of physical insight leading to new and deep mathematical
> theorems.

The punchline is that mathematicians should take seriously underlying
mathematical structure of the real world, even in non-real world situations.
In essence, nature is a superb mathematician.

1.8. This Book

There is a problem with this revolution of physical mathematics. How can any
mere mortal master both physics and mathematics? The answer, of course, is
you cannot. This book is a compromise. We concentrate on the key underlying
mathematical concepts behind the physics, trying at the same time to explain
just enough of the real world to justify the use of the mathematics. By the
end of this book, I hope the reader will be able to start understanding the work
needed to understand Yang-Mills.

1.9. Some Sources

One way to learn a subject is to study its history. That is not the approach we
are taking. There are a number of good, accessible books, though. Stephen
J. Blundell's *Magnetism: A Very Short Introduction* [4] is excellent for a
popular general overview. For more technical histories of the early days of
electromagnetism, I would recommend Steinle's article "Electromagnetism
and Field Physics" [62] and Buchwald's article "Electrodynamics from
Thomson and Maxwell to Hertz" [7].

Later in his career, Abraham Pais wrote three excellent books covering
much of the history of twentieth century physics. His *Subtle Is the Lord: The
Science and the Life of Albert Einstein* [51] is a beautiful scientific biography
of Einstein, which means that it is also a history of much of what was important
in physics in the first third of the 1900s. His *Niels Bohr's Times: In Physics,
Philosophy, and Polity* [52] is a scientific biography of Bohr, and hence a good
overview of the history of early quantum mechanics. His *Inward Bound* [53]

is a further good reference for the development of quantum theory and particle physics.

It appears that the ideas of special relativity were "in the air" around 1905. For some of the original papers by Einstein, Lorentz, Minkowski, and Weyl, there is the collection [19]. Poincaré was also actively involved in the early days of special relativity. Recently two biographies of Poincaré have been written: Gray's *Henri Poincaré: A Scientific Biography* [27] and Verhulst's *Henri Poincaré: Impatient Genius* [67]. There is also the still interesting paper of Poincaré that he gave at the World's Fair in Saint Louis in 1904, which has recently been reprinted [54].

At the end of this book, we reach the beginnings of gauge theory. In [50], O'Raifeartaigh has collected some of the seminal papers in the development of gauge theory. We encourage the reader to look at the web page of Edward Witten for inspiration. I would also encourage people to look at many of the expository papers on the relationship between mathematics and physics in volume 6 of the collected works of Atiyah [3] and at those in volume 4 of the collected works of Bott [5]. (In fact, perusing all six volumes of Atiyah's collected works and all four volumes of Bott's is an excellent way to be exposed to many of the main themes of mathematics of the last half of the twentieth century.) Finally, there is the wonderful best seller *The Elegant Universe* by Brian Greene [28].

2

Maxwell's Equations

Summary: The primary goal of this chapter is to state Maxwell's equations. We will then see some of their implications, which will allow us to give alternative descriptions for Maxwell's equations, providing us in turn with a review of some of the basic formulas in multivariable calculus.

2.1. A Statement of Maxwell's Equations

Maxwell's equations link together three vector fields and a real-valued function. Let

$$E = E(x,y,z,t) = (E_1(x,y,z,t), E_2(x,y,z,t), E_3(x,y,z,t))$$

and

$$B = B(x,y,z,t) = (B_1(x,y,z,t), B_2(x,y,z,t), B_3(x,y,z,t))$$

be two vector fields with spacial coordinates (x,y,z) and time coordinate t. Here E represents the electric field while B represents the magnetic field. The third vector field is

$$j(x,y,z,t) = (j_1(x,y,z,t), j_2(x,y,z,t), j_3(x,y,z,t)),$$

which represents the current (the direction and the magnitude of the flow of electric charge). Finally, let

$$\rho(x,y,z,t)$$

be a function representing the charge density. Let c be a constant. (Here c is the speed of light in a vacuum.) Then these three vector fields and this function
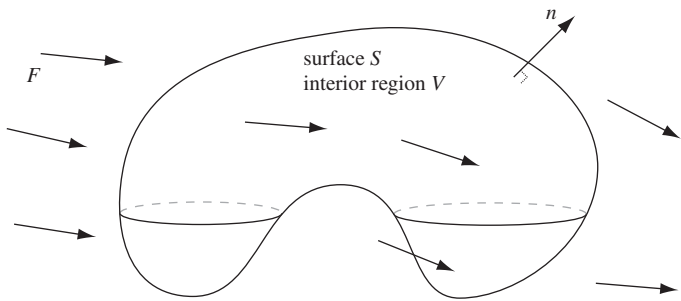
Figure 2.1

satisfy Maxwell's equations if

$$\operatorname{div}(E) = \rho$$
$$\operatorname{curl}(E) = -\frac{\partial B}{\partial t}$$
$$\operatorname{div}(B) = 0$$
$$c^2 \operatorname{curl}(B) = j + \frac{\partial E}{\partial t}.$$

(Review of the curl, the divergence, and other formulas from multivariable calculus is in the next section.)

We can reinterpret these equations in terms of integrals via various Stokes-type theorems. For example, if V is a compact region in space with smooth boundary surface S, as in Figure 2.1, then for any vector field F we know from the Divergence Theorem that

$$\int\int_S F \cdot n \, \mathrm{d}A = \int\int\int_V \operatorname{div}(F) \, \mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

where n is the unit outward normal of the surface S.

In words, this theorem says that the divergence of a vector field measures how much of the field is flowing out of a region.

Then the first of Maxwell's equations can be restated as

$$\int\int_S E \cdot n \, \mathrm{d}A = \int\int\int \operatorname{div}(E) \, \mathrm{d}x\mathrm{d}y\mathrm{d}z$$
$$= \int\int\int \rho(x,y,z,t) \, \mathrm{d}x\mathrm{d}y\mathrm{d}z$$
$$= \text{total charge inside the region } V.$$