Chapter 1

Maxwell's Equations

Our whole progress up to this point may be described as a gradual development of the doctrine of relativity of all physical phenomena. Position we must evidently acknowledge to be relative, for we cannot describe the position of a body in any terms which do not express relation. The ordinary language about motion and rest does not so completely exclude the notion of their being measured absolutely, but the reason of this is, that in our ordinary language we tacitly assume that the earth is at rest.... There are no landmarks in space; one portion of space is exactly like every other portion, so that we cannot tell where we are. We are, as it were, on an unruffled sea, without stars, compass, sounding, wind or tide, and we cannot tell in what direction we are going. We have no log which we can cast out to take a dead reckoning by; we may compute our rate of motion with respect to the neighboring bodies, but we do not know how these bodies may be moving in space. – James Clerk Maxwell, 1876.

Starting with Maxwell's beautiful theory of electromagnetism, and inspired by it, physicists have made tremendous progress in understanding the basic forces and particles constituting the physical world. Maxwell showed that two seemingly very different forces, the electric and magnetic forces, were simply two aspects of the 'electromagnetic field' . In so doing, he was also able to explain *light* as a phenomenon in which ripples in the electric field create ripples in the magnetic field, which in turn create new ripples in the electric field, and so on. Shockingly, however, Maxwell's theory also predicted that light emitted by a moving body would travel no faster than light from a stationary body.

Eventually this led Lorentz, Poincaré and especially Einstein to realize that our ideas about space and time had to be radically revised. That the motion of a body can only be measured relative to another body had been understood to some extent since Galileo. Taken in conjunction with Maxwell's theory, however, this principle forced the recognition that in addition to the rotational symmetries of space there must be symmetries that mingle the space and time coordinates. These new symmetries also mix the electric and magnetic fields, charge and current, energy and momentum, and so on, revealing the world to be much more coherent and tightly-knit than had previously been suspected.

There are, of course, forces in nature besides electromagnetism, the most obvious of which is gravity. Indeed, it was the simplicity of gravity that gave rise the first conquests of modern physics: Kepler's laws of planetary motion, and then Newton's laws unifying celestial mechanics with the mechanics of falling bodies. However, reconciling the simplicity of gravity with relativity theory was no easy task! In seeking equations for gravity consistent with his theory of special relativity, Einstein naturally sought to copy the model of Maxwell's equations. However, the result was not merely a theory in which ripples of some field propagate through spacetime, but a theory in which the geometry of spacetime itself ripples and bends. Einstein's equations say, roughly, that energy and momentum affect the *metric* of spacetime (whereby we measure time and distance) much as charges and currents affect the electromagnetic field. This served to heighten hopes that much or perhaps even all of physics is fundamentally *geometrical* in character.

There were, however, severe challenges to these hopes. Attempts by Einstein, Weyl, Kaluza and Klein to further unify our description of the forces of nature using ideas from geometry were largely unsuccessful. The reason is that the careful study of atoms, nuclei and subatomic particles revealed a wealth of phenomena that do not fit easily into any simple scheme. Each time technology permitted the study of smaller distance scales (or equivalently, higher energies), new puzzles arose. In part, the reason is that physics at small distance scales is completely dominated by the principles of *quantum theory*. The naive notion that a particle is a point tracing out a path in spacetime, or that a field assigns a number or vector to each point of spacetime, proved to be wholly inadequate, for one cannot measure the position and velocity of a particle simultaneously with arbitrary accuracy, nor the value of a field and its time derivative. Indeed, it turned out that the distinction between a particle and field was somewhat arbitrary. Much of 20th century physics has centered around the task of making sense of microworld and developing a framework with which one can understand subatomic particles and the forces between them in the light of quantum theory.

Our current picture, called the standard model, involves three forces: electromagnetism and the weak and strong nuclear forces. These are all 'gauge fields', meaning that they are described by equations closely modelled after Maxwell's equations. These equations describe *quantum* fields, so the forces can be regarded as carried by particles: the electromagnetic force is carried by the photon, the weak force is carried by the W and Z particles, and the strong force is carried by gluons. There are also charged particles that interact with these force-carrying particles. By 'charge' here we mean not only the electric charge but also its analogs for the other forces. There are two main kinds of charged particles, quarks (which feel the strong force) and leptons (which do not). All of these charged particles have corresponding antiparticles of the same mass and opposite charge.

Somewhat mysteriously, the charged particles come in three families or 'generations'. The first generation consists of two leptons, the electron e and the electron neutrino ν_e, and two quarks, the up and down, or u and d. Most of the matter we see everyday is made out of these first-generation particles. For example, according to the standard model the proton is a composite of two up quarks and one down, while the neutron is two downs and an up. There is a second generation of quarks and leptons, the muon μ and muon neutrino ν_μ, and the charmed and strange quarks c, s. For the most part these are heavier than the corresponding particles in the first generation, although all the neutrinos appear to be massless or nearly so. For example, the muon is about 207 times as massive as the electron, but almost identical in every other respect. Then there is a third, still more massive generation, containing the tau τ and tau neutrino ν_τ, and the top and bottom quarks t and b. For many years the top quark was merely conjectured to exist, but just as this book went to press, experimentalists announced that it may finally have been found.

Finally, there is a very odd charged particle in the standard model, the Higgs particle, which is neither a quark nor a lepton. This has not been observed either, and is hypothesized to exist primarily to explain the relation between the elecromagnetic and weak forces.

Even more puzzling than all the complexities of the standard model, however, is the question of where *gravity* fits into the picture! Einstein's equations describing gravity do *not* take quantum theory into account, and it has proved very difficult to 'quantize' them. We thus have not one picture of the world, but two: the standard model, in which all forces except gravity are described in accordance with quantum theory, and general relativity, in which gravity alone is described, not in accordance with quantum theory. Unfortunately it seems difficult to obtain guidance from experiment; simple considerations of dimensional analysis suggest that quantum gravity effects may become significant at distance scales comparable to the **Planck length**,

$$\ell_p = (\hbar\kappa/c^3)^{1/2},$$

where $\hbar$ is Planck's constant, κ is Newton's gravitational constant, and c is the speed of light. The Planck length is about $1.616 \cdot 10^{-35}$ meters, far below the length scales we can probe with particle accelerators.

Recent developments, however, hint that gravity may be closer to the gauge theories of the standard model than had been thought. Fascinatingly, the relationship also involves the study of *knots* in 3-dimensional space. While this work is in its early stages, and may not succeed as a theory of physics, the new mathematics involved is so beautiful that it is difficult to resist becoming excited. Unfortunately, understanding these new ideas depends on a thorough mastery of quantum field theory, general relativity, geometry, topology, and algebra. Indeed, it is almost certain that *nobody* is sufficiently prepared to understand these ideas fully! The reader should therefore not expect to understand them when done with this book. Our goal in this book is simply to start fairly near the beginning of the story and bring the reader far enough along to see the frontiers of current research in dim outline.

We must begin by reviewing some geometry. These days, when mathematicians speak of geometry they are usually referring not to Euclidean geometry but to the many modern generalizations that fall under the heading of 'differential geometry'. The first theory of physics to *explicitly* use differential geometry was Einstein's general relativity, in which gravity is explained as the curvature of spacetime. The gauge theories of the standard model are of a very similar geometrical character (although quantized). But there is also a lot of differential geometry lurking in Maxwell's equations, which after all were the inspiration for both general relativity and gauge theory. So, just as a good way to master auto repair is to take apart an old car and put in a new engine so that it runs better, we will begin by taking apart Maxwell's equations and putting them back together using modern differential geometry.

In their classic form, Maxwell's equations describe the behavior of two vector fields, the **electric field** $\vec{E}$ and the **magnetic field** $\vec{B}$. These fields are defined throughout space, which is taken to be $\mathbb{R}^3$. However, they are also functions of time, a real-valued parameter t. The electric and magnetic fields depend on the electric **charge density** ρ, which is a time-dependent function on space, and also on the electric **current density** $\vec{\jmath}$, which is time-dependent vector field on space. (For the mathematicians, let us note that unless otherwise specified, functions are assumed to be real-valued, and functions and vector fields on $\mathbb{R}^n$ are assumed to be **smooth**, that is, infinitely differentiable.)

In units where the speed of light is equal to 1, **Maxwell's equations** are:

$$\begin{aligned}
\nabla \cdot \vec{B} &= 0 \\
\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} &= 0 \\
\nabla \cdot \vec{E} &= \rho \\
\nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} &= \vec{\jmath}.
\end{aligned}$$

There are a number of interesting things about these equations that are worth understanding. First, there is the little fact that we can only determine the direction of the magnetic field experimentally if we know the difference between right and left. This is easiest to see from the **Lorentz force law**, which says that the force on a charged particle with charge q and velocity $\vec{v}$ is

$$\vec{F} = q\,(\vec{E} + \vec{v} \times \vec{B}).$$

To measure $\vec{E}$, we need only measure the force $\vec{F}$ on a static particle and divide by q. To figure out $\vec{B}$, we can measure the force on charged particles with a variety of velocities. However, recall that the definition of the cross product involves a completely arbitrary right-hand rule! We typically define

$$\vec{v} \times \vec{B} = (v_yB_z - v_zB_y, v_zB_x - v_xB_z, v_xB_y - v_yB_x).$$

However, this is just a convention; we could have set

$$\vec{v} \times \vec{B} = (v_zB_y - v_yB_z, v_xB_z - v_zB_xv_yB_x - v_xB_y),$$

and all the mathematics of cross products would work just as well. If we used this 'left-handed cross product' when figuring out $\vec{B}$ from measurements of $\vec{F}$ for various velocities $\vec{v}$, we would get an answer for $\vec{B}$ with the opposite of the usual sign! It may seem odd that $\vec{B}$ depends on an arbitrary convention this way. In fact, this turns out to be an important clue as to the mathematical structure of Maxwell's equations.

Secondly, Maxwell's equations naturally come in two pairs. The pair that does not involve the electric charge and current densities

$$\nabla \cdot \vec{B} = 0 \qquad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0,$$

looks very much like the pair that *does*:

$$\nabla \cdot \vec{E} = \rho \qquad \nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{\jmath}.$$

Note the funny minus sign in the second pair. The symmetry is clearest in the **vacuum** Maxwell equations, where the charge and current densities vanish:

$$\nabla \cdot \vec{B} = 0 \qquad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0,$$

$$\nabla \cdot \vec{E} = 0 \qquad \nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = 0.$$

Then the transformation

$$\vec{B} \mapsto \vec{E}, \qquad \vec{E} \mapsto -\vec{B}$$

takes the first pair of equations to the second and vice versa! This symmetry is called **duality** and is a clue that the electric and magnetic fields are part of a unified whole, the electromagnetic field. Indeed, if we introduce a complex-valued vector field

$$\vec{\mathcal{E}} = \vec{E} + i\vec{B},$$

duality amounts to the transformation

$$\vec{\mathcal{E}} \mapsto i\vec{\mathcal{E}},$$

and the vacuum Maxwell equations boil down to two equations for $\vec{\mathcal{E}}$:

$$\nabla \cdot \vec{\mathcal{E}} = 0, \qquad \nabla \times \vec{\mathcal{E}} = i\frac{\partial \vec{\mathcal{E}}}{\partial t}$$

This trick has very practical applications. For example, one can use it to find solutions that correspond to plane waves moving along at the speed of light, which in the units we are using equals 1.

Exercise 1. *Let* $\vec{k}$ *be a vector in* $\mathbb{R}^3$ *and let* $\omega = |\vec{k}|$. *Fix* $\vec{\mathbf{E}} \in \mathbb{C}^3$ *with* $\vec{k} \cdot \vec{\mathbf{E}} = 0$ *and* $i\vec{k} \times \vec{\mathbf{E}} = \omega\vec{\mathbf{E}}$. *Show that*

$$\vec{\mathcal{E}}(t, \vec{x}) = \vec{\mathbf{E}}\, e^{-i(\omega t - \vec{k} \cdot \vec{x})}$$

satisfies the vacuum Maxwell equations.

The symmetry between $\vec{E}$ and $\vec{B}$ does not, however, extend to the non-vacuum Maxwell equations. We can consider making ρ and $\vec{\jmath}$ complex, and writing down:

$$\nabla \cdot \vec{\mathcal{E}} = \rho, \qquad \nabla \times \vec{\mathcal{E}} = i(\frac{\partial \vec{\mathcal{E}}}{\partial t} + \vec{\jmath}).$$

However, this amounts to introducing magnetic charge and current density, since if we split ρ and $\vec{\jmath}$ into real and imaginary parts, we see that the imaginary parts play the role of magnetic charge and current densities:

$$\begin{aligned} \rho &= \rho_e + i\rho_m, \\ \vec{\jmath} &= \vec{\jmath}_e + i\vec{\jmath}_m. \end{aligned}$$

We get

$$\nabla \cdot \vec{B} = \rho_m \qquad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = -\vec{\jmath}_m,$$

$$\nabla \cdot \vec{E} = \rho_e \qquad \nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{\jmath}_e.$$

These equations are quite charming, but unfortunately no magnetic charges — so called **magnetic monopoles** — have been observed! (We will have a bit more to say about this in Chapter 6.) We could simply keep these equations and say that ρ and $\vec{\jmath}$ are real-valued on the basis of experimental evidence. But it is a mathematical as well as a physical challenge to find a better way of understanding this phenomenon. It turns out that the formalism of gauge theory makes it seem quite natural.

Finally, there is the connection between Maxwell's equations and special relativity. The main idea of special relativity is that in addition to the symmetries of space (translations and rotations) and time (translations) there are equally important symmetries mixing space and time, the Lorentz transformations. The idea is that if you and I are both unaccelerated, so that my velocity with respect to you is constant, the coordinates I will naturally use, in which I am at rest, will differ from yours, in which you are at rest. If your coordinate system is (t, x, y, z) and I am moving with velocity v in the x direction with respect to you, for example, the coordinates in which I am at rest are given by

$$\begin{aligned} t' &= (\cosh\phi)t - (\sinh\phi)x \\ x' &= -(\sinh\phi)t + (\cosh\phi)x \\ y' &= y \\ z' &= z, \end{aligned}$$

where ϕ is a convenient quantity called the **rapidity**, defined so that $\tanh\phi = v$. Note the close resemblance to the formula for rotations in space. The idea is that just as the x, y, and z components of position are all just aspects of something more important, the position itself, space and time are just aspects of a unitary whole, *spacetime*.

Maxwell's equations are invariant under these Lorentz transformations — indeed, this was the main fact that led Einstein to special relativity! He realized that Maxwell's equations predict that *any* unaccelerated observer will measure light moving in *any* direction in the vacuum to have the *same* speed. Mathematically speaking, the point is that if we have a solution of Maxwell's equations and we do a Lorentz transformation on the coordinates together with a certain transformation of $\vec{E}$, $\vec{B}$, ρ and $\vec{j}$, we again have a solution.

For example, suppose that we do a Lorentz transformation of velocity v in the x direction, as above. The precise recipe for transforming the charge and current densities is

$$\begin{aligned} \rho' &= (\cosh\phi)\rho - (\sinh\phi)j_x \\ j'_x &= -(\sinh\phi)\rho + (\cosh\phi)j_x \\ j'_y &= j_y \\ j'_z &= j_z. \end{aligned}$$

Note that ρ and $\vec{j}$ get mixed up together. In fact, we shall see that they are really just two aspects of a single thing called the 'current', which has ρ as its component in the time direction and j_x, j_y, j_z as its components in the space directions.

The formula for transforming the electric and magnetic fields under the same Lorentz transformation is somewhat more complicated:

$$\begin{aligned} E'_x &= E_x \\ E'_y &= (\cosh\phi)E_y - (\sinh\phi)B_z \\ E'_z &= (\sinh\phi)B_y + (\cosh\phi)E_z, \end{aligned}$$

$$\begin{aligned} B'_x &= B_x \\ B'_y &= (\cosh\phi)B_y + (\sinh\phi)E_z \\ B'_z &= -(\sinh\phi)E_y + (\cosh\phi)B_z. \end{aligned}$$

The most important message here is that the electric and magnetic fields are two aspects of a unified 'electromagnetic field'. Also, we see that the electromagnetic field is more complicated in character than the current, since it has six independent components that transform in a more subtle manner. It turns out to be a '2-form'.

When we have rewritten Maxwell's equations using the language of differential geometry, all the things we have just discussed will be much clearer — at least if we succeed in explaining things well. The key step, which is somewhat shocking to the uninitiated, is to work as much as possible in a manner that does not require a choice of coordinates. After all, as far as we can tell, the world was *not* drawn on graph paper. Coordinates are merely something *we* introduce for our own convenience, and the laws of physics should not care which coordinates we happen to use. If we postpone introducing coordinates until it is actually necessary, we will not have to do anything to show that Maxwell's equations are invariant under Lorentz transformations; it will be *manifest.*

Just for fun, let us write down the new version of Maxwell's equations right away. We will explain what they *mean* quite a bit later, so do not worry if they are fairly cryptic. They are:

$$\begin{aligned} dF &= 0 \\ \star d \star F &= J. \end{aligned}$$

Here F is the 'electromagnetic field' and J is the 'current', while the d and $\star$ operators are slick ways of summarizing all the curls, divergences and time derivatives that appear in the old-fashioned version. The equation $dF = 0$ is equivalent to the first pair of Maxwell's equations, while the equation $\star d \star F = J$ is equivalent to the second pair. The 'funny minus sign' in the second pair will turn out to be a natural consequence of how the $\star$ operator works.

If the reader is too pragmatic to get excited by the terse beauty of this new-fangled version of Maxwell's equations, let us emphasize that this way of writing them is a warm-up for understanding gauge theory, and allows us to study Maxwell's equations and gauge theory on curved spacetimes, as one needs to in general relativity. Indeed, we will start by developing enough differential geometry to do a fair amount of physics on general spacetimes. Then we will come back to Maxwell's equations. We warn the reader that the next few sections are not really a solid course in differential geometry. Whenever something is at all tricky to prove we will skip it! The easygoing reader can take some facts on faith; the careful reader may want to get ahold of a good book on differential geometry to help fill in these details. Some suggestions on books appear in the notes at the end of Part I.