Raymond Brock

Quarks,
Spacetime,
and the Big Bang

Michigan State University

Published by Michigan State University

Michigan State University

First printing, June 2017

Contents

“Nature Loves to Hide”

Heraclitus

Preface:

Quarks, Spacetime, and the Big Bang

QS&BB is a book designed to accompany a general education course of the same name that I've taught at Michigan State University for a number of years. Why? Well, there's a story there.

The North American approach to university education is nearly unique in the world. Citizen-students come to college in order to become proficient in a focused few areas of study (your "major") but are also broadly educated in many other areas ("general education"). So an English major would dive deeply into literature but also take courses in maybe physics, astronomy, chemistry, biology, geology, history, anthropology, psychology, etc. Likewise a physics major would study physics and mathematics, but also biology, literature, psychology, and so on. Every U.S. campus manages this deep-plus-broad approach to higher education in its own way.[1]

Creating courses for non-specialists in the sciences is especially challenging, but it's important because many of society's big problems are scientific at their roots.[2] An informed citizen needs to understand some scientific facts, while also appreciating the scientific method: all too often, controversy swirls as much around what is or isn't "science" as it does to the details. How best to do this in physics?

There are many physics courses for non-science college students. The traditional course is often called "Physics for Poets," which is a conceptual (less mathematical) version of the otherwise full-physics curriculum taught to science and engineering students. But there are other paths which teach physics by shining a light on particularly interesting topics in accessible presentations.[3]

[1] This approach to higher education is credited to the Harvard University president Abbott Lawrence Lowell who began transforming undergraduate education in 1909. Under him, fields of concentration (majors) were established along with required sampling of courses outside of majors, the distribution requirement. "A well-educated man must know a little bit of everything and one thing well." affected college education across America to this day.

[2] Climate change. Energy production. Evolution and big bang in schools. Nuclear power. Nuclear proliferation. NASA. NIH. Vaccination. Pandemics. Weather. Health effects (or not) of common radiation sources. Peer review. Basic versus applied research. And so on.

[3] Many physics departments will offer astronomy courses (or of course, astronomy departments will when they exist), physics of music, physics of energy issues, physics of light, and so on. Our department is no different in that respect. By the way, 50,000 students take college-credit astronomy every year in the United States!

What QS&BB Isn't

This book is not a comprehensive survey of all of physics. A student will not be expected to solve many of the standard "physics class" problems—QS&BB is, intentionally, mostly conceptual. Many topics which would be in a conventional course are not covered here, or are touched on lightly. For example, there is no chapter on thermodynamics nor on energy production or climate. Motion and forces are only presented for one-dimensional situations and only sufficiently to appreciate relativity. Electricity and magnetism are covered in a descriptive way, with only a few quantitative examples. "How things work" is sometimes covered, but less so than from the usual survey course.

We cut a strategic path through "classical" areas of physics in order to accumulate the concepts, quantities, and vocabulary that would apply to a conceptual appreciation of relativity and quantum mechanics, both of which are the jumping-off points to our two main topics.

The level of scientific literacy among college-educated young adults in the United States always ranks among the top two or three among all nations of the world. This research has been done over decades by Professor Jon Miller of originally, Northwestern University and Michigan State University, and now the University of Michigan. In an article for the Association of American Colleges & Universities ("What Colleges and Universities Need to Do to Advance Civic Scientific Literacy and Preserve American Democracy" `https://www.aacu.org/node/2139`) he explains why U.S. results are so positive: "The answer is college science courses." He goes on to note that "The United States is the only country that requires all college students to take one or more science courses as a part of a general education requirement. In a series of statistical analyses using structural equation analyses of both cross-sectional and longitudinal data, I have shown that exposure to college science courses is a strong predictor of civic scientific literacy in young adults and in adults of all ages (Miller 2010a, 2010c)."

What QS&BB Is

My aim is to help you appreciate two of the more exciting "fundamental" topics in physics: particle physics and cosmology. You'll come to appreciate our current picture of how our universe began and what open questions continue to motivate thousands of us around the world. Once we've passed through a gentle introduction to motion, collisions, electricity, and magnetism, the light-algebraic approach evolves into a more conceptual narrative where we tackle modern-day topics. The Chapter 1 describes how the book—and the Michigan State course—are organized in more detail.

I emphasize biography. We'll meet intellectual giants whom everyone has heard of, but also our professional scientific heroes whose images are *not* on T-shirts. The history of physics and astronomy is full of unusual people—and a lot of just plain folks—and I'm eager for you to think of us without white coats and strange manners. We're regular people who chose career paths that are a little outside of the mainstream. But we're not so special except that we are privileged to be supported by the public in order to do our work.

I'm an experimental particle physicist and I've been teaching physics to physics majors and especially non-science students for more than three decades and I have fun doing it. I'm lucky enough to be continuously supported by you[4] for my research in particle physics for three decades and I'm grateful. In some ways, this book and course are in partial repayment for that support.

[4] The National Science Foundation, specifically.

I've never met anyone who didn't share my curiosity in wanting to know how the universe works. Even after a lifetime immersed in these matters daily, I'm constantly in awe at how beautiful it all is and how lucky we are to know as much as we do. I enjoy talking about it and teaching some of the details.

Figure 1: You can find more about me at http://www.pa.msu.edu/~brock/. You'll get to know me as I tell you stories in the pages that follow. Unfortunately, I'll not be able to meet you!

I'm not stuffy. I've tried to write here like I teach, which is informally and hopefully without pretense. I'm deadly serious about the science and passionate about the subject matter. But I also like to have fun and hopefully I'll make you smile every once in a while as we work to grasp complex ideas. Stay with me, and you'll be able to explain Special Relativity at parties just like I can![5]

[5] Wait. That's not necessarily a selling point.

QS&BB Organization

I've organized QS&BB into four Parts.

0. **Tools** We'll use minimal mathematics and the next chapter stands alone as a refresher (and hopefully, a calming influence) for all that we'll need to follow QS&BB.[6]
1. **Physics and Cosmology of my Grandparent's Generation.** Before the turn of the 20th century, known physics included the well-confirmed physics of Newton's mechanics, optics, and the relatively new electromagnetism. These subjects form the language for all of the 20th and 21st centuries and are the individual points of departure for the revolutions to come. We'll need to establish our foundations in these subjects.[7]
2. **Physics and Cosmology of My Parent's Generation.** From 1900 through the 1950's everyone was becoming comfortable (as much as possible!) with the quantum mechanical and relativity theories...and their merging in Relativistic Quantum Field Theory. These subjects are our theories, and our models all respect their rules.[8]
3. **Physics and Cosmology of My Generation.** Since the discovery of the fact that the universe is filled with microwaves left over from the big bang and that two of the most different-looking theories are actually a part of a single story, we've been hard at work on puzzles that these discoveries create. This is our work today.[9]
4. **Physics and Cosmology of Your Generation.** We are intensely pursuing a number of observational puzzles and inspired and compelling theoretical ideas. We will look to the future.[10]

[6] Chapter 2.

[7] Chapters 3 through ??

[8] Chapters 14 through ??

[9] Chapters ?? through ??

[10] Chapters ?? through ??

Okay. I lied. Five parts, but the first one doesn't really count as an actual part.

The Nitty-Gritty of QS&BB

Here's how QS&BB is going to work. As you read through the book you'll see a number o f repeating features: Goals, Biography, Sides, Flags, Notebooks, Diagrammatica, and the Crank. Let's see what these each are.

Goals

The first section of every chapter will itemize three categories of goals that I hope you'll achieve. After completing each chapter, I hope you will:

- **Understand**. This will often mean some facility with a set of calculations and/or graphics interpretation. It means that you've followed a simple mathematical argument interactively (see Notebooks, below). For example to **Understand** a recipe means that you've prepared a meal using it. It doesn't mean that you created it.
- **Appreciate**. This is less quantitative than **Understanding**. To **Appreciate** a recipe you would realize that to sweeten it you'd add sugar, but not actually do it or even predict exactly how much.
- **Familiarize**. This is a fly-by of some story or feature of a bit of our physics story. To be **Familiar** means that you know to go to Mr Google for information, because you can't remember the details before that step. Continuing with the food analogy, you might be **Familiar** with the idea that recipes for chocolate cookies exist, but you'd need the web or a cookbook in order to **Appreciate** or **Understand** one.

Biography

I fear that you might think of physics as strange symbols and dry prose memorialized between the covers of big books and journals. But at its most basic, physics is about people. Scientists carry on daily tasks, most of which are routine. But every once in a while, *exceptional people* accomplish exceptional things—they see some phenomenon or interpret some idea differently from everyone else.[11] This is a stressful place to be! Our heroes—the ones in textbooks—pursue their visions sometimes at personal cost.

I've found that sometimes the content of the physics stays in students' memories because they associate it with the people, so rather than stick a little scientific biography in a sidebar like many books, I highlight the people. The second section of each chapter includes a story: "A Little Bit of Einstein" (or someone) will introduce you to someone you've heard of ("A Little Bit of Einstein,"A Little Bit of Newton," and so on) or someone maybe you've not ("A Little Bit of Huygens," "A Little Bit of Kepler," "A Little Bit of Dirac," and so on).

Although many of these folks are pretty special—and indeed some were a little odd—most were just everyday people with skills. That's most of us.[12] My colleagues and I have different skills, no fancier than those required of many other jobs. I'd muck up the preparation of a legal opinion and you wouldn't want me to treat you for an illness. Those are skills practiced by others. We're moms and dads, mow the yard, and fix dinner just like everyone else. But we have these heroes to whom we're professionally connected[13] our chapters will highlight them. I hope you enjoy this part of QS&BB.

[11] Everyone I work with is smart. But there have been some scary-smart people in the history of science and I'd like for you to meet many of them.

[12] Perhaps you're not surprised at my impatience with the "mad scientist" image. Marty McFly's friend, Doc Brown, is my least favorite example of a scientist.

[13] A fun exercise that all of us have played at some part in our lives is to trace our Ph.D. degree supervisor, to his or hers, and so on back in history. For example, mine was Lincoln Wolfenstein. His was Edward Teller, who came from Werner Heisenberg, who in turn came from Arnold Sommerfeld, who came from Ferdinand von Lindemann, who came from Felix Klein, who came from Julius Plücker, who came from Christian Ludwig Gerling, who came from Carl Friedrich Gauss who came from Johann Friedrich Pfaff who came from Johann Elert Bode who came from Johann Georg Büsch who came from Johann Andreas Segner who came from Georg Erhard Hamberger who came from Johann Adolph Wedel who came from Georg Wolfgang Wedel who came from... well, you get the idea.

Sides

Pay attention to what appears in the side margins. To your left are examples of the items that will appear regularly.. Footnotes[14] will be there, for easy reference. Side comments—sometimes even serious ones—will be placed in margin notes. Think of them as little, tiny essays. And there will be three kinds of named sidenotes: definitions, equations, and constants.

There's a lot of jargon in this business and so I'll call out words or phrases that you'll need to keep in mind for later use. Those will get the name definitions, just like dictionary.

There are also a handful of equations that will be useful and so when one of them appears in a margin, take it seriously. You'll need it. In fact, as you'll see below, I'm serious about taking notes and frankly copying the definitions and equations in a notebook, which you'll add to with each chapter, would be a good reference for you and an extremely important part of mentally processing what you write. So: write them for exercise and for safe-keeping.

[14] Here's a footnote.

Just a regular margin note here.

Definition: **Some word.**
Followed by the definition of that word.

Equation: **T-shirt equation.**
$E = mc^2$

Constant of nature: **A constant of nature..**
Gallon = 4.0 quarts.

Flags

Though our coverage is largely historical,[15] we'll come across ideas and concepts that will play various important roles as we move through the decades. I call these "flags" and they appear in the text, and then will be recalled at the back of each chapter so they will all be in one place. There are four kinds of flags:

[15] And hopefully, sometimes hysterical.

A concept is just what it sounds like: an important idea worth highlighting. *Key Concept 1*

An observation is an experimental fact of profound consequence. *Key Observation 1*

A question is just that: something that we need to understand. *Key Question 1*

Then there is a particle-flag. We'll be accumulating a number of particles as we go along and I will provide this table each time. For example, the electron was discovered in 1895 and the particle-flag for it will read:

Particle 1

Electron

symbol:	e				
charge:	$-1e$	mass:	$m_e = 0.511\ \mathrm{MeV/c^2}$	spin:	$1/2$
category	fermion			category	elementary

Notebooks

There is much of this account that I *don't* want you to "read" in the normal way. I want you to walk through the book—like the phone book—with your fingers doing the walking.[16] One thing I've learned over a few decades of teaching smart students who study subjects that are not mathematical (you?) is that if you come to the university as a freshman to major in, say Political Science or English or Psychology... that initial semester of college might be the first time in 13 previous school years in which you aren't taking a math course. At that point, after about a year away, you might find that your math muscle has atrophied. Trust me, I'm a doctor. You do have a math muscle and it needs periodic exercise to keep it fit.

I'm convinced that your brain is wired directly to your fingers.[17] Unless you've spent many years at this, you really can't *read* mathematics like you might read a history textbook: you have to interact with it. There is an enormous cognitive benefit from tactile reading: forming the symbols and numbers along with the text and allowing the logic to happen in your brain *by writing it out*. So this book will urge you to participate in the mathematical story-telling and I've got two ways for you to do it.

The Pencil.

The first way is by following along with your fingers: Buy a spiral-bound notebook into which you'll record your reading notes.[18] Then, when you're reading, you're using a pencil.

[16] There used to be this book. It had phone numbers, names, and addresses in it. The Phone Company's slogan was "let your fingers do the walking." This seems a century ago.

Just like I can't do 100 pushups any more—and I'd be pretty anxious if I were asked to do that in front of a class—I know that you might not be able to do some mathematics that you once were able to do! That's the famous "math anxiety."

[17] Or is it only my brain?

[18] Or your instructor might wish for you to use the template at the end of this chapter for your work. Notice that the "Pencil" has a number and that would be transferred to your paper.

When I get to a point in the text where need you to use that direct connection from your fingers to your brain, I'll indicate it with:

Pencil 0.1. ✎

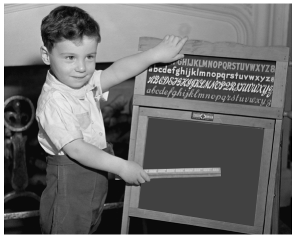
What will follow the pencil will be short sections of content where you need to drill down a little deeper than what just passively reading will do for you. To me that means, start recording detailed notes. In fact I'm happy if you even *just copy* the numbers and formulas and that will be good enough. It will still penetrate your brain...in a good way.

When it's done, I'll congratulate you with a thumbs-up and you can go back to just reading.

I guarantee you that if you don't do this and simply kick back and read without pencil in hand, what comes after will mean less. Further, I can guarantee you that if you *do do this*, the logic of the mathematics and the inevitability of the narrative will be escorted to your brain and be there when you need it later.

The Let's Do It

The second way is more active and requires you to actually write along with me. For example, I will sometimes come across an algebraic equation that needs to be manipulated a little or evaluated by plugging in numbers in order to keep going with the narrative. Or I'll have a graph that we need to look at for a specific number or an ordering exercise that will inform the narrative. When this happens, you'll see a little boy pointing at his blackboard and some instructions...your clue to get out your pencil again, but this time on your own for a minute. That tapping sound you'll hear is me waiting for you to write it out.



You Do It 0.1.
introduction/SolvingNewton1

This is an example of the kind of thing that you'll see: Newton's Gravitational Law is $F = G\frac{mM}{R^2}$. Please solve for G

The idea is that first *try it on your own*, then you can click on the blackboard and I'll walk you through it in a movie. Even if you simply copy what I write symbol by symbol, there's still a huge benefit to your understanding the physics. It will be in your brain, through your fingers. I *want* you to copy my work!

> **Wait.** *I know how to read. Do I really have to do this?*
>
> **Glad you asked.** *No, of course not. But if you can absorb what's coming without your pencil connecting to your brain then you're a lot smarter than I am. Take a chance. Write in your book. I won't tell.*

One more thing. The title under the picture includes a part of a directory. For this example, it's "introduction/SolvingNewton1." If you cannot click on the image, you can go to the movie by from the QR code in the margin (which provides the root https://qstbb.pa.msu.edu/storage/QSBB_WebManuscript/) and appending the final address location or directly by typing it all into a bookmark. The complete path to the movie is then the combination of the root (which is common throughout the book) and that portion in the title of each You Do It caption. For this one it would be:
https://qstbb.pa.msu.edu/storage/QSBB_WebManuscript/introduction/SolvingNewton1 .

Get it?

Figure 2: The QR code points to the root directory for movies.

Digrammatica

We will need many diagrams. Sometimes these will be graphs of characteristic physical quantities (like distance versus time). Sometimes, these will be diagrams of phenomena (like an electric field). Sometimes these will be iconic items that go together in useful ways, like Feynman Diagrams. Rather than interrupt the flow in the narrative, I'll follow that chapter of interest with a special kind of chapter which

I'll call *Diagrammatica*.[19] The contents of Diagrammatica chapters will be little more than a descriptive inventory of the diagrams of interest. Don't expect much lyrical prose in the Diagrammatica chapters. They're all business.

[19] The name is actually borrowed from a venerated little book on Feynman Diagrams by Nobel Laureate and University of Michigan Physics Professor, Martinus Veltman ("Tini"), *Diagrammatica: The Path to Feynman Diagrams (Cambridge Lecture Notes in Physics).*

Turning the Crank

Finally, in a course like this the emphasis is not on the details of calculation but on the conceptual ideas. But calculations do happen and I think we should be able to identify what goes into a particular calculation and what comes out. In Chapter 2 I'll talk a little bit about models and the scientific process. Every prediction includes the following components:

- A Hard Core of unquestioned assumptions, models, data, and so on. A modern publication in aerodynamics doesn't need to go back and justify the use of Newton's laws of motion. It's assumed to be correct. So there is always a Core.
- Sometimes a prediction requires mixing data with mathematics. So an input might include Data.
- Every prediction is a prediction of a model, sometimes as a test of the model and sometimes as a test of an experiment. So the primary input are the ingredients of a Model.
- Most calculations involve a strategy of how to proceed using the Core and the Model.
- Then, there is a result! A prediction can be purely mathematical (we'd say "theoretical") and so the calculation really is a test of the logical consistency of the Model (does it "hang together"). Usually though, we expect the outcome to predict the results of some measurement.

I know that you've all used the phrase "turn the crank." The assumption is that somewhere someone simply followed through with the rules of a mathematical calculation. Well, a crank is so 19th century! I'll repeatedly use a graphic of a nonsense circuit that uses a little fictitious microprocessor[20] which is doing the crank-turning. Figure 3 is my silly image which will emphasize the inputs, what's being tested, and the conclusion. We'll take it for granted that someone with the right expertise can turn that crank, just like a computer might. You'll see how this works in the next chapter and then in many to come.

[20] I've plopped on top of my nonsense circuit an FPGA (Field Programmable Gate Array) from Xilinx Corporation. This is their newest model, the UltraScale+™.

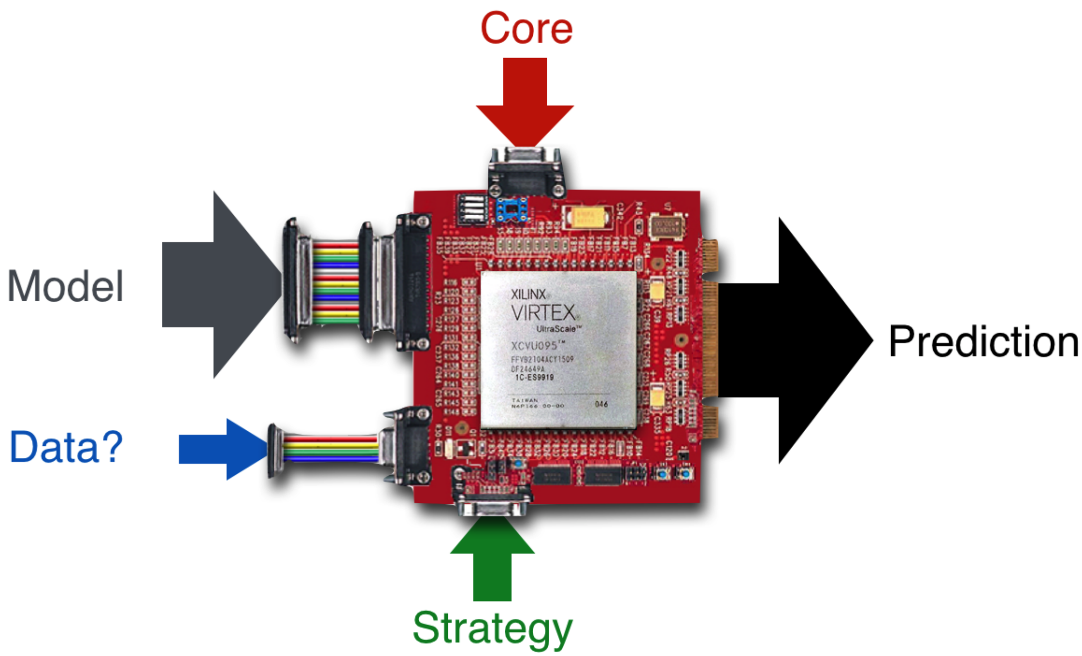


Figure 3: Our QS&BBcrank. The inputs are the Core, the Model, and sometimes Data. The outputs is some prediction. The Xilinx FPGA is essentially a little computer-on-a-chip used in many industrial and research applications, including those designed at MSU for our CERN ATLAS experiment.

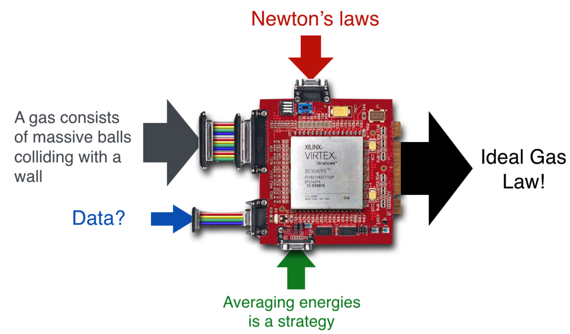


Figure 4: Newton's laws were not questioned, and so the Core. The Model was that a gas is a collection of tiny, massive balls that collided with the walls, and the strategy was to not treat each of them individually, but to average over their motions.

Here's an example. In the early 18th century Newton's ideas about momentum and mechanics were being tried out on various phenomena. Daniel Bernoulli, a part of the most dysfunctional scientific family in the history of physics[21] had the idea that maybe the pressure that gases exert on a container were a function of collisions that hypothetical gas molecules exert on the walls of the container. This idea was expanded on later and actually resulted in a new understanding that temperature is nothing more than the average kinetic energy of a gas. This explained Boyle's Law, which maybe you remember from high school. It says that $PV =$ constant. Figure 4 is how I would short-circuit the calculation that one would go through to reach this conclusion. Get it?

[21] Look them up! `http://www.daviddarling.info/encyclopedia/B/Bernoulli.html`

What I need from you is an open mind and your pencil. Work the examples, do the Pencil-and-Thumb fill-ins, and enjoy our exploration of Outer and Inner Space.

Let's go to work!

Chapter 1

Introduction

Studying the Smallest and the Largest

The Large Hadron Collider, looking south across Lake Geneva and the Swiss Alps

"In the beginning, the universe was created.
This has made a lot of people very angry and been widely regarded as a bad move." *Douglas Adams*

We're about to follow a Big Story —the "just so" story of the beginning of the universe. Yes, that one: Everything. The plot of this story seems to have all sorts of twists and turns that we're still unraveling. Surprises await.

Of course, the details are where the devil resides and they are fiercely complex. So much so that two entirely different scientific communities are currently deployed to battle with nature: those of us who work on the "outside" and those who explore the "inside." The outside crew are astronomers and astrophysicists. They measure and characterize the constituents and nature of the cosmos. They look *out.* The inside teams mimic the earliest picoseconds of the universe by recreating its incredibly hot, adolescent conditions in laboratories here on the Earth. These are the particle physicists and they look *in.* This is the story of both.

What's the smallest real thing that you can know about? For people of my *grandparents' generation*, the sophisticated answer would be "what you can see." I was born in the year 1950, and so my grandparents would have been children at the dawn of the 20th century which is when physics got interesting. They would have been taught that to claim existence for an object that the naked eye could not see was absurd. Chemists spoke of atoms, but were disdainful of anyone who thought they were real. They were just a shorthand picture for how to visualize elements. Physicists were even less flexible.

"Quarks"? "Leptons"? Lots of jargon and I'll keep it all straight for you as we go along. For now, quarks are itsy-bitsy pieces of the proton and leptons include the electron and others.

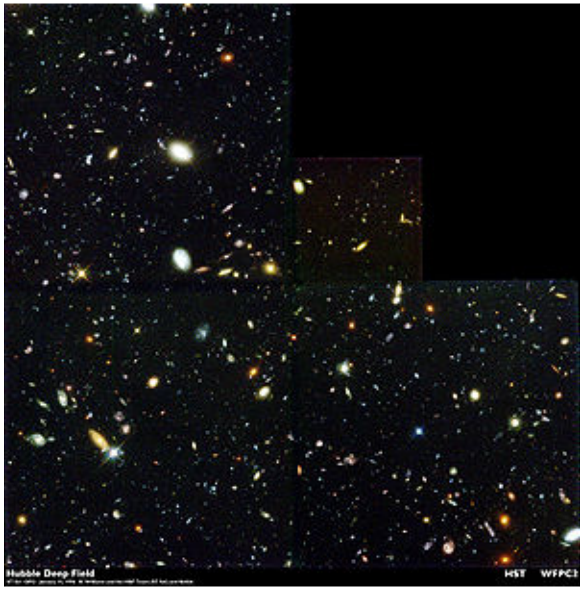


Figure 1.1: The so-called "Hubble Deep Field" view of a tiny spot in the sky, filled with 3,000 galaxies.

For people of my *parents' generation*, the answer would have been "protons, neutrons, and electrons." The atom had been thrust into believability around the turn of the century, and then refined during the next two decades. But the neat planet-like picture of the atom was where it all stopped for many.

In *our generation*, the answer to the "smallest" question has been "quarks and leptons" ...but we fully expect that they are not the end of the "smallest" story. We're hard at work, even as we speak with brand new tools to explore further than ever before.

In *your generation*? The sky's the limit! We've hints at solving some old puzzles and we'll undoubtedly find new ones. We're developing and deploying amazing new instruments and theoretical ideas now rub shoulders with not just nature, but philosophy and the deepest questions asked by humans. Your generation is going to see amazing things.

Through decades of intense experimentation and imaginative theorizing, the tiniest bits of reality are turning out to be a fascinating collection of objects. In the 1950s and 1960s, we just stood back and tried to catch the hundreds of particles that our experiments spit out at us. New particles every year! Names that nobody could remember. Hundreds of them, which was ludicrous! Didn't nature have some plan?

The good news is that we've uncovered a model that's a very good picture of how much of the fundamental particles of the universe work together and we've been exploring it since the 1970s. We've knitted that earlier mess together into a coherent picture of the entities themselves as well as the rules that govern how that stuff behaves.

But we're unhappy. Our grand synthesis of the Tiniest Bits Story—called the Standard Model — now looks a little shakey. While it's been the gold-standard of the successful scientific theory, we expect that *new* tiny bits are lurking in our experiments and we will be astonished if nothing shows up as we dig deeper.[1] This new anticipation would have been met with blank stares only a couple of decades ago.

So much for inside effort.[2]

Okay. So what's the biggest real thing you can know about? For people of my *grandparents' generation* the learned answer to this question would be "the size of the Milky Way," which they would have been taught constituted the whole universe. Everything visible in the night sky was thought to

[1] These new states of matter might be: "additional quarks, the Higgs Boson, Supersymmetric Particles, Weakly Interacting Massive Particles (WIMPs), Dark Matter particles. . . ," all famous candidates for future discovery. Of course whenever we get too cocky, nature plots to surprise us with something completely unexpected—more often than we'd like to admit! So, we're instinctively wary of being too sure of what's coming.

[2] See the frontispiece of this chapter!

be a part of one big, but still cozy cluster of stars which we see to be densest around the southern sky (from North America). Not only was my grandparents' universe compact, it was supposed to be permanent—static and unchanging— built of three kinds of objects: planets, stars, and clusters of stars. Stars twinkled, planets were steadfastly bright, and clusters of stars were fuzzy, indicative of their presumed distances from us. Sure, they all moved with regularity during each night and shifted slightly in a year, but the large scale structure of my grandparents' universe was simple: a nice, intimate, dependable universe.

For people of my *parents' generation*, the universe suddenly became huge. Those fuzzy clusters were found to be other galaxies outside of the Milky Way which are surprisingly far from us—we're not alone in our comfy galaxy. They were taught about thousands—we now know, billions—of others, of which the Milky Way is a relatively modest and ordinary example. But, the real shocker was the overthrow of the static universe of my grandparents era. My parents' universe was found to be flying apart—expanding—at a breakneck speed. No longer a tight-knit, stable thing...the universe is now huge and reckless.

The really unsettling piece of news for *my generation* is that the Big Questions of antiquity are now legitimate scientific research programs: Was there a beginning to the universe?[3] Are we alone? Will the universe end? Are there other universes? Was there anything *before* The Beginning? What drives the expansion of our universe to accelerate? The outside crowd thinks big thoughts now and this is a development of only the last couple of decades.

[3] There was a battle royal between two competing models of the universe in the 1950s. The first was dubbed by a proponent of the second, the "big bang"—not as a compliment. The second model was called the "Steady State" model. We'll talk more about these later. This battle raged until I was in high school.

When I was in graduate school, a professor told me that Cosmology was "physics knitting." Not any more! Cosmology in my and especially *your generation* is going to be flat-out amazing!

1.1 An Auspicious Beginning

Yes. The observable universe had a beginning, and quite a beginning it must have been: it was a roiling mess of radiation and elementary particles at temperatures never to be seen again. Everything that *is* would have been confined into a size smaller than the smallest particle we know of.[4] Unthinkably dense and with growth that was stunningly rapid, our early universe defies imagination. It's so outrageous that comprehending it seems a job for fiction and not science, yet my generation has also found ways to explore it: we probe it through direct telescope observations and we remake it in particle collisions. This is the blending of the outside with the inside pictures that motivates me.

[4] Maybe. Maybe not.

Wait. *I don't believe in the big bang. You appear to, but Isn't what you think just another "belief"? Aren't we each entitled to our own beliefs?*

Glad you asked. *"Believe" is a tricky word that we all use, although in our context, we should be clear. When I say "I believe in X," treat that as shorthand for the sentence: "X is highly confirmed by experiments and X is likely to survive foreseeable experimental tests." If I'm an expert in the field of X, then I have the obligation to describe those experimental tests. If I'm not an expert in X, I should expect that an expert could also enumerate its experimental successes in detail. There are dos and don'ts about this in science. About scientific belief, I can't do three things: 1) I can't say that I believe in X because I want to, 2) I can't say that I believe in X because my gut or a "feeling" tells me to, and 3) I can't say that I believe in X because a non-expert or an ancient text tells me to. Likewise, I can't say that I don't believe in X for any of those same three reasons. Stay with me. What I'll show you are amazing things and a record of success that's hard to ignore. Science is a process as well as a collection of theories and models!*

Quarks, Spacetime, and the Big Bang (which I'll affectionately refer to as "QS&BB") tells the interleaved stories of the two sciences of particle physics and Cosmology and how they have come to be blended together into a believable picture of how we all came to be. We're deep into the narrative—the plot is well understood, the characters are developed, and a "can't put it down" fever has set in. We're eager to see how it comes out and we're doing experiments all around the globe—and in orbit *above* the globe and in deep underground laboratories *inside* the globe— to push ourselves to the story's climax.

Definition: particle physics.
The study of the smallest bits of energy, matter, and the rules that govern their interactions.

1.2 The Inside Game: Particles and Forces

Sure, we've learned a lot in the last four decades about the Particle side of this story—my whole professional life. But, what's particularly interesting about this coming decade in "Elementary particle physics," (aka "EPP") is that we've reached an impasse. We have bushel baskets full of theories about what should come next, but we're starved for new data which will direct us on how to sort out the various theories. You and I are going to explore that situation because new data are coming in right now at extraordinary international laboratories. The coming decade is going to be interesting.

Definition: Cosmology.
The study of history and the future of the whole universe.

The inside story is that of EPP or as it's often called, just "particle physics," while the outside story is that of "Cosmology." We'll travel these narratives sequentially from their common beginnings.

The particle physics side is a well-established field practiced by about 10,000 of us in nearly every country and with major labs on four continents: North America, Europe, Asia, and Antarctica.[5] We build ac-

[5] Antarctica's a continent, right? Lots of experiments at the South Pole.

celerators to provide beams of electrons or protons and crash them together. We then collect the debris from those collisions in gigantic "detectors" that allow us to unravel the resulting debris.[6] Or we build detectors that are exposed to cosmic particles. EPP is one of many sub-disciplines in physics, but it's a little different. Urgent questions in most scientific areas have evolved and sometimes new specialties emerge.[7] In contrast, while particle physics has become specialized and sophisticated, its goals have always been intensely focused on two questions:

[6] Chapters **??**, **??**, and **??**.

[7] For example, Nuclear Physics and particle physics were practiced by the same people until the 1950s when they naturally split into two different subfields of physics. One group pursued the intricacies of more and more complex nuclei and the other pursued the complexities of the simplest objects. Each approach requires specialized devices and each separate theoretical tools.

What are the most elementary particles in nature? *Key Question 2*

What fundamental forces act among those elementary particles? *Key Question 3*

We think that getting closer and closer to answering them which will lead us to a deep understanding of the early universe. Paradoxically: understanding the tiniest things in nature will help to understand our "origins" which have been debated and argued for 2,500 years.

Box 1.1 A little philosophy

By the way, do you see how these two key questions are different? The first one asks about the existence of "things." An inventory. The second question asks about physical laws *among* the things. We're realists, which is to say, we think that things are real and that our theories are about real processes. These two ideas are still debated in philosophy and scientific realists would refer to them as "entity realism" and "theory realism." The former is more easily defended than the latter. But, we're not philosophers. We're scientists and we believe that the discovered laws of nature are factual statements about how things work. Enough of this.

These first two questions were stated carefully, so let's take them apart: "elementary particles," and "fundamental forces" are both specific concepts in my world that have different meanings from normal peoples' worlds! How about parts?

1.2.1 What's An "Elementary Particle"?

The most basic qualification for some entity of nature to be "elementary" is...that it has no parts. Most things have parts: stars, trees, molecules. Even an atom has parts—the nucleus, which is made up of protons and neutrons, and the atomic electrons.[8] The electron? No parts. It's elementary.[9] So, an atom is not elementary and not a subject of our investigations in particle physics.

[8] Chapter **??**

[9] So far.

An elementary particle is a bit of matter and energy that has no constituent parts. *Key Concept 2*

Elementary ≡ no parts is a simple idea.[10] But, as you'll see, a persistent theme of this book is to emphasize how simple ideas about nature can become wonderfully complex.

[10] The symbol ≡ means "defined as" or "equivalent to."

"Here's your moment of zen": "Simple means complex"!

But what a particle is...well, that's actually complicated. We learned in the last century that particles aren't the nitty-gritty of reality because when we combine the theory of quantum mechanics with the theory of relativity, we find that stuff in atoms, nuclei, people, electrons and stars—everything— is actually the consequence of a set of continuous, wiggling *quantum fields*.[11]

[11] Chapters 12, **??**, **??**, and **??**, and **??**

Wait. *Fields bring to mind something that's spread out, but atoms are individual things. How are they related?*

Glad you asked. *That's right. Fields are indeed spread out. Imagine a wheat field waving in a summer breeze, as far as you can see. Fields imply waves and collective motions of quantum fields are not unlike this image. But if a field is the fundamental substance of everything, then that's indeed unsettling since a field is everywhere, but a particle that comes from it is "there." We'll talk a lot about fields and try to reconcile these two pictures. Be patient.*

In spite of this field-reality, I have to admit to the mental crutch of *particles*. For EPP it's easiest to mostly use the mathematical language of particles and that language came to us from Richard Feynman.[12] So, one side of my brain is full of the sophisticated symbols and manipulations of the relativistic quantum field theory that precisely describes this stuff. But the other side of my head is full of images of billiard balls bouncing off of one another: colliding particles.[13] In any case, QS&BB will cover this growing awareness of the importance of fields and how particles make themselves known.

[12] Chapters 6.5, **??**, **??**, **??**, and **??**

[13] By the way, It's not an entirely satisfying picture since in order for this analogy to be precise, my mental quantum billiard balls should also randomly decay into other billiard balls—or into baseballs or bananas,— should pass right through other billiard balls, and even spontaneously leave my pool table and appear on someone else's! But, we have to cling to some picture in our heads and that's mine.

1.2.2 What Is a Fundamental Force?

"Force" is one of those words that has many colloquial meanings.[14] But in physics a force is a precise concept—a noun and not just a verb. Here's the simplest notion of a force, which came from Newton and with embellishments, still works today: if you alter the motion of something, you exerted a force on it.

[14] "You can't force me to eat that!" (You thought I would refer to Star Wars, but I'm better than that.)

Everyday Forces

You and I deal with three kinds of forces every day. Let's talk a little about all three...and then how the forces in particle physics are different from these.

First, take regular pushing—whether it's a push that's through muscles against something, or the push of a tire (or your shoe) against the road—this mechanical thing-to-thing contact seems instinctively to be direct. Solid against solid. You might be satisfied with the phrase "mechanical force" as all you need to say and you'd be consistent with its modern usage in engineering. Write-in this kind of force in pencil for now.

Then of course dropping things can be a bad habit, but whatever you mishandle, it always go to the floor. On the street we'd say that gravity pulls things down to the Earth. But here again, the actual situation is quite a bit more complicated than "gravity made me do it." QS&BB will take us through many of the stories about gravity and you'll be amazed at how that idea has changed. We think Mr. Einstein's picture is closest...right now.

Finally, what about a magnet? Surely at one point in your young life, you've played with a pair of magnets and marveled at the fact that they seem to "communicate" with one another. Without touching, and without any obvious connection between them, a force is transmitted through thin air. Hand, here's another one that doesn't need direct contact to alter motion: your hair's state is affected on a cold, dry day by a comb—your 'do rearranges itself as if by magic without actually touching the apparent cause of the hair motion—a statically charged comb.

One of the neat stories we'll uncover is that the relationship between your hair's unruliness in January and your dog's photograph sticking to your refrigerator is an intimate one: they are both examples of a single force, the "electromagnetic force" and understanding that will take us into Albert Einstein's young life.[15]

[15] Chapters 12, ??, and 14

Here's a well-kept secret: the mechanical thing-on-thing pushes and pulls of everyday life are actually electromagnetic: the reason your hand doesn't go right through the box you're pushing is because the electrons in your hand are repelled by the electrons in the surface atoms of the box and so you. . . push it.[16]

[16] The reason you don't pass right through the floor is due to the same electrostatic force.

So that first kind of everyday force that I warned you write in pencil? You can erase that now. We only deal in two forces in our human-sized, everyday lives: gravity and electromagnetism.

1.2.3 Particles, Forces, and An Amazing Theory

Enumerating particles is like physics-stamp-collecting: find them and sort them for similarities and differences. But that leaves out much of the story, since we need to know how these particles interact with one another. We need to know about the *forces among them.*[17] These sorts of forces are special and abstract.

[17] Chapter ??

Nature is pretty economical. If there are 12 kinds of particles in the universe, you might guess that maybe there's one force, or 6 or 12. But, it turns out that there appear to be only 4. In our everyday lives we encounter half of them. The others act behind the scenes.

The forces are different from one another in two ways. First, not all forces "see" all particles. An electric force only notices electric charge, so anything that's neutral (like a neutron...or your body) will not be yanked around by the presence of an electric field. The second way they're different is their strengths. Your dog's photo stays on the refrigerator and doesn't fall to the floor because the force of gravity is very much weaker than the force of electromagnetism.

Besides electromagnetism and gravity, the other two forces are called, get ready, the Weak Force and the Strong Force. They are, as you might guess, weaker and stronger than some others.[18] We'll talk a lot more about these later, but from weakest to strongest, the forces order themselves:

[18] We have a lot of fun naming things in particle physics.

1. the Gravitational Force,
2. the Weak Force,
3. the Electromagnetic Force, and
4. the Strong Force.

The role of QS&BB? Describing how we learned there were four forces and how they function in the universe.

1.2.4 The Standard Model of Particle Physics

One of the amazing accomplishments of the last three decades in particle physics is that we now rely on a theory, called colloquially "the standard model" that predicted the existence of new particles (which we found) but also explains how 3 of the 4 the forces of nature originated.[19]The standard model is like no other in the history of physics as it pretty much accounts for everything.[20]

It describes all known elementary particle interactions—on Earth, in cosmic rays, supernova explosions, and the earliest moments of the universe. It also is the mathematical blueprint for how atoms are held together, how nuclei bind, how molecules are constructed: it's the scientific platform on which all of physical science stands.[21]

But there's more: the Standard Model tells a story of the big bang that shows that forces aren't forever. The forces we know now, were born out of entirely different forces as the universe cooled. Even *mass* didn't exist as a concept until these forces changed. It's quite a remarkable intellectual achievement, this Standard Model. I'll construct it for you to admire along with my colleagues and me!

[19] Gravity shuns the Standard Model because of its quantum mechanical roots. Gravity stands alone and that bothers everyone.

[20] In fact, it's so special to us that I'm going to capitalize it. From now on: **Standard Model.** Rolls right off your tongue.

[21] Of course the complexity of nuclei, chemicals, organic and inorganic molecules is very specialized and so the sciences of Nuclear Physics, Chemistry, and Biology are themselves quite complicated and beautiful. But the basic laws that are at work deep down, are those of the Standard Model.

1.2.5 Particle Confusions

Our Standard Model now has no missing pieces. The last bit was revealed in 2012 with the announcement that we had found a strange particle called the "Higgs Boson" in our experiments at the Large Hadron Collider at CERN. But we're still not happy for two reasons. First, there are experimental reasons: something's going on in the universe that causes galaxies to move oddly (see below) and something's going on with nothing— the vacuum, which we tend to think of as related to that theory of fields that I described above. The second reason we're not happy is the Standard Model has some formal features about it that don't quite sit right with us. The mathematical instructions that come with the Standard Model require us to do an odd thing to get it to work, and we're pretty sure that this odd thing should have a deeper purpose and not be as *ad-hoc* as it seems.[22]Let's go large.

[22] Want to know what that odd thing is? We take an equation, and we change the sign of one piece from negative to positive. No particular reason. . . except that it works. Stay tuned, you'll see.

1.3 The Outside Game: The Big Bang

As I've indicated, the big news of the 20th century is that our cosmos had a birthday. Astrophysicists have made huge strides in the last three decades with amazing instruments operated on Earth and launched into orbit around the Earth. Results keep pouring in: our universe had a Beginning.

Definition: Astrophysics.
The study of the dynamics and the origins of astronomical objects.

Stand back and think about the implications: this is the most remarkable scientific discovery in history. Of all of the ways people have thought about their place in the world, over thousands of years there was only speculation and myth about a possible Beginning. After decades of patient research, we now know: there was a time—before which there was nothing. Suddenly, in the blink of an instant, space, time, and the energy of matter and radiation were born and then the whole mess cooled eventually. Evolving Into suns, planets, and us.

Some would call this later version, Physical Cosmology in order to distinguish it from the precursor story-telling. (I'm looking at you, Wikipedia.) But we'll just call it plain, old Cosmology.

From the creation stories to the "just-so fables," humankind used mythology and belief to orient itself with the universe they could see. Even then, there was the strong sense that the whole of the universe was bigger than what humans could imagine. Cosmology is an old, old subject, but it only became a *science* in the last century.

Well, we don't just "imagine" any more. We measure. Cosmology is a new science; it became one in the hands of Albert Einstein in the early twentieth century. Things didn't go quite as he'd planned, as we'll see. But he laid the groundwork for a human-based study of the universe using mathematical rules rather than mythology or belief. Today it's among the most exciting branches of all of physics.

The two basic questions that modern cosmology tries to understand the answers to are these:

What are the past and future histories of the universe? *Key Question 4*

What are the ingredients of the universe? *Key Question 5*

These questions area alsoo carefully stated. So let's unpack "history of the universe" and "ingredients."

1.3.1 Histories of the Universe?

You know the meaning of "Universe," right? It's...well...it's everything. At least that's what it used to mean. We'll consider a growing suspicion is that a *universe* might be a relatively local object and that there might be room for an interpretation of the whole cosmos that could incorporate other *universes*.[23]

[23] Now, did you ever think that there could be a plural of that word?

Perhaps you've read about a "multiverse," a speculative idea in which ours is just one of an infinite number of universes which are born and die spontaneously and for eternity. All of them would have different physical laws and so different particles and varying potential for life. So that's one side of a contentious argument in physics. To opponents, the multiverse is speculation that's beyond wild.[24] In QS&BB we'll talk about why the multiverse is a topic for science seminars and not just comic books. On this, we'll be agnostic. Just the facts, ma'am.

[24] For some, even reckless and unscientific.

Let's try to define what our universe would entail. Our universe is

1. the one in which we (or our original elements) reside,
2. the one where the same physical laws work throughout, and
3. the one that had the big bang that our evidence points to..[25]

[25] It ain't much, but it's home.

Certainly, the past history of the universe is the hot[26] topic in all of cosmology.

[26] No pun intended.

Past History

Our inference to the need for a beginning—a big bang—comes from a) the fact that the universe is expanding, b) that we therefore infer that it was smaller in the past, and importantly, c) that we have a plausible, predictive model that describes this situation. Both the fact of the big bang and the stories that led us to this conclusion are fascinating and QS&BB spends quite a bit of time unraveling them.[27] But just how this happened is a matter of urgent research.

[27] Chapters **??**, **??**, and **??**

We can play the universe-movie-camera backwards in our models to what we call the big bang. In the conventional model of cosmology we can reliably predict[28] the times at which atoms were formed, then

[28] post-dict?

back to when nuclei would have formed, and then further back to when protons and neutrons would have been formed. At that point the universe would have been unbelievably hot and dense and only consisted of the most elementary of particles. This birthday of matter is about a picosecond after the big bang: when the universe was about

$$0.000000000001 \text{ seconds old.}$$

Let's call this the "Electroweak time," which we'll study later.[29] We believe we have a good explanation for the universe's evolution from that tiny fraction of a second to the present 13.8B years,

$$430,000,000,000,000,000 \text{ seconds.}$$

.

As tiny as the Electroweak time is, it's still not

$$0 \text{ seconds!}$$

In fact, we don't know how to describe zero since there's a limit, before which, if we keep pushing our models, mathematics fails us with infinities. That's called the "Planck time," 10^{-43} seconds, or...dare I write it?...

$$0.0000000000000000000000000000000000000000001 \text{ seconds.}$$

This is a defining point. Before the Planck time, the very concept of "time" would not exist. But between the Planck and Electroweak times—between the point where we can make mathematical models to the time where we think we start to understand? A lot had to happen in that tiny, tiny time window.

We've hypotheses of what might have happened in that instant and tests we can perform to lend credence to them. [30] So the past history of the universe is an active area of research, world-wide. Theory and experiment in astrophysics and particle physics all work together on this. The good news is that from the Electroweak time forward until now, we can explain how just about everything evolved. The bad news is that before that point, we come up against the 800 pound gorilla-question: **what banged** in the big bang?

Now put on your seatbelt. Could there have been a "before the big bang"? The general consensus is "yes" and the front-runner model—what some have called "an amendment to the big bang"—is called *Inflation*. This 30 year old idea predicts that at about 10^{-35} seconds (not yet as early as the Planck time) the universe went through a phase transition, not unlike when water boils. Before that point, there was only the vacuum...a bubble of nothing. After the inflationary event, radiation and particles were created and our universe evolved until today.

[29] Chapter ??

[30] To complicate matters, space and time are intimately involved in this event in ways that we can't yet rigorously pin down.

This going from *nothing* to *something*—dubbed the “ultimate free lunch” by inflation’s inventor—is heady stuff. But it’s testable stuff. And it’s bizarre stuff since inflation is part of the inspiration for the far-out notion that ours is only one of a “multiverse.” This hypothesized infinite collection of other vacuum bubbles would be eternal (time wouldn’t exist) and would be spawning other universes for all eternity. Some might become full-fledged universes with particle and laws amenable to making stars, galaxies, and carbon-based life. Some might not.

Future History

So having teased you with our past, what is our future history? Well, I’m playing a word-game with you since we’ll see that in physics the direction of time becomes a different sort of thing than our regular use of that word. But the eventual fate of the universe has been a matter of mathematical modeling since the 1960s. The universe could logically

- expand forever;
- stop, shrink, and collapse; or
- slow down and become static.

Nobody was prepared for the surprise of 1999.

The results of determining the distances to a particular type of supernovae led to the conclusion that not only is the universe expanding, but that expansion appears to be *accelerating*. Something seems to be pushing space to stretch faster and faster and we’re not sure what it is. Taken at face-value, the future seems grim for this universe. At some point the expansion will be so fast that light would not be quick enough to be able to travel from one galaxy or star to another. Every celestial object will become isolated. Anyone left alive on any planet in this universe would see only... **black**. It will be a lonely place.

Another future history comes from competitors for whom after the universe’s birth and then Big Bang-ish evolution would lead to a contraction of space, all the way to an eventual collapse (the “Big Crunch”). And then the whole process would start over: the universe would be cyclic. An endless repetition of groundhog day cosmic repeats. In this scenario there is no unique beginning, but rather an endless series of beginnings.[31]

So you can see that while the knowing the past and future of the universe are age-old quests their unraveling might be puzzles that humans can actually solve. Our two most compelling models are physically different and even *philosophically different*! Inflation assumes that time had a beginning, while in the cyclic picture time is perpetual—it never starts and never ends. Appreciating the details of these and other advances are a part of the QS&BB mission.

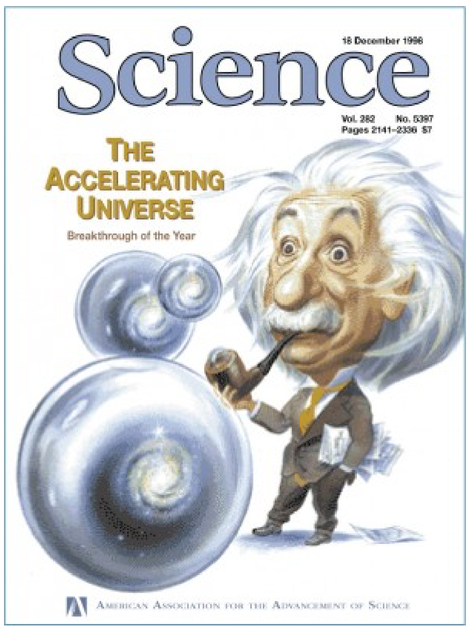


Figure 1.2: sciencemag

[31] This model is also consistent with the accelerating universe, but ascribes the cause differently from inflation.

1.3.2 Ingredients in Our Universe?

In order to inventory the ingredients of your world, you just look around you. Houses, clouds, Earth, the Moon, the Sun, stars, and so on. But the ingredients that I'm speaking of are cosmic. The universe is incredibly big—and we'll get a sense of that—but the smoothed out average amount of actual stuff is actually quite small, not much more than about 3 protons per cubic meter. So the overall density of the universe is minuscule, pretty smooth, and pretty much dominated by hydrogen atoms. So cosmic ingredient number one? The simplest element of all. All of interstellar and intergalactic hydrogen was born out of the big bang. All of the other elements[32] are made in stars.

[32] except for tiny traces of helium and lithium

An inventory of the other cosmic ingredients beyond hydrogen depends on the epoch in which we make the list. During our current era, our accounting would include stars like the Sun, planets and exo-planets,[33] galaxies, a few spectacularly destructive stars (supernovae), and some stellar and galactic black holes. In an era thirteen billion years ago, galaxies wouldn't have been on the scene (but there would still be lots of hydrogen) and thirteen and a half billion years ago, there would have only been particles and radiation (hydrogen wouldn't exist yet). At about 300,000 years after the big bang (about 13.5 B years ago) the universe shined and then cooled and we're now surrounded by a measurable remnant afterglow (called the Cosmic Microwave Background, or CMB) just above the frequency of your microwave oven and studying it has been the mission of a number of famous satellite experiments.

[33] These are planets that are in orbit around other stars.

So understanding the evolution of the ingredients of the universe is an important undertaking, backed up with very sophisticated computer modeling and very precise satellite observatories. That 13.5 B year CMB mark is about the limit of our astronomical looking-back. Understanding the ingredients of earlier times requires a new partnership.

Because...the cosmic ingredients around the time of the Electroweak time would have been just the most elementary of elementary particles. Some we know about, others would have been different and evolved into our familiar set, and still others are only now hypothetical but discoverable in our experiments. I hope it's obvious by now that QS&BB will be focused on how our well-known particle-ingredients influenced this early time, but also what additional kinds might be found in our coming particle physics experiments.

Figure 1.3: vacuum

Beyond particles, galaxies, stars, and other normal things, we're confused by some very exotic cosmic ingredients. For example whatever it is that has grabbed a-hold of galaxies to make them rotate way differently from how we expect them to. Their motions suggest that they're (we're!) surrounded by unseen (not shining) stuff that gravitates but doesn't radiate: Dark Matter is our intriguing name for this stuff.

Finally, the most fascinating ingredient of the universe seems to be nothing. That is, the unseen force that seems to be pushing everything into that newly discovered accelerated expansion, might be a feature of the vacuum.[34] When we don't know what something is, we name it! "Dark Energy" is the placeholder name for the mysterious something that also is a target of frantic experiments and theoretical work.

[34] We'll talk a lot about the vacuum, which until this discovery was the province of particle physics. Now both cosmology and particle physics intellectually own nothing!

1.3.3 Cosmological Confusions

In Cosmology we face some flat-out observational or *experimental* embarrassments. For example, when we add up all of the mass-energy of all of the objects that we can see using all of our observational tools (optical telescopes, infrared telescopes, microwave satellite telescopes, radio telescopes, etc.), 95% of the mass of the universe is missing. No kidding.

A part of the missing stuff appears to be that Dark Matter ingredient (about 30%) and the rest seems to be made up of the mysterious Dark Energy ingredient. When you take the paltry 5% of shining stuff and add in these two Dark ingredients, it actually works out to 100%! This is a major victory for the "standard model of cosmology" or the "hot big bang model" (two names) and getting there is a part of the QS&BB story.

But we're confused about what Dark Matter and Dark Energy actually are. Embarrassed even. So there are major programs all over the Earth to study them.

Want something even stranger? Where are the antimatter galaxies? We don't see any evidence of relic antimatter in the universe. Only matter—the stuff we're made of. So either the universe began with an artificially enhanced matter dominance—an "initial condition" that is not scientifically acceptable—or at some point the originally *symmetric* matter-antimatter soup became our *antisymmetric*, matter-dominant outcome. And the list of puzzles goes on. Let's now play together.

1.4 Particle Physics and Cosmology, Together

After 50 years of successes and surprises in both fields, one thing is clear: the reality of a big bang means that there was a period when the universe consisted of only particles and forces. No protons, atoms, stars, galaxies, or Starbucks. Just elementary particles and the forces among them.

As I noted, that epoch was less than 0.000001 seconds long, but it was critical since the particles and forces were created just before it and what happened after was determined *by* the ingredients and rules of that period. What's more, we suspect that the set of forces *then* was different from those we know of *now* and that the set of primordial elementary particles might have included whole species that we've not yet found in terrestrial experiments.[35]

[35] As a tantalizing tease each of the cosmological problems above has candidate particle physics solutions!

These eras are not connected by a single story thread—yet. But they must be! No physicist thinks that the universe is governed by two contradicting sets of rules. So we have a lofty goal: we're working toward a model of *everything* about the universe from the big bang through to today. Theories abound, but experiment will decide. We can explore the earliest moments of the universe with the most powerful telescopes, but in order to investigate the times earlier than about 3 minutes after that Beginning, we need to do experiments in laboratories on our Earth. It's a bold extrapolation: by colliding protons head-on at very high energies, we're reproducing that early hot cosmic cauldron.

> **Wait.** *How do you know that this is the right connection to make? Maybe the conditions in the big bang were totally different than those in proton collisions?*
>
> **Glad you asked.** *It's a plausible story, and, frankly a nice one. But as pleasing as it is, we have to test it and what's neat about the state of affairs right now is that particle physicists are joining astrophysics collaborations and astrophysical measurements are directly testable in our labs on Earth. This approach could be wrong! But we have to pursue it with a vengeance since the stakes are so high.*

We're currently mounting experiments in both EPP and Cosmology that are going to hit these issues squarely in the next couple of decades. Their results will completely change the way we think. Textbooks will be rewritten. If the first 40 years of the twentieth century were wacky, the first couple of decades of the twenty first are likely to be amazing.

In my professional lifetime, these two fields have become kin. Theoretical and experimental advances (or surprises) in one field directly affect the other and visa versa.

That said, the stakes are so high that we can add a third focus for EPP:

How did elementary particles and their forces affect the evolution of the universe? *Key Question 6*

Like the ancient Ouroboros, the snake eating its own tail. Cosmology—the science of the biggest— is dependent on the science of the smallest, particle physics, and *vice versa*. That's our story: Elementary particle physics and Cosmology are now united in a single path of discovery and this book will show you how.

QS&BB is not old "dead white guy physics"! It's all new and the details are still being worked out so we're going to be talking about matters of very current interest. If you make it through with me, you'll be in a good position to appreciate the surprises when they start to occur at the Large Hadron Collider, Fermilab's LBNF and DUNE, Mu2e, g-2, numerous underground laboratories, as well as the Planck Explorer, James Webb Telescope, the Fermi Gamma-ray Space Telescope, and other space-based laboratories. They'll be in the newspaper (if we still have newspapers). You wait.

Figure 1.4: Ouroboros

Chapter 2
Everyone Needs Tools

A little math


René Descartes by Franz Hals, circa 1649

René Descartes (1596-1650)

"When I imagine a triangle, even though such a figure may exist nowhere in the world except in my thought, indeed may never have existed, there is nonetheless a certain nature or form, or particular essence, of this figure that is immutable and eternal, which I did not invent, and which in no way depends on my mind." *Meditations on First Philosophy* (1641)

It's always amazing to me, just how much we depend on the collaborative work of a handful of people from the 1600s. There must have been something in the water....in France, Italy, Britain, and Holland because this was a time of genius and courage. From people in this period—a number of whom we'll become familiar with—we received a way of thinking about, talking about, and poking at the world. René Descartes is one of my particular favorites. Let's learn a little bit about him.

2.1 Goals of this chapter:

- Understand:
  - Simple one-variable algebra.
  - Exponential notation.
  - Scientific notation.
  - Unit conversion.
  - Graphical vector addition and subtraction.
- Appreciate:
  - The approximation of complicated functions in an expansion.
- Be familiar with:
  - Descartes' life.
  - The importance of Descartes' merging of algebra and geometry.

2.2 A Little Bit of Descartes

The 17th century and just before saw a proliferation of "Fathers of –" figures: Galileo, the Father of Physics; Kepler, arguably the Father of Astrophysics, and Tycho Brahe, the Father of Astronomy. But the Granddaddy...um...Father was René Descartes (1596-1650), generally considered to be the Father of Western Philosophy and a Father of Mathematics.[1] If you've ever plotted a point in a coordinate system, you've paid homage to Descartes. If you've ever plotted a function, you've paid homage to Descartes. If you've ever looked at a rainbow? Yes. Him again. If you ever felt that the mind and the body are perhaps two different things, then you're paying homage to Descartes and if you were taught to be skeptical of authority and to work things out for yourself? Descartes. But above all—for us—René Descartes was the Father of analytic geometry.

He was born in 1596 in a little French village now called, Descartes.[2] By this time Galileo was a professor in Padua inventing physics and Caravaggio was in Rome inventing the Baroque. Across the Channel Shakespeare was in London inventing theater and Elizabeth had cracked the Royal Glass Ceiling and was reinventing moderate rule in England. This was a time of discovery and dangerous opinion when intellectuals began to think for themselves. That is, this is the beginning of the end of Aristotle's suffocating domination as The Authority on everything.[3]

[1] Who's your daddy, indeed.

[2] Coincidence? What do you think.

[3] After all, by the time St. Thomas absorbed Aristotle into Catholic dogma, he was called The Philosopher.

Descartes' mother died soon after childbirth when he was only a year old and he was raised by relatives. His' father was an upper-middle class lawyer who spent little time with his children.[4] He was sent to a prominent Jesuit school at the age of 10 and only a decade later emerged from the University of Poitiers with the family-expected law degree. Apart from his success in school, the most remarkable learned skill was his lifelong manner of studying. He was sickly as a child and had been allowed to spend his mornings in bed, a habit he retained until the last year of his life.[5]

One of the benefits of his schooling was a program to improve his physical conditioning, enough so that he became a proficient swordsman and soldier—he wore a sword throughout his life as befitting his status as a "gentleman."[6] And yes, he was essentially a soldier of fortune. During the decade following his graduation, he would alternate his time between combat assignments in various of the innumerable Thirty Year's War armies and raucous partying in Paris with friends.[7]

Somewhere in that period Descartes became serious and decided that he had important things to say. He wrote a handful of unpublished books and maintained a steady correspondence with intellectuals in Europe, becoming well-known through these letters. Catholic France and of course Italy, were becoming intolerant of challenges to Church doctrine and he moved to the relatively casual Netherlands in 1628. Mostly a good move: he'd been inspired by Galileo's telescopic discoveries and became a committed Copernican and in 1633 was completely spooked by the Italian's troubles with the Inquisition.[8] However, he had trouble with some evangelical protestant leaders in Holland.

Little did Descartes know that he was a mathematical genius. After study as a "mature" student at the University of Leiden, he found that he could solve problems in geometry that others could not. His devotion to mathematics and especially the rigor of the deductive method stayed with him and turned him into a new kind of philosopher. The logic of deduction and the certainty of mathematical demonstration were his philosophical touchstones.

Remember "deduction"? All squirrels are brown; that animal is a squirrel; therefore, that animal is brown kind of arguments? The important thing about this string of phrases is not that animal's color, but that the conclusion *cannot be doubted* if the two premises are true. Since Plato, "What can I know for sure?" was an essential question. For that particular Greek, things learned through your senses are untrustworthy. Only things you can trust are ideas which are eternal, outside of space and time. For other famous Greeks, you learn about the world through careful observation. Famously, Descartes convinced himself that he had discovered a method to truth: whatever cannot be logically doubted, is true.

[4] When Descartes' father died, his brother failed to notify him (he found out through one of his correspondents) and he decided he was too busy to attend the funeral. Not exactly a close family. The similarities with Newton's childhood are striking.

[5] There's a story there...

[6] He still worked in bed every morning until noon.

[7] He was a talented gambler, as befitting a mathematical mind.

[8] That year, one of his major books, *The World*, was ready for publication, but he delayed it until after his death. In *World*, he expounded Copernicanism, but also provided for a reason why the planets circled the sun. A mechanism that Newton demolished with gusto.

2.2.1 Descartes' Philosophy

This is not the place to teach the huge subject of Descartes' philosophy. But there are two aspects of his work that directly influence the development of physics: what can we know and what is the nature of the natural world.

Descartes believed he'd found the formula for determining what's true: when an idea is clear and distinct, which means incapable of being doubted, then you can believe it. His method was to keep doubting *everything* until you reach a point in this thought-process that can't be doubted.[9] The point he reached was the recognition that *he* was doing the doubting. Since that can't be doubted, then what he's learned that's true is: thought exists. One more step to *I exist*, because it is I who is doing that thinking: "Cogito ergo sum"[10] was his bumper sticker for truth.

[9] In this way you reduce a complex problem to a more manageable one. . . one of his essential components to his "analytic philosophy."

[10] "I think, therefore I am." Words to live by.

The rest of his argument is a little shaky but this is the beginning of dispassionately and vigorously analyzing a philosophical problem, setting a high bar for argument. Of course, Medieval thinking was not friendly to the idea that everything can be doubted. The Bible and pretty much all that Aristotle wrote was off-limits. In fact, under the rules of thought not only could neither source be doubted, those sources were the only authority used to determine truth and falsity. Descartes pretty much changed that in philosophy.

He called his method "analytic" and it's essentially applying mathematical problem solving strategies to philosophical questions. Hence, history's assignment of paternity to him for Western Philosophy.

He said later that he made this discovery about doubt while still a soldier and holed up on a snowy night alone in a remote cabin. Sometimes his military escapades were real combat, but mostly it seems like he had a lot of leisure time.

For our purposes, what he decided were that true things about the world could be obtained through pure thought. This is the "Rationalist" philosophy of which he is the king. This is in the spirit of Plato, but unlike Descartes, he gave up on the sensible world as simply a bad copy of the Real World, which is one of Ideas..."out there" somewhere. By contrast, Descartes asserted that there are two substances in the universe. One is mind and the other is matter. Understanding the universe means gaining knowledge of both by blending thinking with observing.

We'll see that physics takes some inspiration through Descartes' approach. Theoretical physicists are often motivated by knowledge gained through thought—and always mathematics—and many work as if those thoughts are representing the world.

Definition: Rationalism.
The only test of and source of knowledge is reason.

Definition: Empiricism.
All knowledge originates in experience—through experiment and observation.

This two-part universe is now called Cartesian Dualism and was all the rage when Newton was a student. But the important thing to take away from this is that Descartes is the proud proponent of the notion that true knowledge can be obtained purely through thought. The counter to this Rationalist belief is Empiricist belief, that knowledge can only be obtained through observation (and in modern form, experiment).

The other aspect of Descartes' philosophy that matters[11] is his notion of Mechanism. The Renaissance was saturated with ideas of nature that we'd consider magic. Nature was infused with occult properties,

[11] no pun intended. . . sort of.

that it is almost alive with "active principles," even human-like in ways. Of course, astrology, alchemy, signs and numerology, Cabala, black magic and white natural magic, and so on were aspects of organized occultism. But it went deeper. People lived lives, tended the sick, and found explanations for natural phenomena based on the assumption that what we would call inert natural objects were alive and possessed magical powers. This continued a long-standing philosophical discussion about Qualities. Is the boiling pot hot because it possess the innate quality of "hotness"?

Magical thinking was a threat to the Church and Descartes also subscribed to the growing program of ridding nature of these features. Things in the world are not possessed of innate features like hot or cold, blue or red, and so on. These for Descartes are attributes not innate qualities. "Things" possess...place. Now we'll think a bit later about what constitutes space, but for Descartes and others, space is determined by the extent of objects. In fact the only aspects of matter that are "clear and distinct" (and hence true) are that matter has the properties of spatial extent (length, width, height) and motion.

He needed to have a mechanism to explain everything in the material world. He explained motion as the point-to-point pushing of material objects that we see (planets) by innumerable, small-sized, varied atoms which are indivisible. This "plenum" of stuff is moving, initiated by God, and they preserve that motion as they transmit it to all moving material objects.[12] It's communicated to the planets, through vortices, as in Fig. 2.1 from *The World*.

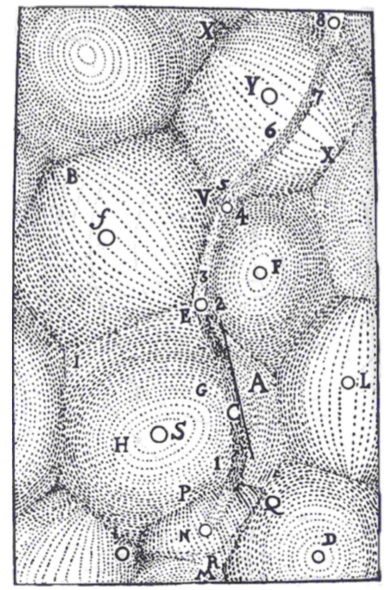


Figure 2.1: plenum

[12] Remember this when we get to momentum and energy!

Likewise magnetism. Boy, that's an occult-ish phenomenon if there ever was one. To Descartes magnetism was propagated by little, tiny left-handed screw-like object that find threaded holes in iron so as to attract or repel. Gravitation is another kind of material experience. First, Descartes hypothesized about a material cause for phenomena and then deduced the consequences.

Descartes paved the way for a reasoned approach to physics, that turns out to have been a part of the story. He motivated Newton and helped European thinkers to find their way to independent ideas, shedding the overbearing weight of Aristotelianism and Church dogma.

But this chapter is devoted to mathematics.

2.2.2 Descartes' Algebra-fication of Geometry

...or geometrification of algebra! Whatever. Descartes brought geometry and algebra together for the first time by reinterpreting the latter and inadvertently, rendering the former less important.[13]

[13] for a while.

Descartes pulled the very new, very unsophisticated new method of "algebra" to a role of supremacy over geometry. He did this by linking the solution of geometry problems—which would have been done with rule-obsessive construction of geometrical proofs—to solutions using symbols. He did this work in a small book called *Le Géométrie* (*The Geometry*), which he published in 1637, the same year he published his *Discourse on Method.*[14]

He instituted a number of conventions which we use today. For example, he reserved the letters of the beginning of the alphabet $a, b, c, \ldots$ for things that are constants or which represent fixed lines. An important strategic approach was to assume that the solution of a mathematical problem may be unknown, but can still be found and he reserved the last letters of the alphabet x, y, z... to stand for unknown quantities—variables. He further introduced the compact notation of exponents to describe how many times a constant or a variable is multiplied by itself.

Prior to Descartes, ab would be the product of a and b but explicitly refer to the area of a rectangle bounded by legs of lengths a and b. a^3 would be the volume of a cube. There would be no such thing as $abcd$ or a^4 because after all, nature has no more dimensions than 3. So the early algebra was confined to a strictly dimensional context. Descartes broke with that and explored equations of higher powers, even showing that equations of higher powers could be reduced to lower power equations and so on until a solution could be found. He did this algebraically and geometrically, side by side. In fact, *Le Géométrie* is just one example worked out after another: it's solutions-oriented. And it's abstract. There's no need to identify "things" to the variables, although one could do so if desired.

Just as arithmetic has addition, subtraction, multiplication, division, and square roots...so to he found geometrical interpretations of these operations. His geometrical description of multiplication—not referring to an area—is instructive of how he did things. FIgure 2.2 shows a figure from *Le Géométrie*. Using his notation, we immediately come upon a new "invention" of his: unity. A line of length "1" could be chosen arbitrarily, and then manipulated.

In Fig. 2.2 I've overlaid red letters in the fashion that Descartes would have, assigning a single letter to represent a line. The lines $\overline{DE}$ and $\overline{AC}$ are both parallel and so the triangles BED and BCA are similar. From elementary geometry, because of their similarity, we would have

$$\frac{b}{d} = \frac{c}{a}.$$

Now he does this clever thing with "1" and assigns the length $\overline{AB}$ to have length 1 so that we have

$$\frac{b}{d} = \frac{c}{1}.$$

and so the product of $cd = b$. No areas. A brand new use of the brand new algebra!

Here's another example from *Le Géométrie*. Supposed you want to find the square root of a quantity. Figure 2.3 is again from his book. His trick here is to assign the distance $\overline{GH}$ to be an arbitrary length x[15] and the distance $\overline{GI}$ to be y. His goal is to compute the $\sqrt{y}$ for this abstract situation. Again, he uses the

[14] *Geometry* can be considered an appendix to the *Discourse*.

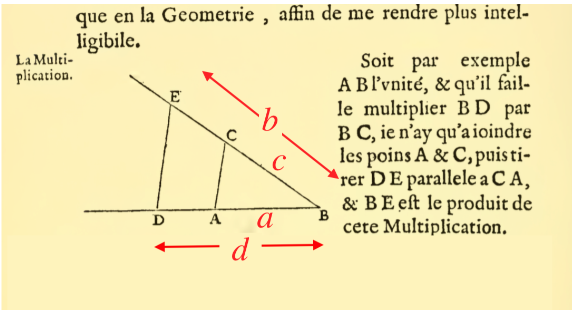
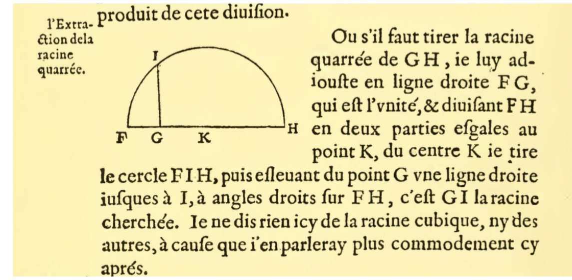


Figure 2.2: geometrymultiply

[15] See? Algebra with unknowns.

"1 trick" and makes $\overline{FG} = 1$. The end result is that $y = \sqrt{x}$ and the problem is solved in general terms and in a way that could be measured with a ruler. Like Euclid would have liked.


produit de cete diuifion.

l'Extraction dela racine quarrée.

Ou s'il faut tirer la racine quarrée de G H, ie luy adioufte en ligne droite F G, qui eft l'vnité, & diuifant F H en deux parties efgales au point K, du centre K ie tire le cercle F I H, puis efleuant du point G vne ligne droite iufques à I, à angles droits fur F H, c'eft G I la racine cherchée. Ie ne dis rien icy de la racine cubique, ny des autres, à caufe que i'en parleray plus commodement cy aprés.

Figure 2.3: root

The early translators of algebra considered equations in two unknowns—some $f(x, y) = 0$—to be impossible. Descartes actually found a way by treating the locus of points on a line as indeterminant, some abstract x. GIven any particular location along x however, another corresponding to the other unknown variable could be identified. He called such a point y and then worked to find solutions to particular problems that might be different depending on what the value of x was...but he did it in a way that was general for any x. This is the first example of what we'd now refer to as an axis. He didn't actually use two axes, but he still solved problems for an unknown y in terms of a parameter x. He called one of these the abscissa and the other, the ordinate.

Mathematicians picked up on these ideas and extended them into the directions that we now love. One of those was John Wallis (1616-1703), a contemporary of Isaac Newton who learned from Wallis enough to construct the general Binomial Theorem.

The use of perpendicular axes, which we call x and y stems from Descartes' inspiration which is why they're called Cartesian Coordinates.

Descartes managed to get himself into a dispute with a Calvinist theologian, Gisbertus Voetius who wanted his university to officially condemn the teaching of "Cartesian Philosophy" as atheistic and bad for young people. Descartes responded by printing a reaction which was posted on public kiosks. This must have been quite a sight! In any case, Descartes began to imagine that his time in the Netherlands was coming to a close. An admirer, the Queen Christina of Sweden, was an intellectual of sorts and invited Descartes to Stockholm to work for her court and to teach her. She even sent a ship to Amsterdam to pick him up. He eventually accepted the position and this was the beginning of the end for him.

She required his presence at 4 AM for lessons. This, from the fellow who had spent every morning of his life in bed until noon! He caught a serious respiratory infection and died on February 11th, 1650 at the age of only 53.

We moderns owe an enormous debt to this soldier-philosopher-mathematician. Both for what he said that was useful and for what he said that was nonsense, but which stimulated productive reaction. In what follows from Section 2.5 there is a direct line from every word back to René Descartes.

2.3 Introduction

In this chapter we'll do some old things and some new things. Some of the old things will be mathematical in nature, while some of the new things will include some terminology and some techniques. I promise that the math will not be hard and we'll get through it together. We'll develop just a few of these tools that we'll return to repeatedly: simple algebra, exponents, unit conversions, and powers of ten. It will come back to you.

But I want to start with some topics which are timely and confusing to non-specialists. What are we doing when we "do" science?

2.4 It's Theory, All the Way Down

Coming.

2.5 The M Word

The language of physics is mathematics, so uttered Galileo a long time ago (although he said that the language of the *universe* is mathematics). Well, he was right and we have no idea why that seems to reliably be the case! So the importance of that realization will become clear as we go, which is partly why I don't want to avoid mathematics altogether. But it will be relatively simple. You've seen everything I'll ask you to do in high school, at the very least. It will be fine. Let me show you.

> **Wait.** *I'm not a math person.*
>
> **Glad you asked.** *Actually, nobody is. Really mathematics is a habit of mind and strategy for how you read. Certainly for what we're going to do. I promise you. Read with your pencil out. Read every line with a mathematics symbol. You'll get it.*

2.5.1 Some Algebra

Our algebraic experience here will be some simple solutions to simple equations. I'll need the occasional square root and the occasional exponent, but no trigonometry or simultaneous equation solving and certainly no calculus. I'll refer to vectors, but you'll not need to do even two-dimensional vector combinations.

Our Algebra will be pretty simple with basically one rule: Whatever you do to the left hand side of an equation, you must also do to the right side and visa versa. Words to live by.

Let me make my point by going back to the Gravitation law and asking a simple scientific question of it.

Wait. *Why bother doing this? Use your words!*

Glad you asked. *There's an economy in using equations, but also a hidden power. The form of an equation that describes something that nature does encodes new information that can be discovered by manipulation...information that would not be obvious in an English sentence.*

Here's what I mean. I keep coming back to Newton's Universal Law of Gravitation which I can indeed describe in a paragraph. Here goes:

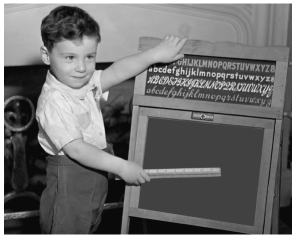
"The force of attraction experienced by two masses on one another is directly proportional to the product of those two masses and inversely proportional to the square of the distances that separate their centers. The constant of proportionality is called the Gravitational Constant which is $6.67408 \times 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$."

There. A perfectly good representation of Newton's Gravitational Law. Lots of writing, so it's inefficient. If I gave it a nickname, say Newton's Law and then used those two words every time I meant to refer to it, you might have to go back and re-read the paragraph again... and again. But what this doesn't do, besides allow you to quickly move through a gravity-narrative, is help you to find out new things about nature.

I mentioned that it's hard to measure G. Why is that? Does the paragraph enunciation of Newton's Law help you to estimate the ease or difficulty of making that measurement? I don't think so. But if we look at it as a formula, we can interrogate it and answer our question.

$$F = G\frac{mM}{R^2}$$

and then use the rules of algebra to ask about G and see what results. Let's do that:



You Do It 2.1.
/toolkit/SolvingNewton

Suppose we want to measure the Gravitational Constant, G. We expect it to be small...it's in the range of 10^{-11}. We have to use the tools available which include a climate and vibration-free lab area that's about 1 meter long and a dime-store spring scale that's incapable of measuring forces less than 0.1 Newtons. Can we make this measurement using any kind of reasonable masses? Does this experiment make sense?

You needed to literally touch equations and move the pieces around in order to gain insight.

Even if it's a part of the text, you should copy it out while you read. Remember, these parts are marked by

Our appetite for algebraic complexity in QS&BB will be limited. For example, we'll not encounter formulas that are much more complicated than these:

$$y = a \times x = ax \quad \text{solve for } x \text{ to get} \quad x = y/a$$
$$y = x + z \quad \text{solve for } x \text{ to get} \quad x = y - z$$
$$y = a \times x + b = ax + b \quad \text{solve for } x \text{ to get} \quad x = \frac{y-b}{a}$$
$$y = \sqrt{a+x} \quad \text{solve for } x \text{ to get} \quad x = y^2 - a$$

You can do this, right? That's about all that you'll need to remember of algebra. Just remember the rule. Then...it's merely a game—a puzzle to solve.

There's an important reason I have chosen to include some mathematics in QS&BB: I'd hate for you to miss...dare I say...a spooky feature of the universe. It behaves as if mathematics is an essential part of how it works.[16]

We'll take it slow with the math, but even a little will add a lot to your understanding. So let's spend the rest of this chapter reminding yourself of things that you would have learned in high school.

[16] There has been this eyes-open discussion in physics for a century now. Is mathematics invented or is it discovered? The former would suggest that it's in some sense, man-made. The latter would suggest that it's a deeply embedded feature of nature. . . to be found out. In 1960 the famous mathematical physicist Eugene Wigner wrote a paper that's still read today called the *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*. Ask Mr Google about it. Almost 30,000 hits, almost all of them "reprints."

2.5.2 The Powers That Be

Once in a while, we'll need to multiply or divide terms that have exponents. There are simple rules for this, but let's figure them out by hand...so to speak. The first thing to remember about exponents is that in a term like x^n, a positive integer n tells you how many times you must multiply x by itself. So:

$$x^1 = x.$$

Here, there's just one x, so: $x^1 = x$.
The second thing to remember is that $x^0 = 1$. There aren't any $x's$ in the product and so all that could be there is 1. Armed with that, let's kick it up a notch.

Suppose I have

$$x \times x$$

You'd be pretty comfortable calling that "x-squared"[17] and from the above, the number of $x's$ there are in that product is two. So

$$x \times x = x^2.$$

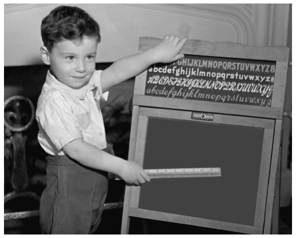
If I add another product, then I'd have $x \times x \times x = x^3$. Get it? Notice that what we've also got in this equation is:

$$x \times x \times x = x^2 \times x^1 = x^3$$

and we've just developed our first rule on combining exponents:

$$x^n \times x^m = x^{n+m}.$$

Now you try it.



What is $x^2 x^1 x^4$?

You Do It 2.2. /toolkit/Exponents

One more time, but different. Another rule:

$$x^{-n} = \frac{1}{x^n}.$$

If the same rule for adding exponents works—and it does—then we can multiply factors with powers by keeping track of the positive and negative signs of the exponents.

[17] From the discussion of Descartes, you can see why the word "squared" is used since this is a legacy of the early linking of algebra with geometry. Ditto for "cubed."

So here's an easy one, first by multiplying everything out:

$$\frac{x \times x \times x}{x \times x} = x$$

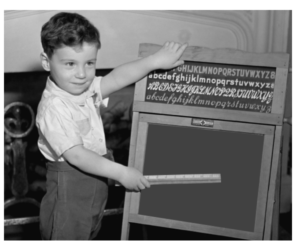
and now by using the powers and the rule:

$$\frac{x \times x \times x}{x \times x} = \frac{x^3}{x^2} = x^3 \times x^{-2} = x^{3-2} = x.$$

One more thing. The powers don't have to be integers.

Perhaps you'll remember that square roots can be written:

$$\begin{aligned}
\sqrt{x} &= x^{0.5} = x^{1/2} \\
\text{so:} & \\
\sqrt{9} &= 3 \quad = 9^{0.5} \\
\text{or:} & \\
\sqrt{\frac{1}{9}} &= \left(\frac{1}{9}\right)^{0.5} = \frac{1}{\sqrt{9}} = \frac{1}{9^{0.5}} = 9^{-0.5} \\
&= \frac{1}{3}
\end{aligned}$$



You Do It 2.3.
/toolkit/ExponentsAgain

What is $x^{-2}x^{1}x^{4}$?

That's it. Now we have everything we need to turn numbers into sizes of...stuff.

2.5.3 Units Conversions

Numbers are just numbers without some label that tells you what they refer to. Now not all numbers have to refer to something, pure number is a respectable object of mathematical research—prime numbers for example have been a topic of research for centuries. Irrational numbers–those that can't be expressed as a ratio of whole numbers, like π, –are likewise objects with no necessary relationship to..."stuff" in our world.

We're concerned with numbers that measure a parameter or count physical things and they come with some reference ("foot") unit that is a customary way to compare one thing with another.[18] Of course not everyone agrees on the units that should be used. Wait. There's *the world*, that agrees on one set and then there's the United States that marches to its own set of units. Thinking of you, feet.

I'll not use Imperial units (feet, inches, pounds, etc.) very much, except to give you a feeling for something that you've got an instinct for. . . like the average height of a person. We'll use the metric system, in particular the MKS units[19] in which the fundamental length unit is the meter (about a yard).

Just like an exchange rate in currency, so many euros per dollar, we'll need to be able to convert, among many different units. All the time.

[18] "Apples and Oranges" is a phrase that refers to units...you need to keep your fruit straight.

[19] This stands for meter-kilogram-second, as the basic units of length, mass, and time. It's a dated designation as the real internationally regulated system is now the International System of Units (SI) which stands for *Le Système International d'Unités*. The French have always been at the forefront of this.

Understand conversions! Conversions are a part of life! At least in QS&BB.

Let's get our bearings. What's a common sort of size in life? How about the height of an average male. Mr Google tells me that's about 5'10". How many inches tall is our average male? Here's the thought-process you'd use to calculate this.

Pencil 2.1. ✎

Three steps:

1. A single foot is 12 inches.
2. So, 5 feet is $5 \times 12 = 60$ inches
3. and the combination is $60 + 10 = 70$ inches.

...which you could do in your head I'll bet. But this simple, almost intuitive calculation uses a more general conversion from one unit to another through the use of a conversion factor. All unit manipulations use a conversion factor, which is just a number,[20] which will be expressed as a ratio or fraction, of the conversion of one set of units ("from") to the new set ("to"). It will appear like this:

[20] . . . a number that's actually like a fancy way to write "1" since it's really relating one thing in a set of units to the same thing in a different set of units.

$$\text{where you're going to} = \left(\frac{\text{to}}{\text{from}}\right) \times \text{where you're coming from}$$

The action is in that bracketed term. It's arranged so that the "from" in the denominator cancels the *units* of the right hand "coming from" term. What's left in the numerator you intentionally set up to have the units of what you are going to, here in step 2 above...we're going from feet to inches. In this case, step 1 defines the bracket and step 2 uses it and in symbols, step 1 says:

$$\left(\frac{\text{to}}{\text{from}}\right) = \frac{\text{number of inches in a foot}}{\text{a foot}} = \frac{12}{1}$$

So armed with this, we can do the conversion of feet to inches.

$$\begin{aligned}\text{five feet in inches} &= \frac{\text{number of inches in a foot}}{\text{a foot}} \times 5 \text{ ft} = \frac{12}{1} \times 5 \text{ ft} \\ &= \frac{12 \text{ inches}}{1 \text{ ft}} \times 5 \text{ ft} = \frac{60}{1} \text{inches} \\ &= 60 \text{ inches.}\end{aligned}$$

There's another way to think about this (which is identical, but just spun differently) which might be useful. You know that you can always multiple any number times 1 and get that back. So in the inches-feet world, we could write:

$$\begin{aligned}1 \text{ foot} &= 12 \text{ inches} \\ 1 &= \frac{12 \text{ inches}}{1 \text{ foot}}\end{aligned}$$

It's that "1" that I want to use to convert 5 feet to inches. We'd do that by writing:

$$5 \text{ foot} = x \text{ inches (looking for } x \text{ here)}$$

$$5 \text{ foot} \times 1 = x \text{ inches (haven't done anything with “}\times 1\text{”)}$$

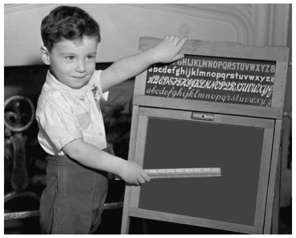
$$5 \text{ foot} \times \frac{12 \text{ inches}}{1 \text{ foot}} = x \text{ inches (used what “1” is here)}$$

$$5 \cancel{\text{ foot}} \times \frac{12 \text{ inches}}{1 \cancel{\text{ foot}}} = 5 \times 12 \text{ inches} = 60 \text{ inches} = x \text{ inches}$$

Notice that we treat units like algebraic terms and can cancel them as if they were symbols or numbers: the "feet" cancel above. That's the neat thing. If you set up the conversion factor right, the units will multiply and divide along with numbers so you can always see that you get what you want. While this is a particularly simple conversion, sometimes we'll need to do some which are either more complicated, or use units that maybe you're not very familiar with. I won't be so pedantic usually, but hopefully you get the point!

Let's do a harder one. *If a furlong is 201.2 meters, how furlongs are there in a mile?* What we know — the "1" as in the above discussion is that 1 furlong = 201.2 m. Then we have to think about it since *miles* is where we start from, not meters. More conversions. How you do this might depend on what you remember. For me[21] what is stuck in my head is that a mile is 5,280 feet and that a foot is 12 inches and that an inch is 2.54 centimeters and that a meter is 100 cm. So I always start there. You might do it differently. So for me, that's 4 conversions, or four brackets along with my fancy "1" that I would use to do this conversion. It's kind of fun. Really.

[21] ...for some reason



You Do It 2.4. /toolkit/FurlongMi

How many furlongs in a mile if there are 201.2 meters in a single furlong?

Did you get that there are 8 furlongs in a mile? If not, click on the little guy and watch me do it. I've collected a number of the useful conversions into graphs which you can use later.

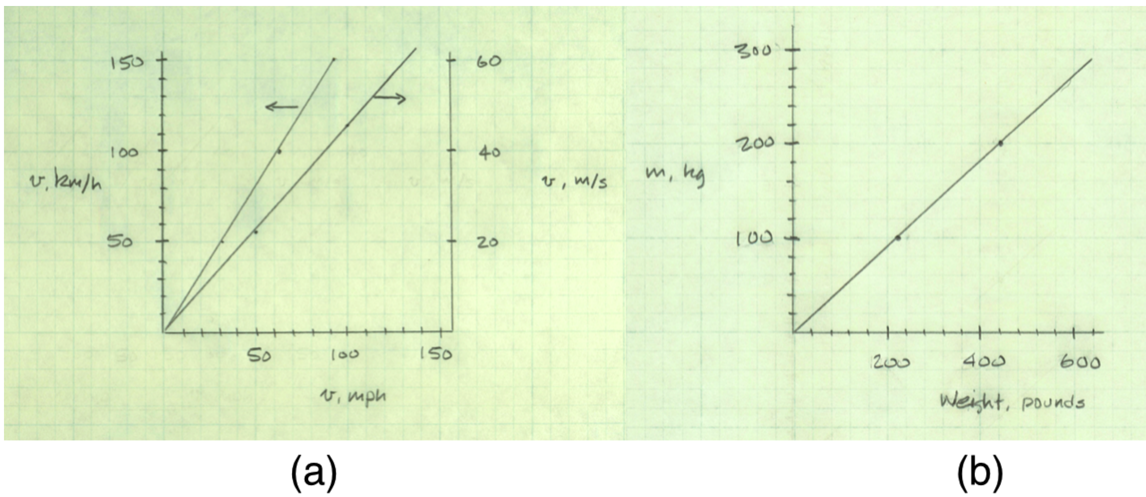


Figure 2.4: The right hand curve shows a constant speed of 4 m/s, holding steady for 10 s. The left hand curve shows the distance that an object will travel at that constant speed as a function of time.

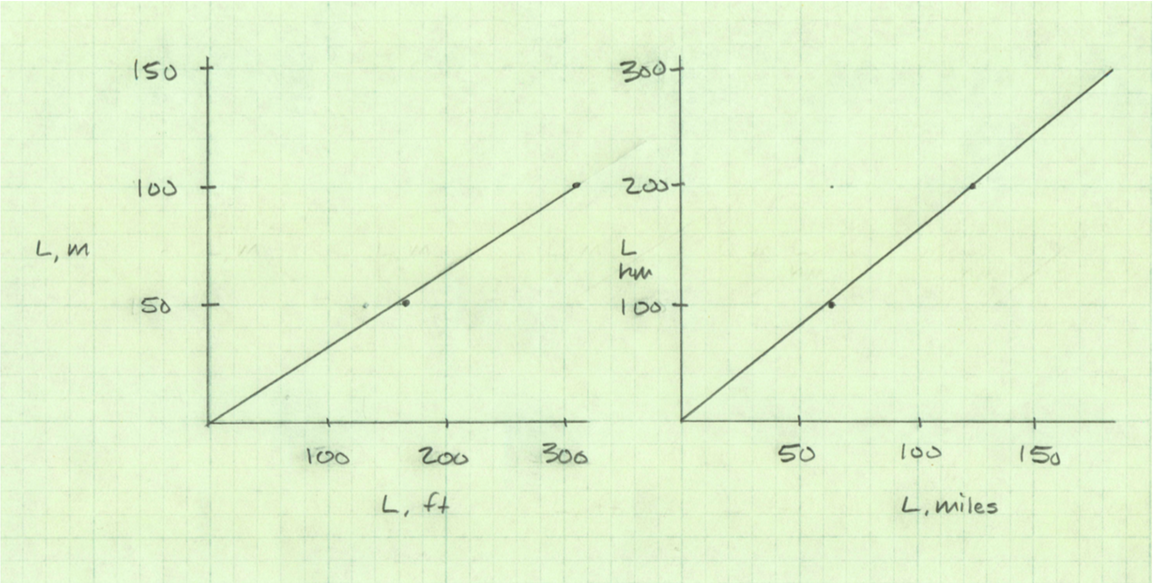


Figure 2.5: The right hand curve shows a constant speed of 4 m/s, holding steady for 10 s. The left hand curve shows the distance that an object will travel at that constant speed as a function of time.

2.5.4 The Big 10: "Powers Of," That Is

One of the more difficult things for us to get our heads around will be the sizes of things, the speeds of things, and the masses of things that fill the pages of QS&BB. Lots of zeros means lots of mistakes, but it also means a complete loss of perspective on relative magnitudes. Big and small numbers are really difficult to process for all of us.

As we think of things that are bigger and bigger and things that are smaller and smaller, where do you start to loose track and one is the same as another? Keep in mind our average-guy height of about a meter and half–for this purpose, thing... "about a couple of meters"–and here is a ranked list of big and small things with approximate sizes:

1. African elephant, 4 m
2. Height of a six story hotel, 30 m
3. Statue of Liberty, 90 m
4. Height of Great Pyramid of Giza, 140 m
5. Eiffel Tower, 300 m
6. Mount Rushmore 1700 m

7. District of Columbia, 16,000 m square
8. Texas, East to West, 1,244,000 m
9. Pluto, 2,300,000 m diameter
10. Moon, 3,500,000 m diameter
11. Earth, 12,800,000 m diameter
12. Jupiter, 143,000,000 m diameter
13. Distance Earth to Moon, 384,000,000 m
14. Sun, 1,390,000,000 m diameter
15. distance, Sun to Pluto, 5,900,000,000 m
16. Distance to nearest star (Alpha Centuri), 41,300,000,000,000,000,000 m
17. diameter of the Milky Way Galaxy, 950,000,000,000,000,000,000 m
18. Distance to the Andromeda Galaxy, 24,000,000,000,000,000,000,000 m
19. Size of the Pisces–Cetus Supercluster Complex, our supercluster, 9,000,000,000,000,000,000,000,000 m
20. Distance to UDFj-39546284, the furthest object observed, 120,000,000,000,000,000,000,000,000 m

Do I need to go any further? Given what I know from my life, I have a pretty good idea of how big #1-8 are. Beyond that, I have no idea how much bigger the Milky Way Galaxy is than the size of Jupiter. It all blends together.

But there's a way: exponential notation. . . using our power rules and the number 10. It's easy.

A number expressed in exponential notation as:

$$\text{a number} \times 10^{\text{power}}$$

Let's think about this in two parts. First, the 10-power part.

The rules above work for 10 just like any number, so 10^n is shorthand for the number that you get when you multiply 10 by itself n times. This has benefits because of the features of 10-multiples, that we count in base-10, and how you can just count zeros. So for example:

$$10^3 = 10 \times 10 \times 10 = 1,000.$$

The power counts the zeros, or more specifically, the position to the right of the decimal point from 1. So if you have any number, you can multiply it by the 10-power part and have a compact way of representing big and small numbers. So, following through:

$$3 \times 10^3 = 3 \times 10 \times 10 \times 10 = 3 \times 1000 = 3000.$$

We can do the same thing with numbers less than 1, by using negative exponents for the 10-power part.

$$0.03 = \frac{3}{100} = \frac{3}{10^2} = 3 \times 10^{-2}.$$

So you just move the decimal place the power-number to the right to go from 3×10^{-2} to 0.03.

The second thing is the number in front that multiplies the power of 10. It's called the "mantissa" and that's all it is. . . a number.

Now that confusing list above can be written in a way that's more likely to allow your brain to compare one with the other, since now you'll immediately see that one thing is 10 or 1000 or so-on times another.

1. African elephant, 4 m
2. Height of a six story hotel, 30 m, 3.0×10^2 m
3. Statue of Liberty, 90 m, 9.0×10^2 m
4. Height of Great Pyramid of Giza, 140 m, 1.4×10^2 m
5. Eiffel Tower, 300 m, 3.0×10^2 m
6. Mount Rushmore 1700 m, 1.7×10^3 m
7. District of Columbia, 16,000 m square, 16.0×10^3 m, or 1.6×10^4 m
8. Texas, East to West, 1,244,000 m, 1.244×10^6 m
9. Pluto, 2,300,000 m diameter, 2.3×10^6 m
10. Moon, 3,500,000 m diameter, 3.5×10^6 m
11. Earth, 12,800,000 m diameter, 12.8×10^6 m, or 1.28×10^7 m
12. Jupiter, 143,000,000 m diameter, 143.0×10^6 m, or 1.43×10^8 m
13. Distance Earth to Moon, 384,000,000 m, 384.0×10^6 m, or 3.84×10^8 m
14. Sun, 1,390,000,000 m diameter, 1.39×10^9 m
15. Distance, Sun to Pluto, 5,900,000,000 m, 5.9×10^9 m
16. Distance to nearest star (Alpha Centuri), 41,300,000,000,000,000,000 m, 41.3×10^{18} m, or 4.13×10^{19} m
17. diameter of the Milky Way Galaxy, 950,000,000,000,000,000,000 m, 950×10^{18} m, or 9.5×10^{19} m
18. Distance to the Andromeda Galaxy, 24,000,000,000,000,000,000,000 m, 24.0×10^{21} m, or 2.4×10^{22} m
19. Size of the Pisces–Cetus Supercluster Complex, our supercluster, 9,000,000,000,000,000,000,000,000 m, 9.0×10^{24} m
20. Distance to UDFj-39546284, the furthest object observed, 120,000,000,000,000,000,000,000,000 m, 120×10^{24} m or 1.2×10^{26} m

So now you can compare and see that the distance from the Earth to the Moon is only a little more than three times the diameter of Jupiter. Now your "mind's eye" springs into action since you can sort of imagine three Jupiters between us and the Moon. With all of those zeros, I couldn't do that!

Powers of 10 have nicknames...Is "a google" really a power of ten?[22] Here's an official table of the names, size, and abbreviation for most of them:

Let's work out an example. Something you can use at a party. I first worked this out for a class when I was in Geneva, Switzerland working at CERN. It was July 4, 2010, which was just another Sunday over there. The United States came into existence on July 4, 1776[23] which was 2010 – 1776 = 234 years ago.

So how many seconds had the United States been around if we start from midnight on July 4, 1776?

Pencil 2.2.

[22] No. The word is Googol and it's $10^{1}00$. The rumor is that the Google founders misspelled it when they incorporated.

[23] Actually, the Declaration of Independence wasn't fully signed until August 2, 1776—my birthday! The day, not the year.

$$\begin{aligned}
\text{234 year per U.S.} &= 2.34 \times 10^2 \ \frac{\text{years}}{\text{U.S.}} \\
\text{86,400 seconds per year} &= 8.64 \times 10^4 \ \frac{\text{seconds}}{\text{year}} \\
\text{So:} & \\
\text{seconds per U.S.} &= 2.34 \times 10^2 \ \frac{\cancel{\text{year}}}{\text{U.S.}} * 8.64 \times 10^4 \ \frac{\text{seconds}}{\cancel{\text{year}}} \\
&= (2.34) * (8.64) \times 102 * 10^4 = (2.34) * (8.64) \times 10^6 \\
\text{seconds per U.S.} &= 20.218 \times 10^6 \\
\text{seconds per U.S.} &= 2.0218 \times 10^7
\end{aligned}$$

Wait. *You mean I treat the words of units as if they were algebraic variables?*

Glad you asked. *Yes. You can do that and even catch mistakes when the products and cancellations don't lead to what you expect. Had I gotten miles times hours, I'd know my actual formula was wrong even before doing it. No charge for this hint. Use it wisely.*

There are a few of things to notice here. First, that's a lot of seconds! Second (get it?), to multiply two numbers together, you separate the mantissas, and multiply them, and the exponents, and add them...separately.[24] Please understand these operations by doing them over by hand. The obvious thing happens when there are negative exponents involved. For example, convince yourself that 15% of the lifetime of the U.S. is 3,032,700 seconds, and do it by treating 15% as

$$15\% = 0.15 = 1.5 \times 10^{-1}.$$

Finally, notice that I canceled the units of "year." You can always do that with units—set them up right, keep them in your equations, and you can quickly find mistakes. Here, the units on the right have to give you the units on the left, which we wanted: "seconds/U.S."

[24] Remember? The "mantissa" in $X \times 10^y$ is X and the exponent is the y.

2.5.5 Graphs and Geometry

One of the amazing mathematical discoveries of the 17th century was that geometry could be tied to algebra through the use of the growing notion of a function. This is almost entirely due to Rene Descartes and Leonhard Euler (1707-1783)[25]

[25] Euler was one of the most amazing mathematicians in history. He did so much that his work is still being analyzed and cataloged today. To him we owe the notion of a function. But he also worked in physical problems like hydrodynamics, optics, astronomy, and even musical theory. While Swiss, Euler lived and worked most of his life in St. Petersburg, Russia.

We will deal with some functions that would be very hard to evaluate on your calculator. But Descartes' gift is that I can show you the graph and evaluation can be done by eye, which is in effect solving the equation. We'll use some simple geometrical relations which I'll summarize here.

Table 2.1: More powers of ten than you ever wanted to know. Except that many of them we need to know.

septillionth	yocto-	y	0.000000000000000000000001	10^{-24}
sextillionth	zepto-	z	0.000000000000000000001	10^{-21}
quintillionth	atto-	a	0.000000000000000001	10^{-18}
quadrillionth	femto-	f	0.000000000000001	10^{-15}
trillionth	pico-	p	0.000000000001	10^{-12}
billionth	nano-	n	0.000000001	10^{-9}
millionth	micro-	μ	0.000001	10^{-6}
thousandth	milli-	m	0.001	10^{-3}
hundredth	centi-	c	0.01	10^{-2}
tenth	deci-	d	0.1	10^{-1}
one			1	10^{0}
ten	deca-	da	10	10^{1}
hundred	hecto-	h	100	10^{2}
thousand	kilo-	k	1,000	10^{3}
million	mega-	M	1,000,000	10^{6}
billion	giga-	G	1,000,000,000	10^{9}
trillion	tera-	T	1,000,000,000,000	10^{12}
quadrillion	peta-	P	1,000,000,000,000,000	10^{15}
quintillion	exa-	E	1,000,000,000,000,000,000	10^{18}
sextillion	zetta-	Z	1,000,000,000,000,000,000,000	10^{21}
septillion	yotta-	Y	1,000,000,000,000,000,000,000,000	10^{24}

Formulas From Your Past

I know that you've seen most of this somewhere in your past! So return with us now to those thrilling days of yesteryear.[26]

[26] Google it!

Equation of a Straight Line

A straight line with a slope of m and a y intercept of b is described by the equation:

$$y = mx + b. \tag{2.1}$$

Figure 2.6 shows such a straight line.

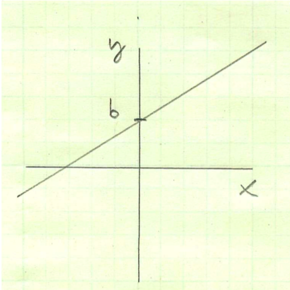


Figure 2.6: straight

Equation of a Circle

A circle of radius R in the $x-y$ plane centered at a (a,b) is described by the equation:

$$R^2 = (x-a)^2 + (y-b)^2. \tag{2.2}$$

Of course if the circle is centered at the origin, then it looks more familiar as

$$R^2 = x^2 + y^2. \tag{2.3}$$

is described by the formula Figure 2.7 shows such a circle.

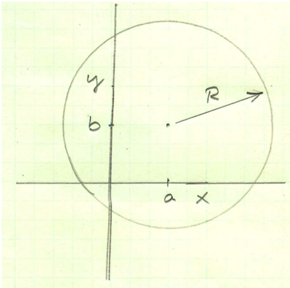


Figure 2.7: circle

Equation of a Parabola

A parabola in the $x-y$ plane with vertex at (a,b)

$$y = C(x-a)^2 + b \tag{2.4}$$

where C is a constant. Figure 2.8 shows a parabola.

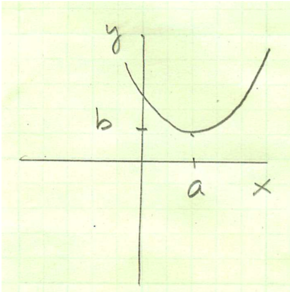


Figure 2.8: parabola

Area of a Rectangle

A rectangle with sides a and b has an area, A of

$$A = ab \tag{2.5}$$

Area of a Right Triangle

A right triangle (which means that one of the angles is 90 degrees) with base of a and height of b has an area, A of

$$A = 1/2ab. \tag{2.6}$$

For a right triangle, the base and height are equal to the two legs. But the formula works for any triangle. Figure 2.9 shows how that works.

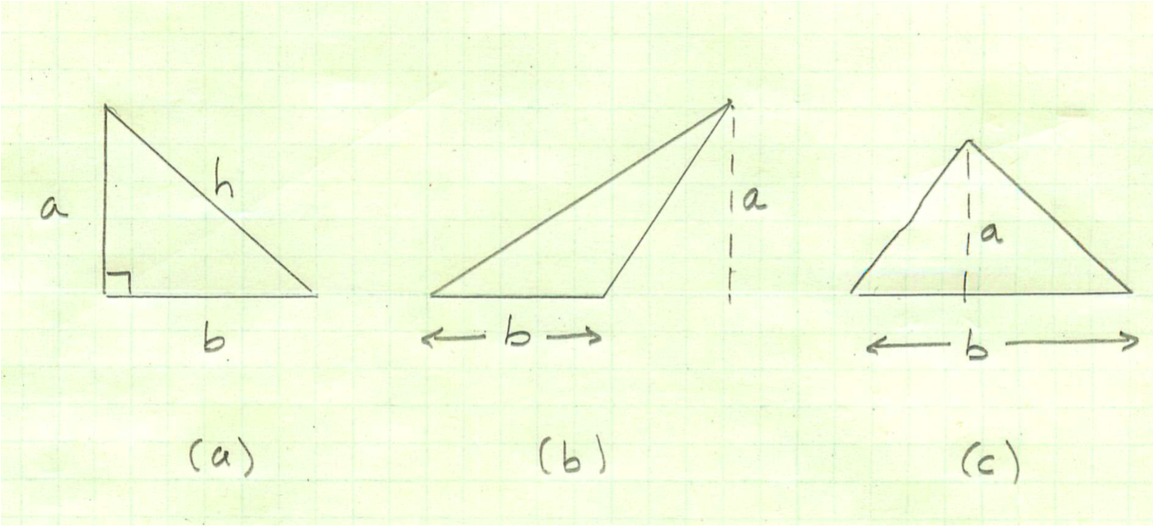


Figure 2.9: triangles

Figure 2.10: You realize that two pizzas is a "circumference"? Because...wait for it...it's "2 pie are." You're welcome. (papajohns)

Area and Circumference of a Circle

For a circle of radius R, the area, A is

$$A = \pi R^2 \tag{2.7}$$

and the circumference, C is

$$C = 2\pi R. \tag{2.8}$$

Pythagoras' Theorem

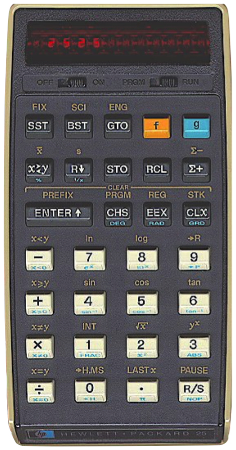
For a right triangle, the hypotenuse, h is related to the lengths of the two sides a and b by the Theorem of Pythagoras:

$$h^2 = a^2 + b^2. \tag{2.9}$$

2.6 Shapes of the Universe

One of the remarkable consequences of the mathematization of physics that began with Descartes is that we've come to expect that our descriptions of the universe will be in the language of mathematical *func-*

tions. Do you remember what a function is? The fancy definition of a function can be pretty involved, but you do know about function machines and I'll remind you how.



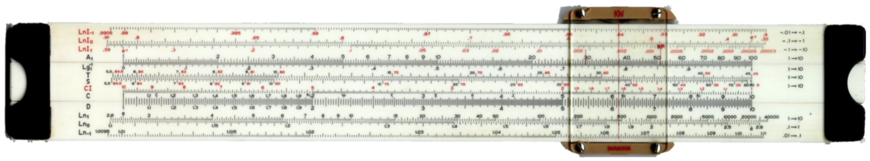


Figure 2.11: Left: the venerable HP-25 programable (!) scientific calculator. Right: a slide rule used for all calculations until the early 1970's. It was not programmable (although it was wireless).

When I was a senior in college, finishing my electrical engineering degree, our department had a visitor from the Hewlett Packard Company. It was either Bill Hewlett or Dave Packard, I can't remember which. But they promised to do away with the slide rule that we all carried around with us everywhere and showed us a brand new product: a portable scientific calculator, that they called the electronic slide rule. This was 1972 and he showed us the first HP calculator, the HP-35. Needless to say, I couldn't afford it—it cost $400— but later in graduate school I bought my first scientific calculator, the HP-25, pictured in Fig. 2.11 along with the slide rule that I carried for four years. Today I've got more processing power in my watch then I had in that calculator. But I'll bet you've got something like it...calculators are nothing but electronic function machines. So in the spirit of Fig. 3, Fig. 2.12 shows the circuit board from the inside of the HP-25 with it's simple processor at the bottom.

%

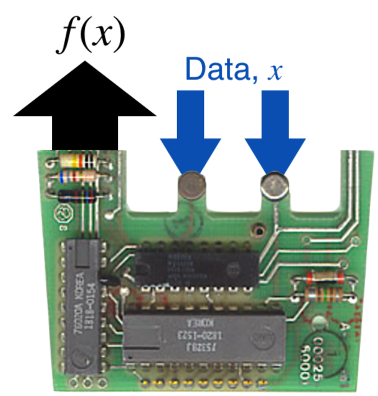


Figure 2.12: The AMI 1820-1523 Arithmetic, Control Timing processor: the heart of a function machine. Adapted for my silly purposes, but I'll bet you won't forget it! The tabs at the blue arrows are actually connected the processor to the keyboard. That's how data get in.

2.6.1 Functions: Mathematical Machines

Figure 2.12 shows what a function does: if you enter data through the keypad—a value of x—and hit the appropriate button, the display shows the value of the function. So if the function was the formula $f(x) = x^2$ and if I keyed in "4" and pushed the x^2 button, the display would read "16," the value of $f(2)$ for that particular function. Notice that it doesn't give you more than one result, and that's a requirement of a function: one result.

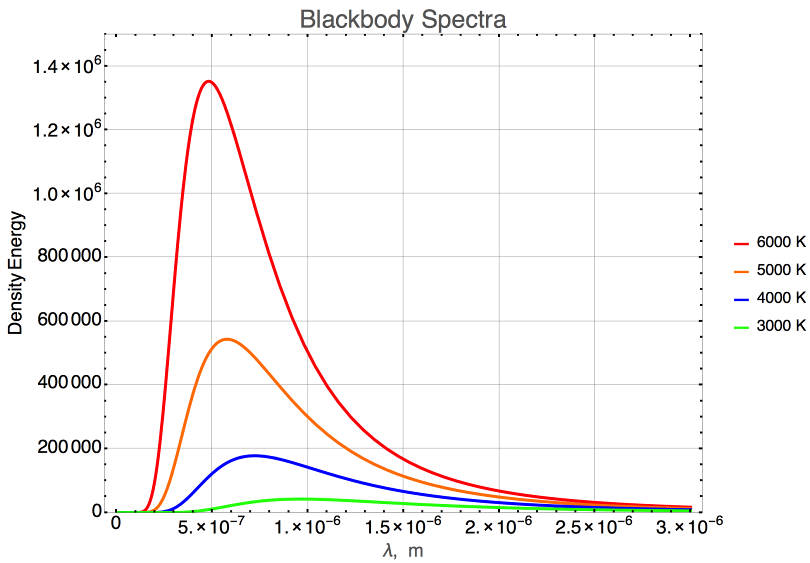


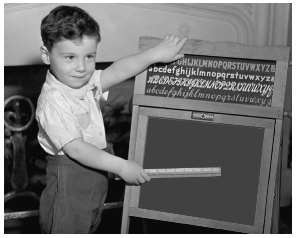
Figure 2.13: blackbodyvarious

So that's all a function is: a little mathematical machine that reports a single result for one or more inputs according to a rule. For us, functions can be represented by a formula, an algorithm, a table, or a graph. In all cases, it's one or more variables x or x & y... or x &y &z... *in*, a rule about what happens to them, and one numerical result *out*.

Your algebra teacher would have called the inputs (e.g., $x, y, ...$) the independent variables, which would have been members of the function's "Domain," and the output (e.g., $f(x, y, ...)$ or often y) the dependent variable, which would have been inside the "Range."

Nature seems to live by functions[27] and since in QS&BB we're all about Nature, we'll need to use functions. We'll solve actual formulas when they're simple functions and analyze plots of functions when they're complicated. For example, Fig. 2.13 is a function of two variables, a wavelength, λ and temperature (the units don't matter here). It's a messy formula which we'll admire, but not derive in Chapter 16. But boy is it an important function. Here the little function machine calculates the value of the energy density of the radiation emitted by an object heated to a particular temperature. If you provide a wavelength and a temperature (in the figure, 3,000, 4,000, 5,000, or 6,000 degrees) to the function, then it reports back to you the value of the energy density that the body radiates. You can evaluate that function:

[27] Why? We don't know.



What is the ratio of the value of the energy densities for one object at 4,000 degrees and another at 5,000 degrees at a wavelength of 1×10^{-6} meters?

You Do It 2.5. /toolkit/GraphRead

There. You just evaluated a complicated function...twice.

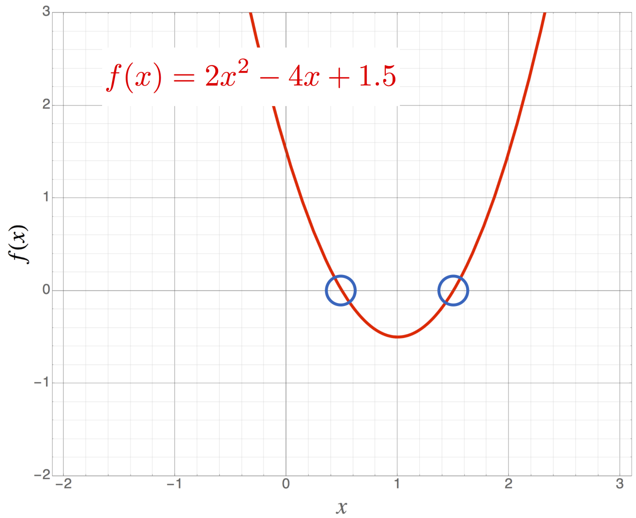


Figure 2.14: The quadratic function $f(x) = 2x^2 - 4x + 1.5$. plotted with blue circles at the points where $f(x) = 0$, the roots.

2.6.2 Polynomials

Many of Nature's functions are in the form of polynomial equations, which are reminiscent of the quadratic equation:

$$f(x) = ax^2 + bx + c. \tag{2.10}$$

You may have "solved" this equation in a number of ways in your algebra classes. What solving means is finding the x's for which the value of the function is zero. There's also a geometrical interpretation of "solving" a polynomial and an algebraic rule for doing it. Notice that the quadratic has the form of the equation of a parabola, so let's look at an example:

$$f(x) = 2x^2 - 4x + 1.5. \tag{2.11}$$

Remember that we can plot functions and Fig. 2.14 is a graphical representation of this function. When you solved a quadratic, you actually found the values of x for which the value of the function value—these are the "roots" of the function—of which there are two which I've called x_1 and x_2. So if we plug either into Eq. 2.10, then we will get $f = 0$..[28]

For quadratic equations, there is also a single formula to calculate the roots directly.[29] If we take Eq. 2.10 as the general form, then the "quadratic formula" you might remember from a former mathematics life is

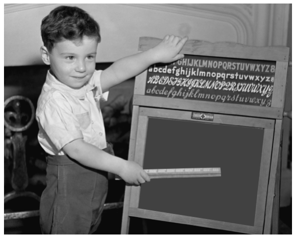
$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{2.12}$$

Of these two solutions: x_1 is for the + sign and x_2 is for the − sign.[30] So for our example in Eq. 2.11, $a = 2, b = -4$, and $c = 1.5$.

[28] Remember that the degree of a polynomial corresponds to the number of roots. For a quadratic, the degree is 2. For a cubic, it's 3 and so on.

[29] For cubics, there is a procedure. For polynomials of higher degree, it's complicated!

[30] Or the other way around—your choice.



For the example quadratic, use the quadratic formula, Eq. 2.12 to find the two roots of the function, Eq. 2.11. Do they match the "solution" you would get by looking at Fig. 2.14?

You Do It 2.6. toolkit/Quad

A polynomial can be of any "degree," which is the highest power of x. Since the middle of the 16th century (Copernicus' time) mathematicians had figured out how to expand any such function for an arbitrary degree, like $(a+x)^n$, where n is a positive integer. This formula would save work since expanding $(a+x)^n$ if n was anything bigger than about 3 is a lot of calculating. Let's expand a quadratic polynomial, that is for $n=2$:

$$(a+x)^2 = (a+x)(a+x) = a^2 + ax + xa + x^2 = x^2 + 2ax + a^2 \tag{2.13}$$

This old magic expansion formula is called the Binomial Expansion for polynomial of degree n—it has $n+1$ terms:

$$(a+x)^n = a^n + na^{n-1}x + \frac{n(n-1)}{2!}a^{n-2}x^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}x^3 ... + x^n \tag{2.14}$$

Until our hero, Isaac Newton came along, n was always a positive integer in this context.[31]

[31] Remember that the $n!$ notation stands for "n factorial." Which is $n! = n(n-1)(n-2)(n-3)...1$

Approximating Functions

Newton began inventing mathematics in the 17th Century and found a way to expand a formula for cases in which n could be anything: a positive integer, a negative integer, or even a fraction.[32] The result was an expansion that has an infinite number of terms! In contrast to how that sounds, it's actually very useful for many physics applications as we'll see.

[32] This was an essential step in the invention of the calculus. . . and the thing that Leibniz learned from Newton and used himself to invent a competing version of calculus. We'll touch on this in Chapter 5.4.1.

Let's take a particular case in which $a=1$ and write it out Newton's idea in the same spirit as Eq. 2.14.

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 ... \tag{2.15}$$

Here's where it will be interesting for physics. Look carefully at Eq. 2.15: each term is proportional to an increasing power of x, x^2, x^3, x^4 and so on. In physics, we can use this to make accurate approximations.[33] Suppose that $x<1$. Then each term gets smaller and smaller since $x^3 < x^2$ and so on if $x<1$...so each

[33] While this sounds like just a work-saver, we'll see that it actually allows us to sometimes gain insight of some tricky physics. Be patient.

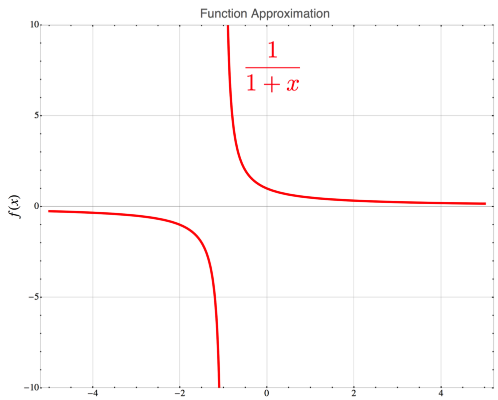
additional term *adds less and less* to the sum before it. Now we've got a little approximation-tool because many formulas that matter in physics look like

$$\frac{\text{something}}{(1+\text{ something tiny})^{\text{ some power}}}$$

or can be rearranged to look like that.

Here's one that we'll use. Let's imagine the function

$$f(x) = (1+x)^{-1} = \frac{1}{1+x}.$$



Let's even plot it, which I've done in Fig. 2.15. Notice that this function becomes infinite when $x = -1$ and that it quickly falls until $x = 0$ and then slowly heads off towards zero as x becomes very large. That makes sense, right?

Now lets expand that function according to the approximation in Eq. 2.15. For this particular function, $n = -1$ and we will keep just the first four terms of the otherwise infinite number of terms:

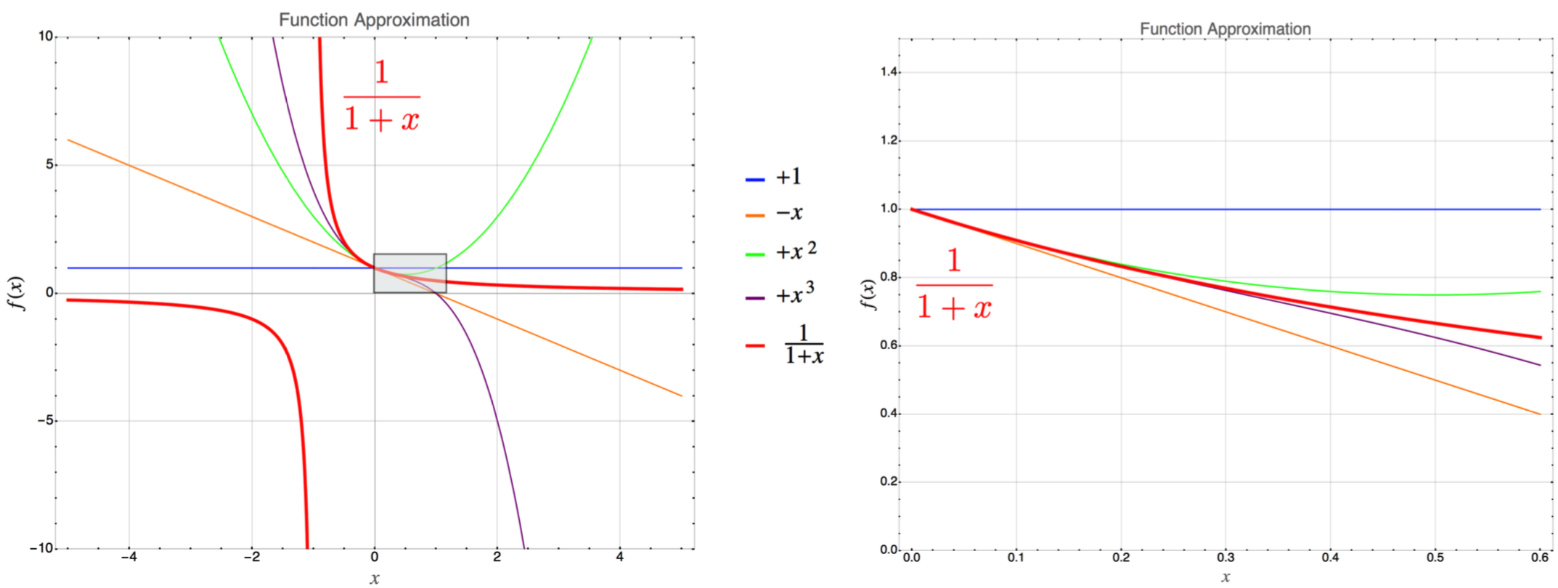


Figure 2.16: See the text for an explanation. The right plot is a blow-up of the left around the gray box.

$$f(x) = \frac{1}{1+x} \approx 1 - x + x^2 - x^3 \tag{2.16}$$

(By the way, the ≈ symbol in Eq. 2.16 stands for "almost equal to.") The right hand side of this equation is really the sum of four different, simple functions. When added together, we'll see that they get closer and closer to the original, depending on how many terms are included. Look at Fig. 2.16. The red curve in the left and right plots is our original function and the colored curves are each getting closer and closer to it.

The blue "curve" is the trivial function that's the first term in Eq. 2.16: $f = 1$. The orange curve takes the second term in Eq. 2.16 and adds it to the first, so it's $f(x) = 1 - x$. The green curve adds the third term, x^2 to the orange curve and so on. The right plot is a blowup of the region in the gray box on the left. Notice that in the region of x which is very small, the few functions are a pretty good approximation to the red. The more terms we might add the further out in x that agreement would continue.

Remember this! It will become important later when we'll encounter functions and approximate them with a few terms of the expansion from Eq. 2.15. Here are the functions that we'll see in the pages ahead:

$$\sqrt{1+x} = 1 + \tfrac{1}{2}x - \tfrac{1}{8}x^2 + \tfrac{1}{16}x^3 - \ldots \tag{2.17}$$

$$\frac{1}{\sqrt{1-x}} = 1 - \tfrac{1}{2}x + \tfrac{3}{8}x^2 - \tfrac{5}{16}x^3 + \ldots \tag{2.18}$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \tag{2.19}$$

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + \ldots \tag{2.20}$$

2.7 Euler's Number

You all know that π is an unusual number. It's simply the ratio of the circumference of a circle to its diameter (see Eq. 2.7) and, the Indiana Legislature[34] not withstanding, it's a number that has a decimal representation that never ends. It's "irrational" and has the (approximate!) value:

$$\pi = 3.1415926536\ldots \text{ forever!} \tag{2.21}$$

There is another irrational number that plays a big role in mathematics, but also in many other areas of "regular" life. It's called "Euler's Constant" although the prolific mathematician Euler didn't first discover it, he discovered many of its unique features and so his name is associated with it. We physicists tend to just call it "e" since that's the symbol that is used to represent it. It has the value:

$$e = 2.71828182845904523536\ldots \text{ forever!} \tag{2.22}$$

[34] Yes, that story is true. In 1897 state legislature representative, Dr. Edward J. Goodwin, a physician who dabbled in mathematics, proposed changing the value of π to 3.2. The bill sailed through the House but was postponed indefinitely in the Senate. It seems that Professor C.A. Waldo at Purdue was horrified enough that he intervened and the bill died.

Euler first used e to understand compound interest. If you invest \$1 at a compounded interest of 100% per year, then at the end of the year your wealth would have been increased by a factor of e. While not many savings plans grant 100% interest, you get the point. It figures into the calculation of any interest rate. I'm going to try to convince you that it appears in many guises.

The importance of e in science comes from the fact that the rate at which e increases or decreases is proportional to *itself*. So if something increases by e^{ax} then the rate at which it increases is ae^{ax}. This leads directly (with some calculus) to the rule for how radioactive nuclei, atomic systems, or elementary particles decay. Suppose we start out with N_0 radioactive nuclei with a "lifetime" called τ at a time $t = 0$, then the number of left after a time t is equal to

$$N = N_0 e^{-t/\tau}. \tag{2.23}$$

So the fraction left is $\frac{N}{N_0} = e^{-t/\tau}$. Figure 2.17 shows two curves for both the exponential decay and

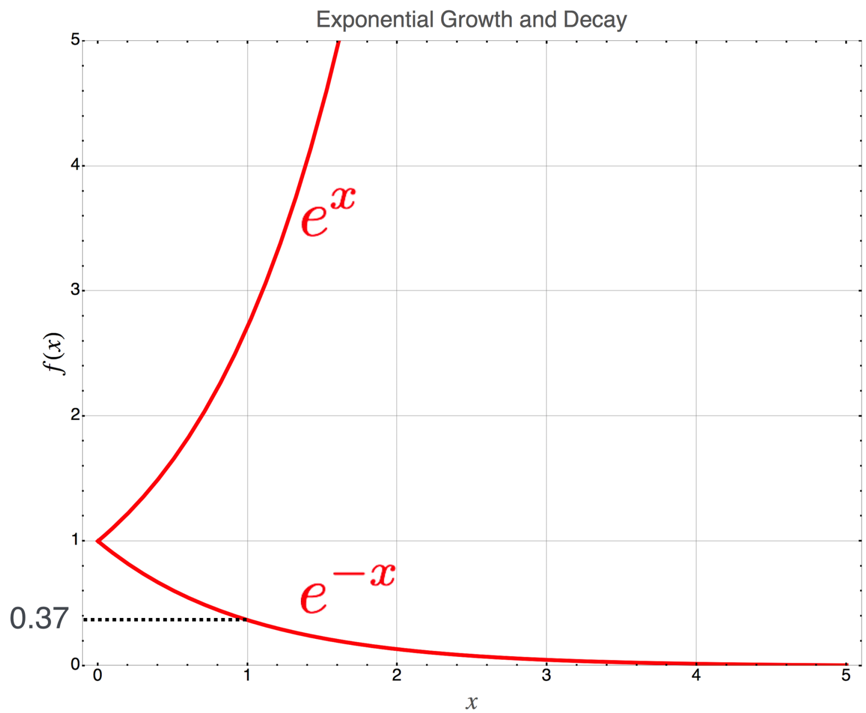


Figure 2.17: exponentials

exponential growth formulas.

But it's not only some sort of modern physics thing. Atmospheric pressure decreases the higher up you go...this is because there's less air above you. So home runs in Denver's Coors Field go further than in Chicago's Wrigley field since Denver is about a mile higher than Chicago. We could pretty closely calculate the density at any altitude using this same formula, but modified for the physical situation. Let's call the density of air at any height above sea-level (y) to be $\rho(y)$. Then if we let $\rho(0) \equiv \rho_0$ then the function that describes the density at any height turns out to be

$$\rho(y) = \rho_0 e^{-y/8000}. \tag{2.24}$$

where the distance above sea level, y is measured in meters. Let's do one more thing and then we can use our curves, even though the axes are just relative numbers. So we could directly ask the fractional change in density:

$$\frac{\rho(y)}{\rho_0} = e^{-y/8000} \tag{2.25}$$

Relative to sea level, then a mile high (1,609 m) makes the right side $e^{-(1609/8000)} = e^{-0.2}$ so we can use the general graph in Fig. 2.17 since we've determined that $y = 0.2$,[35] At that value, read across, we see that the density is reduced to about 80% of what it would be at $x = 0$. So,

$$\frac{\rho(y)}{\rho_0} = 0.8. \tag{2.26}$$

[35] Of course, we're using y in the formula for height, which is often a convention, but it's still playing the role of the x in the general graph.

Not everything in nature decays! Suppose you're a biologist studying bacterial growth. If a particular strain grows continuously at a rate of 5% per day, you could predict the size of the colony after some number of days.[36] The growth in the colony where t is measured in days is given by

$$F(\text{bacteria}) = F_0 e^{Rt} = F_0 e^{0.05t} \tag{2.27}$$

[36] Or, you could measure the increase and write the function that describes it.

where F(bacteria) is the number of bacteria after a time t and F_0 is the number that you started with. For a different bacterium, R would be a different number (a "rate"). If we waited patiently for about a month, say $t = 30$ days, we'd have

$$F(\text{bacteria in a month}/F_0 = e^{Rt} = e^{(0.05\times 30)} = e^{1.5} \tag{2.28}$$

Back to Fig. 2.17 with $x = 1.5$ the top graph reads about 4.4. So if we started with a population of 100, after 30 days it would have grown to $4.4 \times 100 = 440$.

This is what people mean when they refer to "exponential growth"—a very rapid increase in some phenomenon.

2.8 Vectors

We're about to talk about motion, but let's make an important point here that will be obvious. When you're driving on the highway and your (American) speedometer reads "60 mph," it's telling you the *speed* not your direction. Going 80 mph north is as much over the speed limit as going 80 mph east since speed is all the highway patrol radar cares about. (There isn't one speed limit for easterly travel and another for when the road bends north.)

The cops might not care, but you care a lot whether you're traveling north at 60 mph or east, since in order to get where you're going on schedule–your trip depends not only on how fast you go, but in what direction. The difference between *speed* and *velocity* is critical. Not all quantities are vectors...for example, what's the direction of a temperature? But, velocity, space coordinates, force, momentum, electric and magnetic fields, and many other physical quantities have directions as well as values.

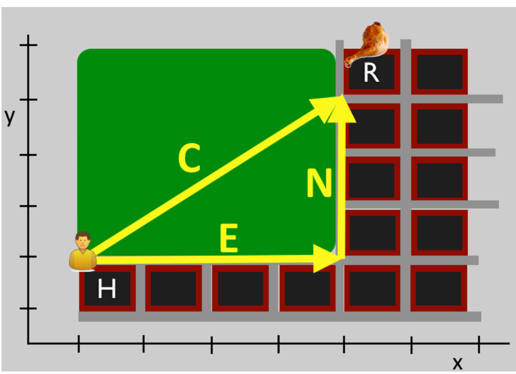


Figure 2.18: The layout showing my hotel (H), the restaurant (R) where there is fried chicken waiting, and the city block structure.

A vector has both a magnitude and a direction — *Key Concept 3*

There's an algebraic way to represent vectors, but we'll not need that. Instead we'll make use of the handy symbol of an arrow: →. The length of the arrow represents the magnitude and of course the orientation and the head of the arrow represent the direction. Arrows can be ⟶, or short →, pointed in different directions, ↖, ←, ↗, etc. Very handy. The magnitude can mean many things, depending on the physical quantity being represented. Obviously, the simplest would be a distance in space, like an arrow on a map or a whiteboard during time-out. That's it.

Here's a way to think about them. Suppose you're in a strange city and you want to know how to get from your hotel to a particular restaurant. You go to the front desk and you're told that you need to walk for 7 blocks, Terrific. Now what? Seven blocks that way? Or, seven blocks the other way! Rather, "walk 4 blocks, east and then 3 blocks north" is more helpful, as you can see in Fig.~\ref{blocks}. (It's just like velocity.)

Now we can go around writing "four blocks east" (or "60 mph north") everywhere, but we need a better notation that packs both *directional* and *magnitude* information into a single symbol so that our hotel-restaurant stroll east is succinctly distinguished from one to the west (and so we don't need to use words in our equations). Traditionally, in print, a vector is represented as a bold letter.[37]

Notation in equations is fine, but pictures of vectors are going to be most useful for us. It's easiest to think in terms of distance vectors. Just like "speed" and "velocity" are related, we can think of "distance" and "displacement" as analogs. So, our hotel tells us that the restaurant is a *distance* of 7 blocks away

[37] There are at least three ways that I can think of to represent vectors. In print, the bold face **x** is most common. On a blackboard, usually people will draw an arrow over the top, $\vec{x}$. And, finally, some people put an underline when they write, $\underline{x}$.

and that its *displacement* is "4 blocks, east and 3 blocks north" and we draw a picture to describe that instruction. Figure 2.18 shows two vectors that do that:

2.8.1 Vector Diagrams

Drawing arrows on a diagram represent a vector with its orientation representing the direction and its length representing the magnitude. Sometimes the length of the arrows are actual length dimensions (like meters, feet, and so on), since a displacement in regular-space is a vector. So, just like a scale on a map, a displacement can be represented as an arrow which is 3 inches long, but where each inch actually corresponds to 1 block (or feet, or miles, or furlongs). But, sometimes a vector doesn't represent a length in space, but some other physical quantity, like a force or a velocity. Now, this can be complicated since you're drawing an arrow that has a length, but you mean it to be something else, like a force. But, it still works geometrically (the arrow still points in space) and we just use a different scale: we might draw an arrow aimed at a box on a diagram that's 2 inches long where every inch corresponds to 2 pounds. So even though it's drawn on a diagram of an object, it represents the application of a force of 4 pounds applied at the point where the arrow is drawn. That's just a visual convenience since the length of the vector in pounds wouldn't have anything to do with any of the length scales in the picture that are lengths or heights.

For a couple of definitions, refer to Fig. 2.19. There are two basic ways to represent vectors, one for print and the other for blackboards (or pencils). The print version is to render the vector quantity as a bold letter. So in Fig. 2.19 the vector on the top is in print **A** and on paper we would write $\vec{A}$.

Two vectors, **A** and **B** are said to be equal if they are *both* the same length *and* point in the same direction. So, as shown **A** = **B**, but neither is equal to **D** even though the length of **D** is the same as that of **A**. Also, we say that **A** = −**C** if the vectors have the same length, but are pointing in exactly the opposite directions. This is shown in Fig.~2.19b. Another standard definition is to represent the magnitude of a vector–its length–using the symbol |**A**|. This quantity is a number, not a vector and so we would say that |**A**| = |**D**|.

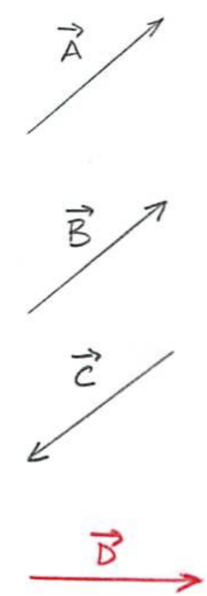


Figure 2.19: Vectors **A** and **B** are equal, and each is equal to −**C** and none are equal to **D**, even though the lengths are all same.

2.8.2 Combining Vectors

If you help me to push on my car, we're each applying a force. The whole reason for the two of us is not so we can bond in a shared accomplishment. That's not a guy thing. No, the reason we do it is that we each supply a force and the car then gets pushed with more force than either of us could supply by ourselves.

That is, our forces add...and maybe we bond a little. So, vectors can be added both in symbols, and with pictures.

We can add vectors together by manipulating the arrows. If in our little moment together, I'm **A** and you're **B** then, the car gets pushed by our combined force as shown in Fig. 2.20(a). However, the car would not know the difference between being pushed by the two of us and by some brute who pushes with the force of our combined effort, which we'll call **C**.

$$\mathbf{C} = \mathbf{A} + \mathbf{B}. \tag{2.29}$$

Pencil 2.3. ✎

To calculate this using pictures, you can place the tail of **B** to the head of **A** and then the displacement from the tail of **A** to the head of **B** is the sum, **C**. This is shown in Fig. 2.20(b), and the replacement of the two forces is shown as Fig. 2.20(c). It's important to realize that the situation (a) and (c) are identical, but you would not put both$ **C**$ and the two **A** and **B** on the same picture. It's one or the other.[38]

[38] Dare I carry my little story this far? It's as if I push on the car, and you push on me. If my arms hold up, we still push on the car with the combined force. But, I'd rather not do it that way, thanks.

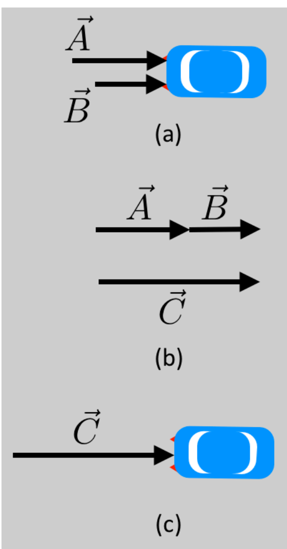


Figure 2.20: (a) Both of us pushing on a car; (b) the combination of our two force vectors; and (c) the replacement of our two independent forces with the combined force. The car doesn't know the difference between (a) and (c)!

Notice, that for doing sums, we can translate vectors around our "space" if we don't change their orientation or length. I did that in the figure.

The car example was all in one dimension, but of course vectors are useful in 2, 3 or more dimensions. Let's go back to our trip to the restaurant from our hotel. What I didn't know, was that there was an open park just behind my hotel, and I could have cut across it to get to the restaurant. That is, an equivalent displacement would have been to follow **C** as shown in Fig. 2.18. That's all the adding of vectors says: a single vector that's equivalent to the operations of the first two. So my trip has two different paths (well, an infinite number):

$$\mathbf{C} = \mathbf{E} + \mathbf{N}$$

Notice that the two vectors don't point in the same direction, so it would be wrong to calculate the distance that **D** represents by just adding the lengths of **E** and **N**. That is, the magnitude of **D**, $|\mathbf{D}| \neq 4+3$. We have to keep the directions and the lengths pointing in their directions separate.

One more way to look at this trip–which resulted in a nice dinner, by the way–would be if we returned to the hotel across that field, then our trip would look like Fig. 2.21.

Notice, that it's different from Fig. 2.18 in that **D** points in the opposite direction from **C**. It's a "round trip" and so the total displacement in a round trip is: zero. In algebra, what this says is:

$$\mathbf{A}+\mathbf{B}+\mathbf{D}=0$$

Any time you can rearrange a set of vectors to give a "round trip," you describe a situation in which there is no net displacement (we went from the hotel, back to the hotel), or if they are forces, no net force, or if they are velocities, no net velocity. It's a balance $\mathbf{A}+\mathbf{B}$ is balanced by its opposite, **D**. The other way to think of this is remembering that we could have gone to the restaurant across the field if we'd known about it. Notice, that then the vector describing that trip would be $-\mathbf{D}$. We replace $\mathbf{A}+\mathbf{B}$ with $-\mathbf{D}$. And, the balance is just the obvious: $-\mathbf{D}+\mathbf{D}=0$. This balancing of vectors will be an important concept to us as we'll see in Chapter 6.5.

Finally, we can also subtract vectors graphically which is easiest to think about if we think about this almost silly statement:

$$a-b=d$$
$$a+(-b)=d$$

This says that the adding the negative of b to a is the same as subtracting it from a. With vectors, this is a little more meaningful. Referring to Fig. 2.21, let's create a vector subtraction.

$$\begin{aligned}\mathbf{C} &= \mathbf{E}+\mathbf{N}\\ \mathbf{D} &= -\mathbf{C}\\ -\mathbf{D} &= \mathbf{E}+\mathbf{N}=\mathbf{C}\end{aligned}$$

So, we change a subtraction of vectors into an addition of vectors by just turning the appropriate one around.

In order to make the negative of a vector, turn it around and reverse its direction. *Key Concept 4*

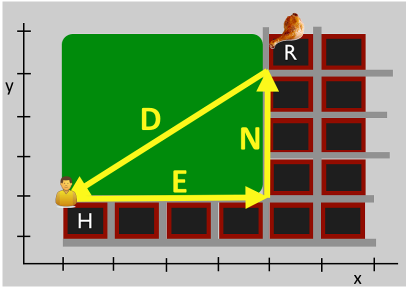


Figure 2.21: The same situation as before, but with the hotel-restaurant trip shown and the restaurant-hotel return shown on the same picture.

2.9 What To Take Away

> "...it is impossible to explain honestly the beauties of the laws of nature in a way that people can feel, without their having some deep understanding of mathematics. I am sorry, but this seems to be the case.
>
> "You might say, 'All right, then if there is no explanation of the law, at least tell me what the law is. Why not tell me in words instead of in symbols? Mathematics is just a language, and I want to be able to translate the language.' ... I could convert all the symbols into words. In other words I could be kind to the laymen as they all sit hopefully waiting for me to explain something. Different people get different reputations for their skill at explaining to the layman in layman's language these difficult and abstruse subjects. The layman searches for book after book in the hope that he will avoid the complexities which ultimately set in, even with the best expositor of this type. He finds as he reads a generally increasing confusion, one complicated statement after another, one difficult-to-understand thing after another, all apparently disconnected from one another. It becomes obscure, and he hopes that maybe in some other book there is some explanation...The author almost made it—maybe another fellow will make it right.
>
> "But I do not think it is possible, because mathematics is not just another language. Mathematics is a language plus reasoning; it is like a language plus logic. Mathematics is a tool for reasoning."
>
> Feynman, R.P. (1965) *The Character of Physical Law* BBC. Reprinted by Penguin Books, 1992

Part I Physics and Cosmology of my Grandparent's Generation

Chapter 3

Motion

Getting Around

Galileo Galilei, *circa*1624.

Galileo Galilei, 1564-1642

"I do not feel obliged to believe that the same God who has endowed us with sense, reason, and intellect has intended us to forgo their use." *Letter to the Grand Duchess Christina*

As he got older, his mouth got him in more and more trouble until he was imprisoned in his own home for the rest of his life. But by then, he'd created physics and defined diverging paths for religion and science. Galileo's life can be segmented into four distinct periods: his young life, education, and university employment at Pisa; his second job at the University of Padua in the Venetian Republic; his return to Florence as the Chief Mathematician and Philosopher to the Grand Duke of Florence; and then his house arrest at his villa outside the city gates. We'll follow his scientific path from falling bodies, to astronomy, to his method of doing science. When we're done, physics will have been born.

3.1 Goals of this chapter:

- Understand:
  - How to calculate distance, time, and speed for uniform and constantly accelerated, linear motion
  - That falling objects all have the same acceleration near the Earth.
  - How to graph simple motion parameters
  - How to read graphs of realistic motion parameters
- Appreciate:
  - The algebraic narratives in the development of the formulas
  - The shape of the trajectory of a projectile
  - That a projectile's motion is made of two components with different accelerations
- Be familiar with:
  - Ideas of motion before Galileo
  - Galileo's life
  - Galileo's experiments with motion

Figure 3.1: The tomb of Galileo Galilei in the Basilica of Santa Croce

3.2 A Little Bit of Galileo

The original Florentine Galileo was a 15th century medical doctor and civil leader of the family Bonaiuti. So significant was this elder Galileo that the subsequently middle-class family renamed itself Galilei and our Galileo Galilei's two identical names was a subtle parental reminder that he was expected to do great things.[1]

Galileo was born in Pisa within a year of the death of Michelangelo and educated at the newly restored University of Pisa. He always considered himself a citizen of Florence, although he lived there only briefly in his early years. His father was a musician—of necessity, a wool merchant in his wife's family business—and determined that his son would be a medical doctor. But as a student, he was disrespectful of his professors as an innate skeptic regarding the natural philosophy taught which conformed to the European standard: Aristotle. What Aristotle said about motion had not made sense to anyone for centuries, but his authority was almost absolute. Galileo—like his father—didn't "do" authority.

While a medical student, Galileo accidentally discovered that he had an aptitude for mathematics that led him to an intense but clandestine program of the study of Euclid with Ostilio Ricci, the Court Mathematician to the Grand Duke of Tuscany. He eventually abandoned medicine, leaving the university a year short of his degree. By this time he was doing original mathematical research (in geometry) and

[1] So venerated, the original Galileo was buried in the Basilica of Santa Croce, eventually the resting place of Machiavelli, Michelangelo, and a pantheon of Renaissance personalities. . . and eventually, our Galileo as well as shown in Fig. 3.1

had caught the attention of scholars in Pisa and Rome. He lived at home for three years and gave private mathematics lessons in Florence and Sienna[2] while he cultivated patrons for help in finding a university position. After some rejections, he succeeded...back at the University of Pisa as a lecturer of mathematics. He wasn't altogether welcomed by his former teachers.

[2] He became a passionate follower of Archimedes' mathematics and invented a more precise way to measure the density of metals, following in his hero's wet footsteps. He also gave invited lectures at the Florentine Academy on the geometry of Hell from Dante's Inferno

His reputation as an original mathematician was growing when his father died and he inherited the responsibility of a significant dowry for one sister and responsibility to provide for the other's. Galileo spent the rest of his life in search of a higher salary, which as a lowly mathematician at Pisa was a factor of four smaller than that of a philosophy professor. He got his break when he was offered the position of Professor of Mathematics at the University of Padua, among the most prestigious universities in Europe and safely in the progressive Republic of Venice. It was at Padua where the magic happened.

3.3 From Here to There

This is important:

Almost everything in physics boils down to: motion.[3] *Key Concept 5*

[3] ...even "boiling"!

Whether it's runners on a track, the cosmic rays piercing us all the time, orbiting planets, electrons in a wire, electromagnetic waves, quark wavefunctions inside of a proton, electrons and holes in a semiconductor, or the stretching of spacetime itself. Everything is about motion.

These first chapters on the physics of my grandparents' generation will establish our language and tools that we'll need in order to pursue the more exotic forms of motion and we'll become skilled at manipulating concepts (and their attendant symbols) like velocity, kinetic energy, mass, momentum, and force. Each of these terms has a 16th to 19th century origin, but each has managed to keep up with the times as layer upon layer of subtlety is discovered about each of them as we dig deeper and discover more.

But at its most basic, it's all about how to get from here to there or from then to now, and to be able to explain how that happened.

3.3.1 A Greek Version of Here to There

The correct understanding of everyday motion was long incoming. Really long. Classical works had been out of reach of Europe until Greek philosophy and science essentially fell into their laps in the form of hundreds of conflicting Arabic translations in the 1300s. Aristotle—eventually referred to as "The Philosopher" —had invented formal logic that taught people how to evaluate arguments. But while the Philosopher's methods were refreshing, his ideas about motion were confused at best—but nonetheless became firmly stuck in the academic and religious communities where they were protected as *philosophy*, not as *science*.[4] Not until the end of the 16th century did Galileo shed Aristotle and lay the groundwork for the first systematic understanding of what it means for something to move. It took three centuries!

For Aristotle, motion[5] was of two sorts: natural and unnatural. Natural motion near the Earth was in a *straight line*, either *down* to the center of the Universe (which he located at the center of the Earth, proportional to the amount of "earthy" composition of the object, and hence its weight) or *up* (proportional to its lightness).[6]

Natural motion beyond the orbit of the Moon was to be *circular* with every extraterrestrial body attached to its own rotating crystalline sphere.

> **Wait.** *Why would they insist on circles for the stars and planets?*
>
> **Glad you asked.** *If you go out on a dark night and watch the motions of the stars in the north...you'll convince yourself that they are moving in circles around Polaris. You'd be wrong about the North Star's involvement, but you'd be pretty sure: circles. So were they.*

These spheres were all nested with common centers and rotated around the Earth to account for the apparently circular orbits that we see from the Earth. They included all of the known planets, the Sun, and the Moon...and even the stars in the outermost shell. Natural motion just happened...naturally, but unnatural motion required a pusher...an active force that was in contact with the object. Therein lay one of the most obvious flaws in his model.

Make no mistake, translation of Aristotle's *Physics* from original Greek, to Arabic, and then to Latin did not make his ideas any less confused than they originally were. Where he got himself into big trouble was with projectiles, like a thrown spear. Since for Aristotle the philosophy came before observation, he had to do an embarrassing dance to explain that when a rock was thrown, the continuous "push" that Aristotle insisted was needed came from the displaced air rushing around behind the rock and pushing it forwards.[7] Everyone knew that this was nonsense, but his authority reigned, and organization of the first medieval universities with Philosophers and Theologians at the top guaranteed that natural science was taught by them and not by the mathematicians and astronomers, whose roles were aimed at more mundane activities like casting horoscopes and designing military weaponry and fortifications.

Galileo was the one person who changed the landscape and uncovered the modern notion of how things move, ridding the intellectual community of Aristotle's baggage. He began his revolution while he was at Pisa where he wrote an unpublished manuscript, *de Motu* ("On Motion"). While he was unsatisfied, one of his conclusions was right on: all objects fall at the same rate, contrary to Aristotle's insistence that

[4] While Aristotle stumbled with physics and astronomy, he really was amazing. He practically inventing biology, zoology, anatomy, psychology, logic,ethics... the list goes on and on.

[5] To Aristotle, objects in nature "moved" according to causes and one had to beat one's common sense into submission in order to allow his explanations of everyday events into the mainstream. Motion for him was a very general thing: anything that changed in time, like when an seed grows into a sapling and then into an oak tree... is "motion." The kind of motion that we think about was "locomotion."

[6] Actually, he classified four kinds of elements: earth, water, air, and fire. Each had its natural place and substances went to that natural place according to the mixture of the qualities of the elements.

[7] Yes. He actually suggested that.

heavier objects fell faster. His data? Not the Leaning Tower of Pisa. That's a myth. He just looked around him. And saw things differently.

3.4 Speed in Modern Terms

To us, motion and its measure—speed—is a simple matter. Our cars and even devices on our wrists readily tell us how far we go, how long it takes us to get there, and the rate at which we do it. In fact, we can be penalized for traveling on our roads at rates that are...too enthusiastic. Speed, or its more sophisticated word-cousin, velocity, is so familiar to us that we hardly pay any attention to just how fundamental this concept is. Let's start this very slowly since some of our more sophisticated physics later will build on a strong underpinning of a few basic concepts. Speed is one such concept because it's a blend of two even more fundamental concepts of space and time.

Just wait until Albert Einstein gets his hands on it.

Speed is a rate. For example, 60 mph describes the change in our spatial position and how long it took to make that change. Like all rates, it's a ratio with respect to time:

$$\text{speed} = \text{"the change of distance divided by the change in time"} = \frac{\text{change of distance in space}}{\text{change of time}} \tag{3.1}$$

Wait. *Everyone knows this. We all drive and we can calculate how long it takes to get home. So why the fuss?*

Glad you asked. *Apart from the fact that trying to understand motion took almost 1500 years, we continue to misunderstand it over and over, as you'll see. We need to start gently with ideas that will seem trivial. But hold on to your hat, since it will get weird. I'll remind you that you thought that this definition was silly.*

Let's make this more compact by inserting customary symbols to get rid of the English words. Here are the rules of the use of motion in QS&BB:

Pencil 3.1.

- We'll limit ourselves almost exclusively to motion in one dimension in space.
- We'll use the symbol v for speed (because customarily, we'll speak of "velocity"... more about this below).
- We'll use the symbol x for distance in one dimension, regardless of which direction it points.
- We'll use the symbol t for time and almost always presume that we set our clocks so that the beginning time of any interval is $t_0 = 0$.
- Oh, and we'll use the subscript $_0$ to indicate the beginning of some time interval "t_0" or location "x_0" in a sequence of events.
- We'll use the Greek symbol Delta, Δ to mean "change of"... this will come up a lot.

Definition: Δ.
means "change of."

So our formula for calculating speed changes from the English sentence in Eq. 3.1 to a mathematical statement becomes:

$$v = \frac{\Delta x}{\Delta t} \tag{3.2}$$

Definition: Velocity..
Velocity (or speed) is the rate of change of distance.

Equation: Velocity.
$v = \Delta x / \Delta t$

Definition: initial quantity.
We will always put a little subscript 0 to indicate that some quantity is the "initial state" of a process. So, x_0 would be the initial position, E_0 might be the initial electric field, and so on.

It's important to think about how we would measure any quantity. For speed, we'd need something that functions like a ruler in order to measure a distance in space and something that indicates time intervals—a clock. So if we're on the interstate, we could imagine the mile-markers along a highway that tell us miles or better, small fractions of miles. We could arbitrarily designate a starting point as $x = 0$ and then, after speeding up, travel—without acceleration—until some pre-determined time, say 2 hours, had elapsed. Of course we would measure time with a clock in the car that begins ticking when the car passes our starting point. Then after precisely two hours had elapsed the clock would cause a camera to take a picture of the mileage sign that was closest to being opposite the car. Let's suppose that the sign read "100 miles." Without putting pencil to paper (this time!) you could quickly calculate the average speed that the car traveled in that time as 50 miles per hour.

Likewise, if I asked you how long it would take to travel north from Detroit 300 miles to the Mackinac Bridge at an average 50 mph, I'll bet you could tell me. You have done this sort of calculation a thousand times and so you would calculate:

You Do It 3.1. Travel60mph

Calculate the time it would take in hours to travel 120 miles at an average speed of 60 mph

or copy the solution

Did you get two hours? Your brain is already doing physics. Let's go "up north."

3.4.1 Calculating a Speed

"Change" and "change-of" always means the difference between where you *are* as compared with where you *were*. Suppose I start out with \$100 and my wife gives me \$50. The change in my net worth is \$50, right? But we can represent this simple transaction as

$$\Delta(\text{my wealth}) = \text{where I ended up} - \text{where I started} = \$150 - \$100 = \$50.$$

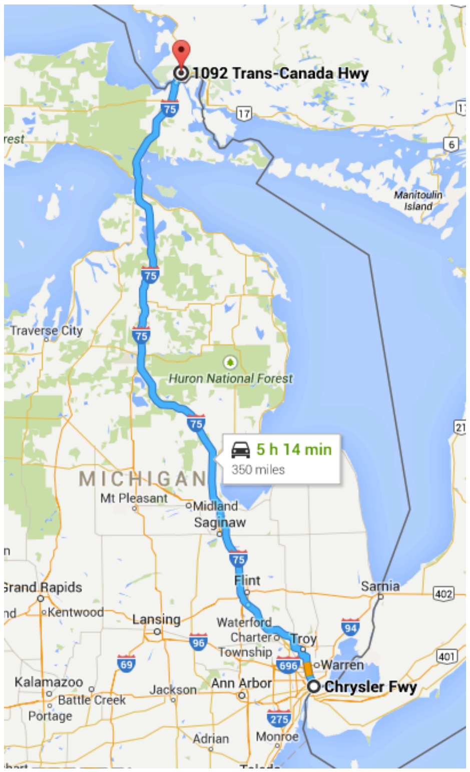


Figure 3.2: A trip from Detroit to Canada.

Remember that we're using the standard notation in which the "initial state" of any quantity will be decorated with a little "0" subscript, like x_0 here. The "final state" will have no subscript and just be x.

In our trip from the bottom to the top of Michigan, we will stay on I-75 as in Fig. 3.2, beginning at the nearest Comerica Park exit in Detroit which is near mile marker 50 at Grand River Avenue. Then we'll go all the way up, across the Mackinaw Bridge to Newberry, Michigan in the Northern Peninsula which we'll say is marker 350. We'll go fast.

So the change in my displacement is

$$\Delta x = 350 - 50 = 300 \text{ miles.}$$

Now, we can write the real velocity relationship:

$$v = \frac{\Delta x}{\Delta t} = \frac{x - x_0}{t - t_0}.$$

where I snuck in a "initial" decoration for the beginning time as well. If I start my clock ($t_0 = 0$) at mile 50 and stop it at mile 350, our average speed is:

$$v = \frac{\Delta x}{\Delta t} = \frac{x - x_0}{t - t_0} = \frac{350 - 50}{5 - 0} = \frac{300}{5} = 60 \text{ mph.}$$

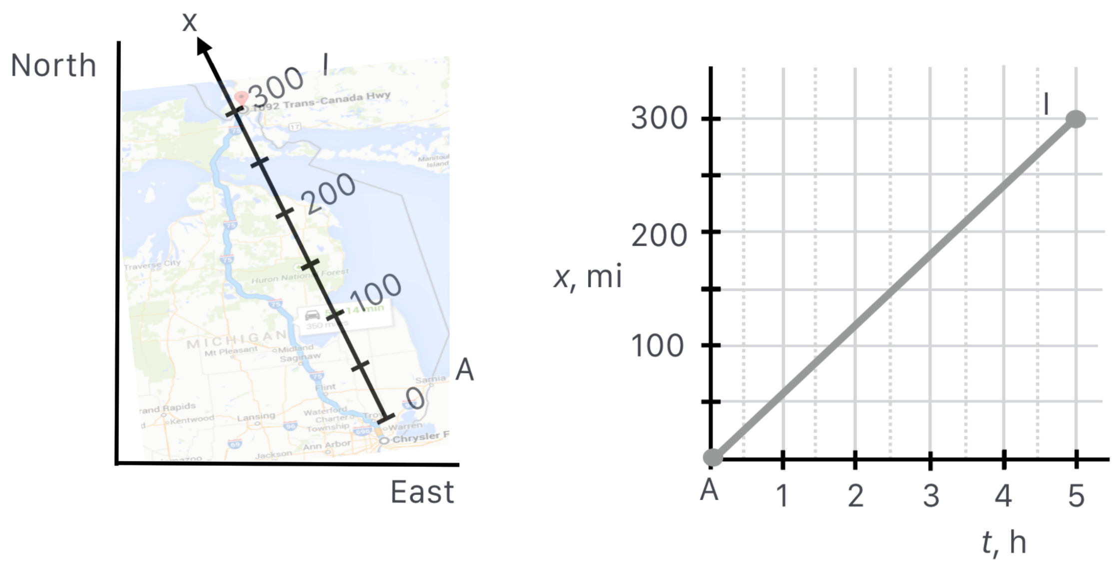
3.4.2 Diagramming Motion

We will need a variety of graphical ways of representing motion...from here to there. A perfect example of such a representation is when you draw a route on a map. The map has "space axes" of east-west (x) and north-south (y) and when you go from one town to another you might draw a along your route in space, recognizing that each mark corresponds to a different time as you move along the road images. So time is represented implicitly on such a graph.

3.4.3 Space Space Graphs

Let's take a realistic trip. A regular map, like in Fig. 3.2, or the display on a GPS system, is a familiar way of looking at travel. Let's make an approximation to that curvy map trip by straighteining out all of the road's curves and bends so that it looks like the approximate straight line in Fig. 3.3 a. Time is still implicit and the single coordinate is a space direction. Notice that I've labeled the axis along our straightened-out route the x axis, increasing from Detroit ($x = 0$, called A) to Newberry $x = 300$ miles, called I).

as that famous crow flies?



(a) (b)

Figure 3.3: Two approximate views of the trip in space (a) and space-time (b).

3.4.4 Spacetime Graphs

Now let's represent the trip in a different way using axes that aren't space "and space" but space and time. In this trip I drove and you fell asleep when we left Detroit and woke up five hours later when we arrived—you should go to bed earlier. You looked at your watch and saw that 5 hours had elapsed and looked at

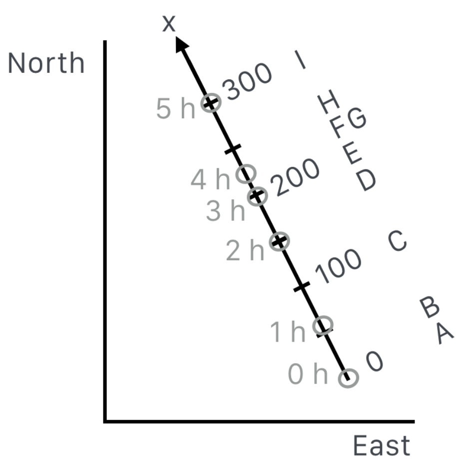


Figure 3.4: A trip from Detroit to Canada "on a napkin." Another, more life-like view of the trip. The open circles are drawn every hour and are not evenly spaced, indicating that the velocity changes. So the Fig. 3.3 b picture is an approximation and simply a global average. The labels refer to places where the speed changed as described later in the narrative.

the odometer and saw that 300 miles had been traversed and you calculated that your trip's speed was $\frac{300}{5} = 60$ mph. You even drew the spacetime diagram for your idea of the trip in Fig. 3.3 b.

First, notice that your simple speed calculation corresponds to the finding the slope of this line. Also notice that you shifted the beginning point to define our distance origin, where $x = 0$. In this case:

$$v = \frac{\Delta x_{AI}}{\Delta t_{AI}} = \frac{300 - 0}{5 - 0} = 60$$

which is a fine thing to have done.

At any point during your nap, the slope of that space-time trajectory would the same as at any other point. So any region in which you calculate the speed by evaluating

$$v = \frac{\Delta x}{\Delta t}$$

over and over gives you the same value. Let's rearrange things slightly and get a little equation...a predictive *model* of our motion:

$$\Delta x = v\Delta t. \tag{3.3}$$

You give me a time, and using the model, I'll tell you where you are. It's a little physics machine.

Remember the equation of a straight line with a slope of m and which passes through the y axis at b (the "intercept")? Sure you do. It's

$$x = mt + b.$$

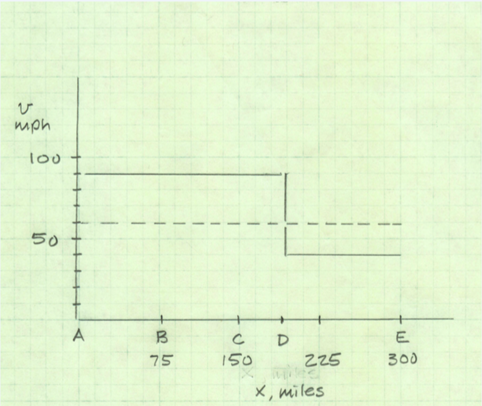
Our spacetime trip plot fits this form with a zero intercept and a slope of v.

Gotta stop to eat once in a while.

Example 3.1

Racing to the bridge.

Question:



Let's suppose that we are able to get out of town early in the morning when the roads are empty. Not that I'd ever do this, but we'll start our trip at a steady speed of 90 mph which we can keep up for 2 hours before traffic slows us down. Then we travel at a slower speed and find that we still arrive 5 hours after we began.

- How far were we able to go before we had to slow down?
- If the overall trip took 5 hours what was our average speed in the second, 3 hour segment?
- What was the overall average speed?

Solution: The first segment, 90 mph for 2 hours, means that we went 180 miles before hitting the brakes. If we traveled for 5 hours in total, then the second segment took 3 hours and the distance left was $300 - 180 = 120$ miles. So the average speed in the second segment is

$$v = \frac{120}{3} = 40 \text{ mph.}$$

Finally, the overall average speed is still the total distance divided by the time that it took. Still 60 mph. Figure 3.1 shows the speed profile as the solid pair of curves. Notice that this is not the average of the averages. There is more distance covered by the fast trip segment than the slower trip segment.

3.5 Acceleration

Figure 3.3 b showed that in one interpretation of our trip we never deviated from 60 mph. Can you drive like that? I can't and I didn't! While you were asleep, I sped up, slowed down, and stopped for sushi. You calculated an *average* velocity for the whole trip, which doesn't care what happened between the beginning and the end.

Nothing beats gas station sushi.

Let's be a little more realistic and invent a trip profile shown in Fig. 3.4. . It's like the previous version but in the open circles I've added where we actually are at each hour...if you'd been awake, you'd have realized that we went pretty fast between about hour 1 and hour 2. You were sound asleep when I stopped for that meal at the 4 hour mark. And so on—the constant-time interval hour marks are closely spaced (meaning little ground covered, and so slow) and spaced more apart (meaning lots of ground covered, and so a high speed).

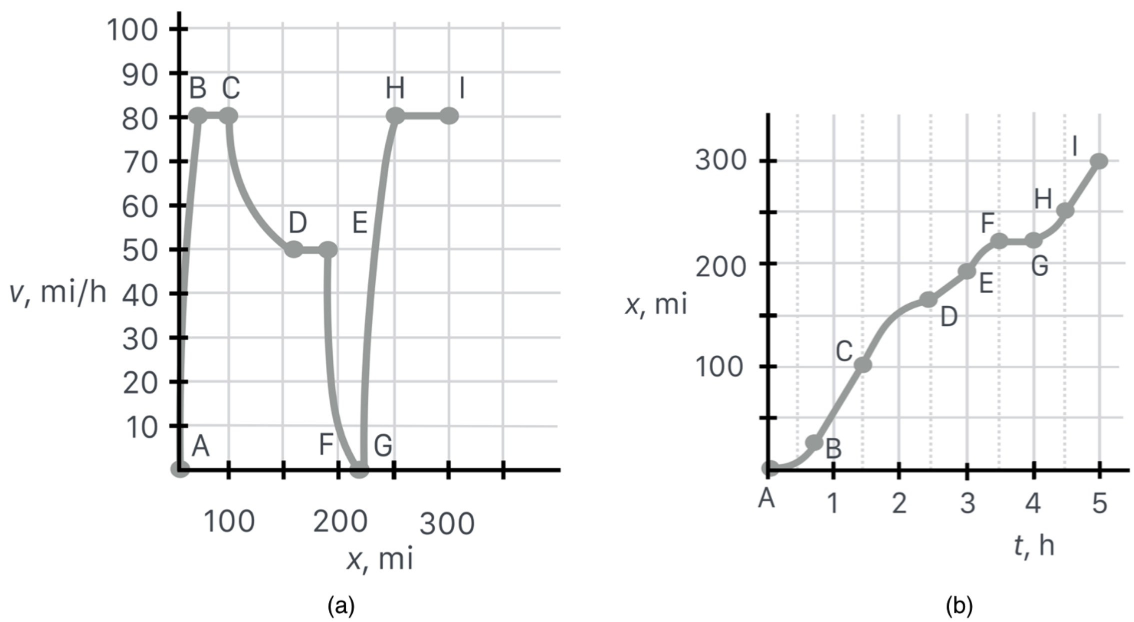


Figure 3.5: The trip is further diagrammed showing how the velocity changed as a function of distance (a) and how the distance traveled changed as a function of time (b). The curved lines indicate where acceleration has taken place and will be described below: for a constant acceleration, the velocity changes like the square root of the distance and distance varies as the square of the time interval.

Figure 3.5 tells the whole story. With a little bit of artistic license, Fig. 3.5 a shows the speed at each point along the road:

- We started at A and accelerated to B, from 0 to 80 mph.
- We drove at that speed from B to C for a while and then started to reduce speed—must have seen a highway patrol car.
- So from C to D we slowly reduced our speed for traffic to 50 mph and held it steady until E when I got hungry.
- So at E we started to slow for an exit and stopped for snack at point F.
- When we got back on the road I accelerated back to 80 mph, to H and stayed at that speed until we got to our destination at I and you woke up.

Figure 3.5 b shows that same trip, but now plotting the space time representation. You can find each of the segments and compare them. From B to C, for example we're traveling fast (the $x-t$ slope is steep) while from D to E, we're traveling slower (the slope is less). From F to G, we didn't change our position at all, but time elapsed while I enjoyed my raw fish.

Finally, for completion, Fig. 3.6 shows how our speed changed as a function of time.

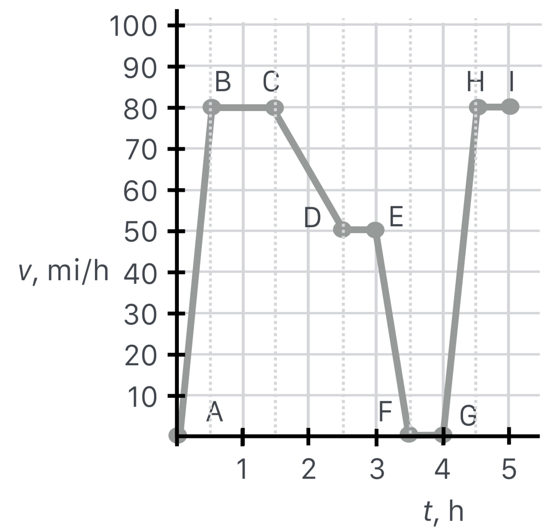


Figure 3.6: A more realistic speed profile for our trip.

Any time a speed changes is an **acceleration**. And just as velocity is the rate at which distance changes, acceleration is the rate at which *velocity* changes. So it's defined similarly:

$$a = \frac{\Delta v}{\Delta t}. \tag{3.4}$$

Let's look at our trip before and after the sushi break and concentrate on how the velocity changes in time in Fig. 3.5. Putting the meaning of Δ back in explicitly, let's look at G-H, when we're refreshed from our sushi break:

$$a = \frac{\Delta v}{\Delta t} = \frac{v_H - v_G}{t_H - t_G} = \frac{80-0}{4.5-4} = \frac{80}{0.5} = 160 \text{ mph per hour.}$$

Where the speed was constant, like B to C, the slope of the $v-t$ graph is zero, so there's no acceleration at all, or more correctly, $a = 0$.

How about the interval C-D? First, notice from the graph that it's going to have a negative slope, the opposite of the slope we just calculated. Explicitly, we can evaluate:

$$a = \frac{\Delta v}{\Delta t} = \frac{v_D - v_C}{t_D - t_C} = \frac{50-80}{2.5-1.5} = \frac{-30}{1} = -30 \text{ mph per hour.}$$

What's the significance of the negative sign? Well, for one thing whenever an acceleration is negative it means the object is slowing down. In vector-world, it means that the direction of the acceleration vector is the opposite from the direction of the motion.

Definition: Acceleration..
Acceleration is the rate at which velocity changes. If it gets higher, it's called acceleration and if it gets lower, it's called deceleration.

The units of speed are easy to remember because we use them every day. They're units of distance divided by units of time, or *per* unit time: miles per hour, feet per second, meters per second (the standard in physics), or kilometers per hour (if you drive in Canada).

Equation: Acceleration.
$a = \Delta v / \Delta t$

But the units of acceleration, while simple to figure out, are a little unusual since we don't use them in everyday life. From Equation 3.4 they would of course be units of *speed* divided by units of time, but since the units of speed are distance per time, then the units of acceleration would be distance per time squared: "meters per second per second"[8] is what we might say out loud, meters/seconds^2 or m/s^2 is

[8] or feet per second per second, or miles per hour per hour

what we'd write. That explains the units from our two intervals as mph per hour.

At any point along the distance-time plot we could determine the "instantaneous speed" by carefully drawing a line that's tangent should the velocity change be varying (not a straight line).[9] Your speedometer is not doing that but rather it's measuring the average speed in time increments that are so small, that to you it looks like you're receiving a report of your speed "right now," but it's not instantaneous in reality.

[9] Remember that a line that just touches a curve perpendicular to it at a point is "tangent."

Let's take a bike ride.

Graphical Kinematics

Let's work out an extended example with realistic numbers. Follow along with me because if you become comfortable with this, what follows will be smooth sailing.[10]

[10] . . . er. . . biking.

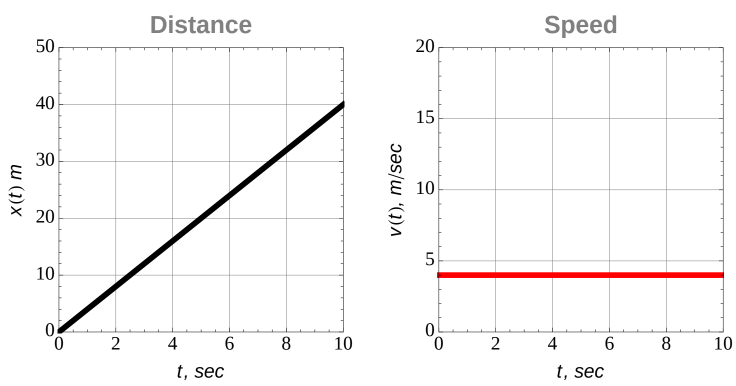


Figure 3.7: The right hand curve shows a constant speed of 4 m/s, holding steady for 10 s. The left hand curve shows the distance that an object will travel at that constant speed as a function of time.

Let's think about a cyclist moving at a constant speed. These two pictures in Fig. 3.7 describe all that you need to know about an object moving at a constant 4 meters per second (about 9 mph—a sprinter's speed, or a relatively slow cyclist's speed). Here we take the initial position to be zero (it could be our house). And, we take the initial time as zero (it could be when we leave). So, to find the distance traveled after 6 seconds by an object moving at that speed we would look at the graph at $t = 6$ s and read about 24 meters. Or, equivalently, we could use the formula and calculate:

$$v = \frac{x}{t}$$
$$x = vt$$
$$x = (4\text{m/s})(6\text{s}) = 24\text{m}.$$

Equation 3.3 has that geometrical meaning we spoke of. Referring to Fig. 3.7, in the first second, in the right plot I see that my speed is 4 m/s and that I traveled about 4 m (left plot). In the next second, my speed is *still* 4 m/s (right plot) and I travel another 4 m (left plot). The *slope* of the quantity x plotted as a function of time t is the constant $\Delta x/\Delta t$ which I pointed out is the *slope* of the curve. That's easy since the speed is constant.

Now, suppose I speed up to pass the cyclist in front of me. My speed increases and I cover ground at a faster rate: during each time interval, I travel more than the time before. Of course now I'm accelerating if I'm going faster during each interval ("decelerating," if I'm going slower). In either case, the state of my motion *is* changing from my previously steady speed. As we'll see, if the state of my motion is changing, there's a force: which I've applied by pumping my legs faster and faster, which is transmitted as a push on the road (Earth) through the rubber tires.

Figure 3.8 shows the predictive power of kinematics when applied to *constantly* accelerated motion. The curve on the right shows a *fixed acceleration* of 2 m/s^2. If this were a car, it's like your foot is on the accelerator in order to maintain a constant force and hence, a constant acceleration on your bike, you need to overcome friction and wind, so maintaining a constant force on the pavement is hard work. This constant acceleration means that the velocity is changing at a constant rate—proportional to time—and that's shown in the middle plot. So, at 6 seconds, your acceleration is 2 m/s^2, and your velocity or speed has increased to 12 m/s from the start. In the mean time the distance you traveled is going up faster: it increases proportionally to the square of the time and this is shown in the left hand plot. If you started

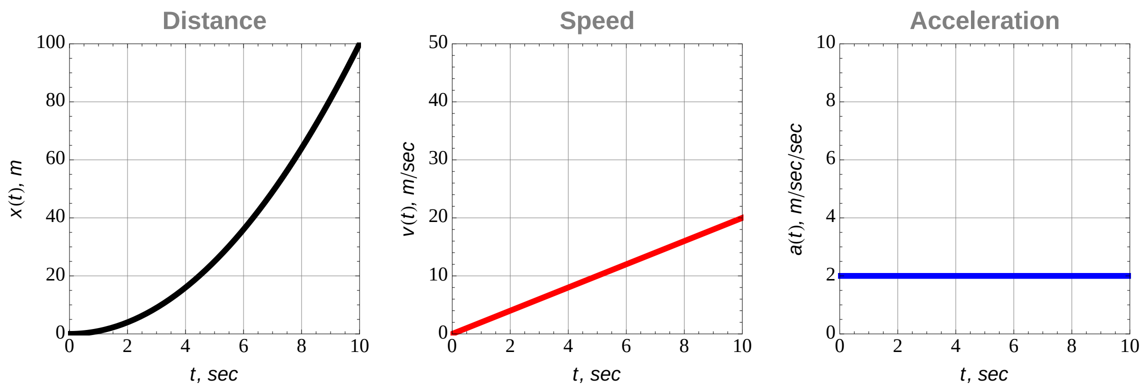


Figure 3.8: This shows the result of a constant acceleration of 2 m/s^2. Each plot is as a function of time: the acceleration, is independent of time—a straight line. The velocity is then a constantly increasing function of time, and the distance increases as the square of time.

from rest, at $x = 0$, by the time 3 seconds have passed, you've traveled 9 meters, and by a total of 6 seconds, you've gone 36 meters, so your distance intervals are increasing with each increment of time.[11]

[11] That's moving right along for a bike—about 25 mph.

Finally, thought of geometrically, the slope of the middle plot is constant—the change of speed with respect to time increases at a steady rate, the value of the constant acceleration. But, now the slope of the left hand plot of distance changing in time is *not* constant—at each successive time interval, it's steeper and that's reflected in the fact that the middle plot value changes at each time and so *instantaneously* (the tangent, if you want to be fancy), the speed is different and so the distance increases.[12]

[12] Notice that the maximum distance in the right hand figure is 100 m, reached in about 10 seconds. Usain Bolt's world record is 9.58 s.

So for the constant acceleration (right hand plot, a = constant), which results in a steadily (linear, $v \propto t$) increasing speed (middle plot), the changing distance covered per unit time increases a lot: the shape of that curve (right plot) is a parabola, $x \propto t^2$:

$$x = \frac{1}{2}at^2. \tag{3.5}$$

Definition: *propto.*
In an equation, *propto* means "approximately equal to."

Equation: **Distance for constant acceleration.**
$x = \frac{1}{2}at^2$

Observing this parabolic, or quadratic, increase in distance with respect to time is the smoking gun for constant accelerated motion.

Now go back to our trip up north and in particular, Fig. **??**. We can see that those constantly accelerated regions are parabolas, turned down for slowing down between B and C, and turned up for constantly accelerating between D and E. Again, that deceleration at the end is not a parabola, so it's not a constant deceleration as can be seen in Fig. **??**

3.5.1 Special Kinds of Acceleration

We can categorize acceleration into three separate categories with what we now know.

Zero Acceleration

No acceleration indicates a constant velocity.

Positive and Negative Acceleration

Positive and Negative accelerations mean very different things here and we'll illustrate this with a simple pair of examples. When you leave the city limits, your speed goes from 35 mph to 55 mph, so the numerator[13] in the definition of acceleration in Eq. 3.4

$$\Delta v = v - v_0 = 55 - 35 = +20 \text{ mph}$$

and this would be a positive acceleration (when you divide by the time it takes), or just "accelerating." But when you come into town from the highway, your speed goes from 55 mph to 35 mph and the numerator in the acceleration equation

$$\Delta v = 35 - 50 = -20 \text{ mph}$$

would be a *negative* acceleration, which is called a *deceleration*. So acceleration is negative or positive by virtue of whether the object is slowing down or speeding up.

[13] Of course the time difference is always positive.

Definition: deceleration.
A negative acceleration which indicates that the speed is getting smaller: slowing down.

Varying Acceleration

One more time for Fig. **??**. At any particular point in time, the value of this graph gives the *instantaneous* speed while the slope at any point gives the instantaneous acceleration. Instantaneous quantities were unimaginable before Isaac Newton, so this would have been a confusing thing even for Galileo. Let's think about what it means to take an average...a simple one, of two quantities.

Suppose we want the average height of two people, one of whom is 5 ft tall and the other, 6 ft tall. You'd calculate:

$$\text{average height } = \frac{5+6}{2} = 5.5 \text{ ft.}$$

The same thinking applies to motion—if, and only if the acceleration is constant. So, in our trip, if at some point we were traveling at 80 mph when we saw the state police car (point B) and 50 mph when we slowed down (point C), what was our average speed during that time? Just like the average of heights, it would be

$$\text{average speed} = \bar{v} = \frac{1}{2}(v + v_0) = \frac{50+80}{2} = 65 \text{ mph},$$

Definition: **average symbol .**
We represent an average quantity with a bar over the top, like $\bar{A}$

where I've introduced another notation that we'll sometimes use: an average of a quantity is represented with a bar over the top.

The average deceleration during that slow-down from Gaylord, MI at point F is over a long time interval, and you'd watch your speedometer steadily reporting a lower and lower speed. It's really not reporting an instantaneous speed, since the electronics reporting what you read on the speedometer is really calculating an average speed over very small digitally-computed time windows.

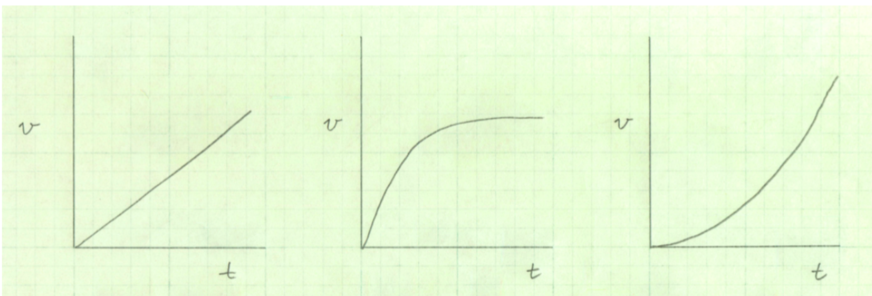
We could calculate over a smaller time interval, and then a smaller one and of course are we really even going precisely at an absolutely constant speed? That is, we could imagine decreasing the time interval all the way to zero, at which point we're doing calculus. We don't need to do that, but we should remember that for any finite time interval, all of the displacement formulas are really averages. But the fact that the slope of the velocity curve in the F to G segment of Fig. **??** is of all different values shows that this is not constantly accelerated or constantly decelerated motion.

Cars are not the only place where we encounter non-constant accelerated motion. How about running? A sprinter comes out of the blocks with a very high change of speed, and so a very high acceleration. As she gets into her stride and become more erect, her effort must fight against wind resistance and besides, her internal quick expenditure of energy would be hard to maintain, so there's a *reduction* of acceleration. Note this doesn't mean a reduction in velocity, just that the velocity increases at a slower rate as the race goes on. Nor does it mean a deceleration...she's still going faster, just at a lesser rate than before.

You Do It 3.2. Sprinter Graph

Which graph of velocity versus time for a sprinter matches the description of a runner from the text?

or copy the solution



(a) (b) (c)

The figure that you just chose shows the increase of speed of a sprinter: start fast and then level off in almost constant speeds. Figure 3.9 shows how the designers in *Madden Football '11* modeled the rate of speed of one of the players in the game.

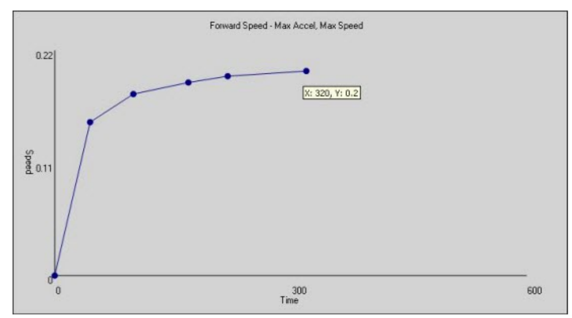


Figure 3.9: Modeling a running back in Madden NFL11 (http://www.operationsports.com/ncaa/utopia/topic/69752-madden-nfl-11-locomotion/

Constant Acceleration

From the above figures and our experience it's plain that everyday motion would likely involve accelerations that vary in time—like the slope of the realistic sprinter graph above— changes almost throughout the run, starting out very high and becoming smaller. But historically and practically there is a special, constant, acceleration.

3.6 Galileo and Free Fall

Galileo is revered in physics for essentially creating two modern sciences (physics and astronomy), and for enunciating the principles of how science should work that are still relevant today. Very early on he recognized that the unnatural/natural motion notions[14] of Aristotle were wrong. For example, Aristotle insisted that two objects made of the same material but of different weights would not hit the ground at the same time if dropped. The heavier one would be more "eager" to go to its natural place (the Earth's center) than the light one. Or an object that was more earthy (like a rock) would fall faster than an object that's less earthy (say a stick) even if they weighed the same...again, the more earthy the object, the more anxiously it tended towards the center of the Earth which was...well, totally earthy. Galileo simply watched that hail stones of differing sizes seemed to hit the ground at nearly the same time which was not consistent with the Aristotelian ideas, and that caused him to "think different."

[14] I actually wrote "motion notions"?

Wait. *This is trivial! You mean nobody else noticed this in nearly 2000 years?*

Glad you asked. *You're right. People did question Aristotle's concepts about motion but not seriously until the beginnings of they medieval university system in Europe. Everyone knew that there were problems with Aristotelian science but the overbearing importance of Aristotle's whole philosophical package and its merging with Catholic theology carried out by St. Thomas Aquinas in the 13th century, meant that questioning it was a hard sell. Indeed, to pick away at just the motion part of his philosophy would weaken other pillars of his system—like the famous Four Causes—and that would have been hard to bear.*

It had "been in the air" for a century that the speeds of falling objects increase either in proportion to increasing *distance* or increasing *time*. All one had to do was watch water drips from a roof...the distances that they travel in a given time are much longer near the ground than when they first leave the edge. So they are accelerating, but which way? Long before Galileo, smart philosophers at Paris and Oxford realized that if, during equal time intervals, the distance of a falling object increased like the squares of the time interval...then the speed must increase proportionally with time, This was pretty sophisticated geometrical reasoning for people without algebra, graphs, or graphical representation of functions. Remember, from what we know from the above discussion, this quadratic time dependence told him that gravity produced a constant acceleration.

Galileo first started to think differently about motion when he was a professor at Pisa. That's where he was supposed to have dropped a wooden and iron ball from the Leaning Tower, although we know that he couldn't have done it the way the story went. His unpublished pamphlet on motion was a mixture of old language and his attempts at a new one, and so was not quite there yet.

3.6.1 Pisa To Padua

His salary was abysmally low and he was clearly ready for a more lucrative position. So after three years in Pisa, when he was offered a job at the University of Padua he took it and stayed for 18 years until 1610. Not only was his salary higher, the Venetian Republic[15] was a much more independent region of Italy. Indeed, unlike Florence in Tuscany where connections to Rome and the Papacy were long-standing, Venice was often in the Papal Doghouse for its independent ways.[16] Here Galileo was a popular professor, he created tools for military use that he was able to sell, including a manual. He settled into a common-law relationship with a woman with whom he had three children.[17] But money was always problematic and they even took in students for tutoring and rent into their crowded home.

While Galileo was in Padua, he resumed some of his earlier experiments with moving objects. His original ideas were unformed but he reworked them as he shed his Aristotelian influences. There he had a workshop and eventually hired a toolmaker.

3.6.2 The Pendulum

A pendulum is a trivial toy, yet it encompasses many key physics ideas about motion and force. Galileo, Huygens, and Newton each made many pendulums and did extensive experiments with them. Because of its immediate simplicity, quantitative measurements are easily done with modest tools. It will come up over and over! Figure 3.10 is a picture of a simple pendulum which I'll refer to periodically over the next few chapters.

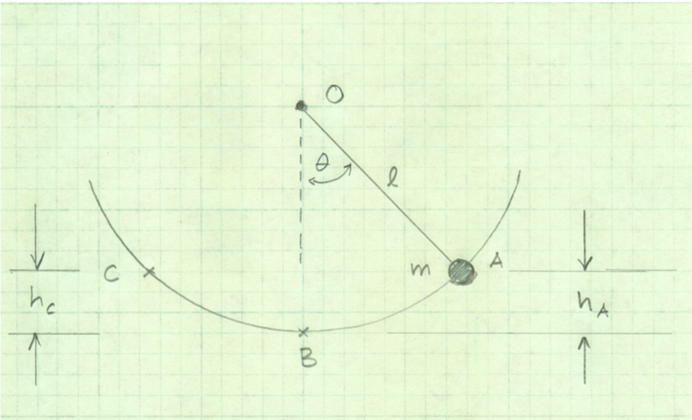


Figure 3.10: A simple pendulum which oscillates in a vertical plane (that is, gravity acts down). There are three points indicated on the circular path that it makes: A is the point of release, a height h_A from the lowest point, B. C is the point of the highest elevation which it obtains on the other side and that height is labeled as h_C. The pendulum bob has a mass of m and the length of the string connecting it to its point of suspension, O, is ℓ. The angle that the string makes with the vertical is θ.

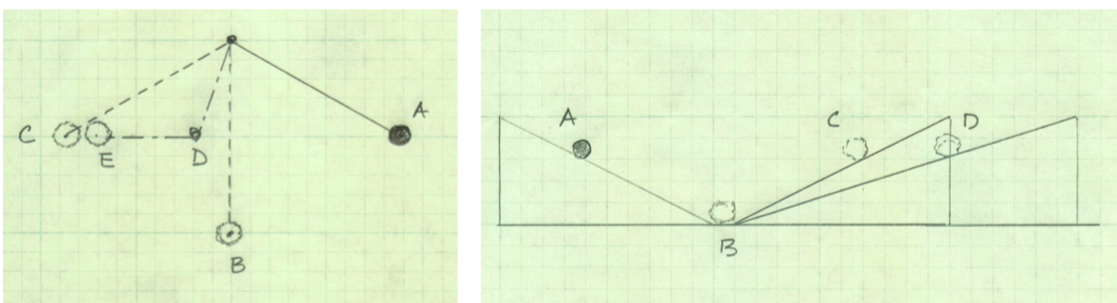
[16] Popes railed against Venice, sometimes excommunicating everyone, sometimes even going to war, but Venice was the most powerful naval power facing the troublesome East and so Rome needed Venice periodically.

[17] For a professor to remain single was not unusual and Galileo was a frequent visitor to Venice and its literary and art community. This was a sophisticated society with thousands of highly literate courtesans. While little is known about the early relationship between Galileo and Marina di Gamba, some have speculated that she might have been such a professional companion. Galileo stayed with her for almost a decade which might suggest that they were intellectually compatible—indeed, given his personality, it would have been surprising were he to spend time with someone not up to his conversational standard. In any case, Gamba moved to Padua to live with him where they had three children, two girls and a boy. (Birth records do not refer to a father.) When he went back to Florence in 1610, he took Livia and Virginia with him (aged 9 and 10) but left four year old Vincenzio with Gamba who remained in Padua and eventually married. He installed the girls in convents, where they would cost no dowry, and brought his son to Florence. Eventually, Vincenzo became a musician and was partially supported by the Pope, as a deference to Galileo. Virginia, later Sister Maria Celeste, relied on him for his influence in supporting her convent. She died in 1633, the year he entered house arrest as an old man. He was devastated.

Among the surprising facts about a pendulum, no matter where you release the bob, it takes the same amount of time to make a complete back and forth. That is in Fig. 3.10, $h_C = h_A$. Except that's not true. Only for very small values of θ (which is the same as making h_A relatively small) does this approximately work. For Galileo, the differences were small enough that he presumed that the pendulum's motion is isochronous, that is, that it swung back and forth in equal times.

Galileo first studied the pendulum in Pisa. The initial striking result that he noted was that the bob will make a complete arc, following a circle, sweeping through point B and come to rest to essentially the same height from which it started, at C. He pushed against this result by putting a peg in the way at D and the bob still returned to the same height, now at E as in Fig. 3.11a. The second thing he noticed was that the time that it takes to make one compete oscillation from the higher point *is the same* as when it was started from the lower position. In fact, he found that the *period of oscillation*—how long for a back and forth trip—only depends on the length of the cord. Not the mass of the bob and not the material of the bob.

This led him to make the leap from pendulums to motion in general and in Padua he began to experiment with inclined planes. Suppose you have two inclined planes opposite one another, like in Fig. 3.11b and you start a ball rolling from the top of the left-hand inclined plane at A, it will fall to B and then rise to pretty nearly the same height on the right-hand one at C. Like the pendulum.



(a) (b)

Figure 3.11: The right-hand figure shows various positions on a pendulum bob which is released from rest at point A. At B it's moving to the left and comes to rest at point C. Point D is a peg that's inserted in the path of the string and causes a sharp change in the trajectory...but it sill comes back to the same height, now at E. The left-hand figure shows a set of inclined planes and positions of a ball released from rest at point A, which passes through B on its way back up the first plane, rising to the level at which it was released at point C. If the slope of the plane is decreased, the ball still rises to the same height, now at D...but further along the incline in order to get back to that height.

Now, make the angle of the right hand plane smaller so that the ball has further to go to reach the same height, now at D. He reasoned, it still would and showed that. Now, here's the genius part: suppose you lower the right-hand plane more and more until it's flat. What does the ball do now? Galileo said it would roll forever. He then packaged that reasoning into a statement that Newton appropriated nearly 60 years later: that an object that's started in motion on a flat plane will continue forever, unless something stops it. No pushing involved. Once it's started, it goes. Notice how far behind Aristotle looks in Galileo's rear view mirror! He's not worried about natural or unnatural motion and he's not concerned himself with *why* the motion continues. Rather he's asking different questions...*How* questions rather than *Why* questions. Aristotle's students would start with a philosophical prejudice and interpret what they saw in accordance with that philosophical system. Galileo threw all of that aside and observed nature without bias or preconceived notions. So if you release the ball in Fig. 3.11 *where would it go?* is his question. Not, *why* would it go. And he'd set up the circumstance and observe it—no Aristotelian would do an experiment. They would think about it in the specific context of the philosophy and passively observe.

The other important thing is that Galileo was fully aware that there were a whole host of impediments to a pendulum going back and forth forever, or even coming back to the original point after one swing. In fact—*and this is important*—he imagined that the real rules of pendulum motion were those of the zero friction limit and that our actual pendula are incremental modifications to the true situation through successive addition of friction, air resistance, stretching of the string, etc. Likewise, with the rolling ball struggling up the right hand incline there are overlays of real impediments to the otherwise perfect motion. The bob nor the balls really, exactly came back to their starting points...just really close. Plus nearly every event would be slightly different. But rather than dwell on the differences, Galileo was the first to decide that there was *something that was the same in every event.*

3.6.3 Free Fall

The big question for Galileo (and many others) was to explain the behavior of falling objects. What rules govern how fast an object falls. Aristotle said that an objects "nature" was the cause and that its determined how fast. Clocks didn't yet exist, and so even Galileo couldn't drop something and measure the time for it to fall, even from very high places.[18]

[18] His observations of the timing of pendulums' periods led to Christan Huygens' invention of the pendulum clock a generation later.

What Did He Do?

But he invented a trick...a way to dilute gravity so that he could make timing measurements using 16th century tools. Suppose instead of dropping a ball, you rolled it down an incline from the same height. He hypothesized that whatever pull a ball felt in free-fall would still be there in the slow descent down an incline. If he made the slope so shallow that the ball would roll ("fall") slowly enough, he could measure it and sneak up up on the rate of falling directly. He found that inclines of only two or three degrees were ideal. Figure 3.12 shows the setup.

Three different distances are indicated for the ball. The ball starts from point A and rolls down the plane to point D along the path ℓ taking some time to get there. We can ask about the separate "motions" in the horizontal and vertical directions since the path is in the plane. So along the x axis, the ball goes from point A to point B and it "falls" in the vertical direction along the y axis from point A to point C. A little bit of trigonometry shows that the rate that the ball travels along ℓ is the same as the rate to fall along the y axis. So all he needed to do was measure how far the ball rolls along ℓ in each increment of time. That same rate would be the free-fall rate. Pretty clever.

Galileo was good with his hands, and armed with the understanding that he needed to reduce the effects of friction and other extra impediments to determining the real rules, he built inclines with finely machined grooves to minimize the effects of friction[19] and then made elaborate schemes to measure the time that it took for an object to "fall" as it rolled slowly down the incline. He used his pulse. When that wasn't sufficiently regular, he hired a string quartet to play at an even meter. He then built a "water clock" that slowly dripped at a regular rate and it was with that, that he was able to see how far a ball would roll during each "tick" of his water clock.

[19] I'm being a little coy here. "Rolling" only happens when friction is at work between the ball and the surface. Think about it. If there were no friction, the ball would slip and it would not translate. What he reduced was irregularities in the slot that the ball rolled in.

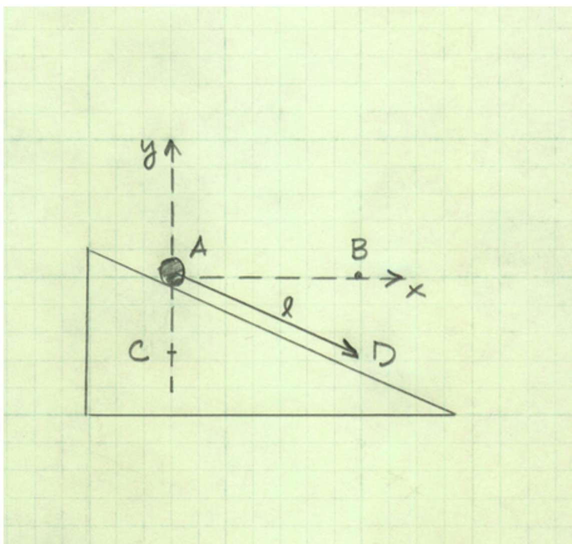


Figure 3.12: Various positions of a ball released at A measured along the incline and separately along the horizontal and vertical directions. See the text for the story.

Galileo kept careful records of his experiments and these have been preserved. Fig. 3.13 shows the aha! moment in one of his pages where he's taking his measurements, scribbling around the sheet, and keeping his records in the upper left hand corner. The late Stillman Drake from the University of Minnesota, the preeminent Galileo scholar, has studied these notebooks and on this particular page, he noticed that the ink used in the column with the squares was different from the ink in the actual data-taking. In fact that ink was consistent with pages that came from entries many days later. Clearly Galileo had taken the data, then pondered the results and realized later that his measurements fit the square-time rule and came back to the page to add them in. Imagine what his emotional state must have been when he realized what he'd done!

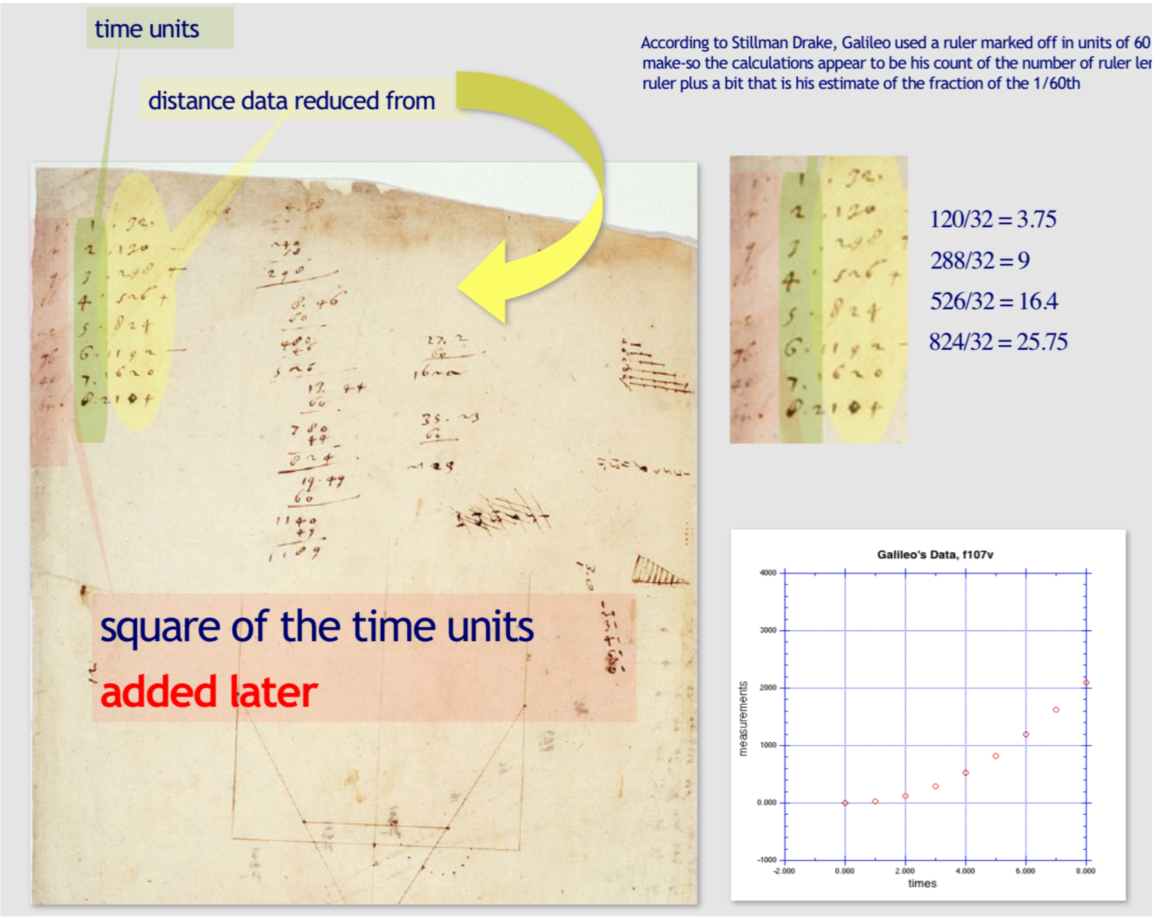


Figure 3.13: Galileo's notes of his inclined plane data, including his "aha moment" where he returns to it after a few days and adds the time-squared calculation.

Little g

What he was discovering was the shape of the left graph in Fig. 3.8 (plotted for his data in Fig. 3.13 as the white inset) that the distance is quadratic in the time traveled, like we saw in Eq. 3.5. Quadratic? He knew that meant a constant acceleration is at work for his rolling objects. He then went one step further. By measuring the results with balls of differing materials and weights he found that they all showed the same result.

Now the big leap...so to speak. He reasoned that the same would be the result of dropping objects vertically. Casting aside 1,500 years of Aristotle's insistence that the heavier, most "earthy" objects would fall much faster, Galileo insisted that all objects would fall with the same acceleration and that the distances would increase like the square of the time.

Today we write:

$$x = \frac{1}{2} g t^2 \tag{3.6}$$

where we give this special, constant acceleration a name, the "acceleration of gravity," g, or "little g"[20] which is roughly 32 ft/sec/sec, or 9.8 m/s^2 which was first measured by Isaac Newton a generation later...using pendulums.

[20] Is it called "g" because of "Galileo"? What do you think?

Objects near the Earth's surface fall with essentially constant acceleration. *Key Observation 2*

Constant of nature: **Acceleration due to Earth's gravity at its surface.**
$g = 9.8$ m/s^2

He then made yet another important leap: In practice, a heavier object might hit the ground slightly before a lighter one, but he correctly reasoned that this was because the effect of the resistance of the air would be a larger retarding force on the little ball than on the heavier one.[21] Now he's doing physics. He was after Nature's "real rule" which was hidden behind the apparent motion, affected by the air. He was exploring what's the same about falling objects, not what was different about them.

[21] Test this yourself. Take two identical pieces of paper and crumple one of them into a tight ball and the other only slightly. Drop them side by side and what happens? The air resistance matters a lot.

Air Resistance

Little g is a large acceleration, but raindrops don't usually bruise us and hail stones don't usually kill. Were there not for air resistance, nice spring showers would be dangerous as rain drops fall a long way! It turns out that things dropped from high up reach a speed in which the viscous drag of air friction pushes back with the same force that the Earth pulls, causing a falling object to reach equilibrium. The result is that the speed of falling becomes constant to what's now called the Terminal Velocity, which depends on an object's size.

For example, a regulation major league baseball will reach terminal velocity in about two and a half seconds after falling about 100 ft. At that point, its speed will be just about 60 mph and will stay at that value. For reference, the "Green Monster" left field wall in Fenway Park is about 40 feet (11 m) high, so you can see that high fly balls would be high enough to reach terminal velocity on their way down.[22]

[22] There have been stunts performed with baseballs dropped from the Washington Monument, skyscrapers, and balloons for players to catch. The speed that a baseball would achieve is not much more than that experienced by a catcher on a high school baseball team. However, famously a professional catcher knocked himself out missing a ball dropped from a blimp.

The saga that he dropped balls from the Leaning Tower of Pisa while a faculty member there, surrounded by all of the members of the student and faculty body was told by one of his ardent followers after Galileo's death. There is no evidence that this public event ever happened, and Galileo himself never described doing such an experiment. This same disciple also told a story that as a child he sat in the Cathedral of Pisa and while bored during mass timed the candelabra overhead and from that experience reasoned his pendulum rule later. The tale had young Galileo making these measurements with a candelabra that were not installed in the cathedral until many years after the event was said to have taken place. So there's much to be wary of in Galileo-lore as told by his young admirers.

While Galileo was on the faculty at Padua, he did other experiments with motion which lay dormant until in 1638 he wrote the second of his great books, *Discourses and Mathematical Demonstrations Relating to Two New Sciences.* He finished this work while in house arrest outside of Florence as an old, sick man. How he got into trouble is a story that we'll leave for our discussion of early astronomy.

The Equations of Constant Acceleration

Acceleration at g is called... well, "gees." And "pulling gees" is a measure of acceleration that fighter pilots must contend with: in order to not black out, humans can tolerate accelerations up to around 6 g's, or $6 \times 9.8 \text{ m/s}^2$ = about 60 m/s^2. That's moving right along as you will see from the Porsche-experience below.

The plots for motion had their origins with Galileo, but their algebraic form waited for the 17th and 18th Centuries to become standard. Our formulas so far can be summarized here: [23]

[23] I've added one for constant acceleration that is easily derived from the others, but I've not done so here. Trust me. I've also generalized the equations that I did use to include the possibility that the initial conditions might not be when $x_0 = 0$ or $v_0 = 0$.

$$\text{(page 92)} \quad x_{\text{ave}} = \frac{\Delta x}{\Delta t} = \bar{x} \tag{3.7}$$

$$\text{(page 101)} \quad v = v_0 + at \tag{3.8}$$

$$\text{(page 104)} \quad v_{\text{ave}} = \frac{1}{2}(v + v_0) = \bar{v} \tag{3.9}$$

$$\text{(page 102)} \quad x = x_0 + v_0 t + \frac{1}{2}at^2 \tag{3.10}$$

$$v^2 = v_0^2 + 2ax \tag{3.11}$$

Obviously, if the motion starts from rest, then the initial velocity is zero ($v_0 = 0$) and we can often easily define our coordinate system so that the initial starting point is set to zero ($x_0 = 0$).

Let's work out what the acceleration would be for cars that you may or may not drive. How long does it take your car to go from 0 to 60 mph? If you own, say, a Mitsubishi Mirage ES, it will take you around 12 seconds to get to sixty. If you own a Porsche 911 Carrera S, it will take you closer to 4 seconds.

Example 3.1

The Slow Ride

Question : What's the acceleration of the Mitsubishi in m/s^2?

Solution:

From the above equations, we can see that if we start from rest, then $v_0 = 0$ and we can calculate:

$$v_M = a_M t_M$$
$$a_M = \frac{v_M}{t_M}$$

But we need metric units, so we can refer to Fig. 2.4 and see that 60 mph is just about 26 m/s. So for 12 seconds, we get

$$a_M = \frac{v_M}{t_M}$$
$$a_M = \frac{26}{12} \approx 2 \text{ m/s}^2$$

Not exactly like falling off a log...or falling off of anything for that matter, since it's almost 5 times less acceleration than 1 g! If you black out while speeding up with your Mitsubishi, it's not because of acceleration.

You Do It 3.3. Mitsubishi

or copy the solution

What is the acceleration of the Mitsubiishi and the Porsche respectively in m/s^2. And what fraction is that acceleration relative to that of gravity, $g = 9.8 \text{m/s}^2$?

These two states of motion—constant velocity and non-constant velocity—are very different. We'll see in our discussion of Relativity just how different, but let me let the cat out of the bag right here. Have you ever been in a train or a car where you're relatively enclosed and can only see a car or train next to you and not the earth? If the adjacent vehicle starts to move you can be suddenly confused: who moved? You might actually not know! I find that unsettling, and my brain quickly tries to figure it out. This is an inherent feature of constant velocity motion: there is no "right" answer as the same unsettled feeling would be felt by a passenger on the other train. Both are equivalent situations relative to one another. There. I've said the word "relative" and you now have the hint of only a part of what's interesting about Einstein's theory of *relativity*.

3.6.4 Going In the Right Direction!

Now we need to add one more piece to the modern story. As we saw in the Tools chapter, distances—displacements—are vector quantities. Because distance is in the definition of velocity, and since distance is a vector, so is velocity. Since velocity is a vector and acceleration is defined in terms of it? Yes, acceleration is a vector too.

Velocity Is a Vector

Here's our first physics-vector. I've been loose in the use of the words speed and velocity. In fact, there's a difference: velocity is a vector, $\mathbf{v}$, and so it includes a magnitude and a direction, and the magnitude of velocity is the speed. Now the definition should really be:

$$\mathbf{v} = \frac{\Delta \mathbf{x}}{\Delta t}$$

Notice, that I've implicitly included distance ($\mathbf{x}$) as a vector quantity, but not time. North and east are different directions and $\mathbf{x}$ makes that clear. Like mass, temperature, and many other quantities, time is not a vector. At least, not yet!

So 60 mph east is not the same velocity as 60 mph north. The speeds are the same, but the velocities are different. And, certainly if you're trying to go north, you don't want to deploy a *velocity* that points east. So both the magnitude (speed, here) and the direction are required to specify a velocity. We simply draw an arrow, the direction of which points in the direction of travel and the length of which is defined by some scale of speed magnitudes.

Now when you're walking I want for you to invent your own speed scale—how many inches equals 1 mph of speed— and then imagine this arrow sticking out of your chest. As you speed up, the arrow gets longer and as you slow down, the arrow shrinks. If you turn left, the arrow turns with you. Everywhere you go, you imagine your very own velocity arrow preceding you. You should also imagine arrows coming out of everyone you see, people on bicycles would have longer arrows and cars would have even longer ones. When they come to a stop sign? The arrow disappears. Keep this in mind as we move on to those length-changing arrows.

Wait. *That's silly. But now I'm seeing arrows everywhere.*

Glad you asked. :)

Acceleration Is A Vector

Since velocity is a vector and acceleration is defined in terms of velocity, it too is a vector:

$$\mathbf{a} = \frac{\Delta \mathbf{v}}{\Delta t}.$$

While the little arrow that represents your personal velocity is always point out in front of you, its direction seems rather arbitrary. It goes where you go, but there's nothing special about the direction. But that's not really the case since our movement is inside of a coordinate system of streets, sidewalks, hallways, etc. and they all have relative directions.

So let's for a minute remember the obvious...if we take our coordinate system to be (arbitrarily) that to go east is to go in a positive direction and west, a negative direction then you could describe your vector that way. If you're walking east, then you could say your velocity is 2 mph, east, or you could say that your velocity is +2 mph. If you're walking to the west, your velocity would be -2 mph. The direction and the sign are mingled. Because we will work in one dimension most of the time, this dual-role for a sign will matter but hopefully be pretty easy.[24]

[24] The alternative is to use a full-on vector notation which is more complication than we will need in QS&BB.

This is important as we consider the direction of an acceleration. Suppose your speed is 60 mph and after a certain time—let's say 2 hours—your speed has increased to 70 mph. Remembering our ordering in the Δ notation of (now - before), then then the magnitude of your acceleration would be:

$$a = \frac{\Delta v}{\Delta t} = \frac{70-60}{2} = 5 \text{ miles per hour}^2$$

Now what if your original speed is reduced because you applied the brakes from 70 mph to 50 mph. Then we'd have:

$$a = \frac{\Delta v}{\Delta t} = \frac{50-70}{2} = -10 \text{ miles per hour}^2$$

What are we to make of that negative sign? Of course it means we're slowing down, and we learned that the term for that is deceleration. But that pesky sign change between accelerate and decelerate is again related to our coordinate system. If we continue to take east as positive, and have both of our circumstances—the acceleration to 70 mph and the deceleration from 70 mph—be easterly motion, then that algebraic negative sign that came from the calculation plays a geometrical role.

Now you have two vectors point out of your car's hood. First we have the velocity vector, which is always positive as long as we're traveling east, albeit getting longer in the one case and shorter in the other. Then we have also an acceleration vector that points to the east when we accelerate (corresponding to the change of the speed being positive) but points west when we decelerate. That's what the negative sign does, it changes the direction of the acceleration vector. Let's see how this works.

What Goes Up, Must Come Down

Armed with Galileo's discovery and our notion of vectors, the acceleration felt by all falling objects near the Earth is constant, we can look at some examples. Let's drop something from the Leaning Tower, in honor of fictional accounts of famous scientists.

Mr. Google tells me that the height of the Leaning Tower is 183 ft. By now, you know how to convert that to meters and it's 57 m. The first thing to realize is that a baseball dropped from the Leaning Tower at this height would likely reach terminal velocity. But let's ignore that for a moment—we will pretend that the atmosphere doesn't exist—and look at Fig. 3.2 for the arrangement.

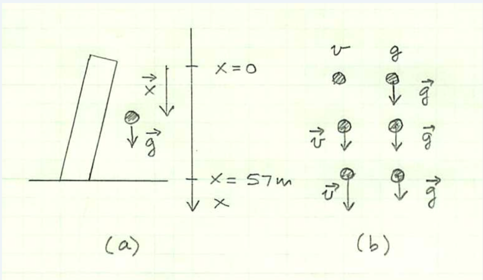
Wait. *Why do you keep doing that? You take a perfectly normal situation and then replace it with a pretend situation and explain that instead!*

Glad you asked. *This is what Galileo taught us: We try to understand the basic rules by which Nature operates and must work very hard to understand why our measurements might be corrupted by effects that mask the details. Sometimes what we can calculate are only the simple situations and we have to add in the complications, like friction or air resistance as small effects. Sometimes we can't calculate these effects perfectly so we have to build...yes, models...to approximate them. But the overall goal is those hidden, underlying regularities that dominate.*

Example 3.2

That Leaning Tower

Question:



Drop an apple from the Leaning Tower of Pisa, all 57 meters.

- Draw a picture that shows the rock on its way down twice...near the top and near the ground: draw a representation of the distance vector, the velocity vector, and the acceleration vector on the rock.
- How fast is a rock moving just before it hits the ground?
- How long does it take to hit the ground?

Solution:

Always draw a picture. We'll take the positive x axis to be pointing down, in the direction of $\vec{g}$ and the direction of v, so both are positive and increasing in magnitude. Figure 3.2 shows the situation and the various vectors. The distance axis is on the left side. On the right side of the figure are the kinematical vectors. The velocity vector gets longer as the apple falls and the acceleration vector stays the same. Notice that there is an acceleration at the top.

We can use various of the equations from the set in Eqs. 3.7. To find the time, for example, we can use $x = x_0 + v_0 t + \frac{1}{2}at^2$. We defined our coordinate system so that $x_0 = 0$ at the top of the tower. And since it's dropped from rest without being thrown down, $v_0 = 0$. And we can then calculate the time when $x = 57$ m:

$$x = \frac{1}{2} g t^2$$

$$2x/g = t^2 \text{ so } t = \sqrt{2x/g} = \sqrt{\frac{(2)(57)}{9.8}} = 3.4 \text{ s} \tag{3.12}$$

In order to calculate the speed at the bottom, we could use a couple of equations. We could use:

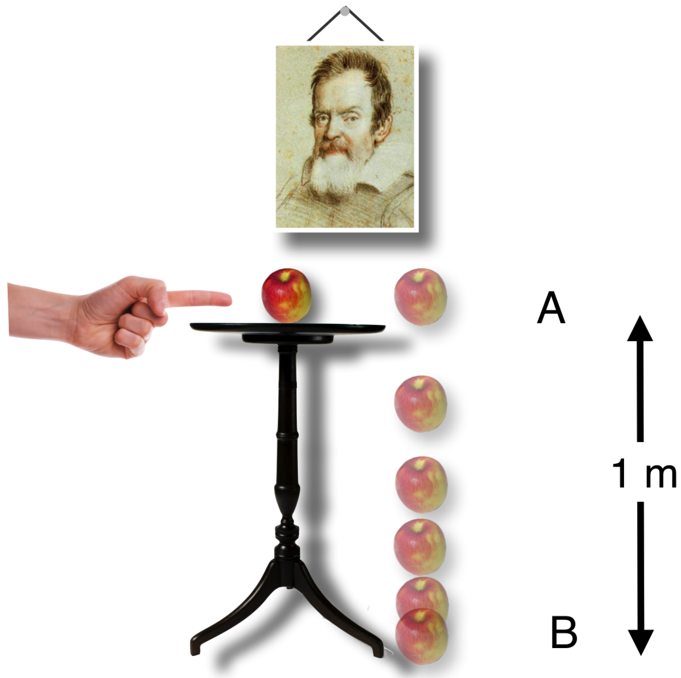
$$v^2 = v_0^2 + 2ax = 0 + 2(9.8)(57) = 1117$$
$$v = \sqrt{1117} = 33.4 \text{ m/s}$$

Now, you do it. Closer to home.

You Do It 3.4. Apple Motion

or copy the solution

Instead of dropping something from the Leaning Tower, drop an apple from a table to the floor, 1 meter down. What's its speed just before it hits? Calculate it in m/s and pretend that $g = 10$ m/s^2...it's close!



In fact, while the above illustrations are elementary, they lead us to practical and interesting questions. For example, suppose I throw a ball straight up. We could analyze the velocity vector on the way up and the way down and the acceleration vector during the same trip. Galileo's assertion that the acceleration is constant for all objects, we interpret as the acceleration vector of constant length pointing *down* throughout the path, up and down. What about at the very top? Is it accelerating? Yes. Is it moving? No.

One of the biggest indignities that Aristotle's motion suffered was trying to explain projectile motion. Galileo has that covered too.

3.7 Projectiles

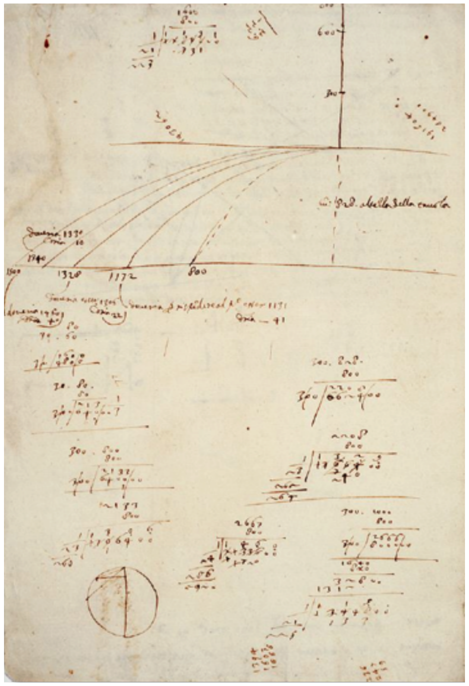


Figure 3.14: Galileo's notes showing his table-floor measurements.

Although we've been dealing with 1-dimensional motion, Galileo wasn't done when he figured out that everything falls at the same, constant acceleration of gravity. He also tackled the other problem that Aristotle messed up: throwing something. With an understanding that objects would happily move at a constant speed without being pushed (his two inclined planes) and that objects fall towards Earth with a common acceleration, he had the genius idea to embed these two different motions together into the motion of a single, thrown object. Let's follow his experiment by looking at his notes in Fig. 3.14.

What you see are various measurements of rolling a ball off the surface of a table and marking where it lands according to how fast it was going. How did he repeat measurements so that each attempt would be the same speed upon lift-off? He rolled them from an inclined plane where he could dial-up the speed he wanted by how far up the plane he let go of the ball as shown in Fig.~3.15. Clever, no?

So let's think like he did. If the speed of the ball at B is v_x, then what's the speed of the ball just at the edge of the table, C? v_x. Now the ball acquires two separate motions that are combined: the ball *continues* to have the horizontal speed, v_x. But as it falls, it starts to acquire a *new vertical speed*, accelerated down by g. The two motions are separate, but together! So throughout the trajectory, the ball's *horizontal speed remains* v_x. Nothing has happened to change that. The ball's *vertical speed increases* as the ball "falls" along this curved path.

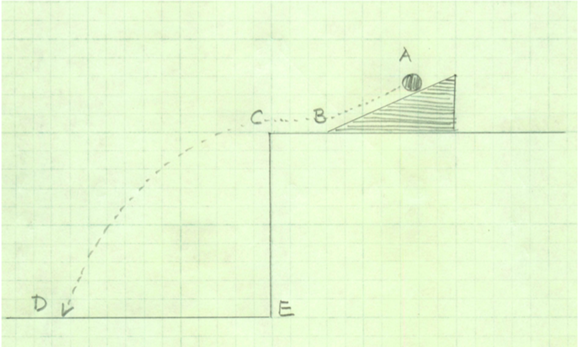


Figure 3.15: This is the arrangement for Galileo's projectile experiment. He released a ball at point A on an inclined plane so for each trial, every ball at B would have the same speed, which would be the speed upon "launch" off the edge of the table at C. The ball then flies through the air and lands at point D.

Now here's the neat thing. Suppose at the same point where the ball leaves the table, another ball is allowed to just drop from that point, say at point C in Fig. 3.15 so that it falls (under the influence of gravity!) to point E. This one has only one kind of motion, vertical and accelerated. In fact, it's accelerated by the same amount as the first one and from the same vertical height, at the same time. So which one reaches the floor first?[25] They both reach the floor at the same time.

[25] This is a "Who's buried in Grant's Tomb question.

This bundling up of two separate motions into one object was pure genius. Nobody had ever conceived of that sort of thing. By analyzing the landing point and some sophisticated solid geometry, he was able to extrapolate to perfect conditions and assert that the trajectory that the ball followed was that of a parabola.

Now let's take this to the act of actually throwing something. Refer to Fig.~3.16.

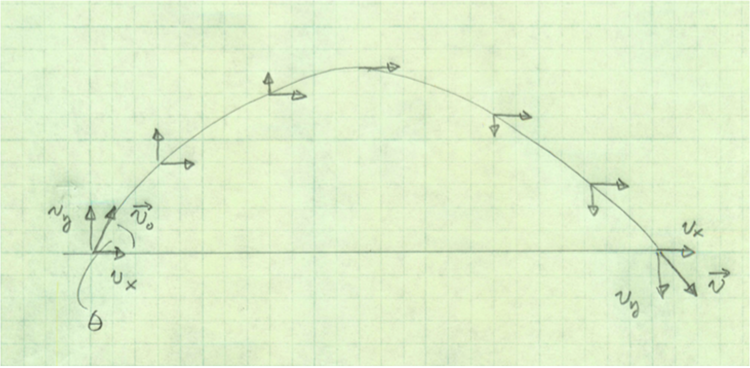


Figure 3.16: This shows the path of a projectile—like throwing a ball—launched at the left and landing on the right. The velocity is shown twice at the beginning and at the end. $\mathbf{v_0}$ is the total velocity when the ball is first thrown, a little out and a little more up. This velocity is also drawn on top in terms of its x and y components. This is drawn again for the landing point. At each point along the way, the components alone are shown.

Here I've put on the diagram the separate components of the ball's velocity, horizontal and vertical. Throughout the path the horizontal velocity that the thrower originally provided (because she didn't throw the ball straight up, but at an angle relative to the horizontal) is unchanged.

But the vertical component of the velocity points up until the top. Then, down on the way, well, down. If you've ever watched a towering home run, you know that the baseball does not follow a perfect parabola. Rather it falls to Earth relatively more quickly than it took off and so the trajectory is shorter down than up. The forces that an actual ball feels are two: there's the force of the Earth's gravity, always down. But there's also the force of resistance that the air presents to the ball and its direction is precisely the opposite of the ball's actual direction. The faster the ball goes, in most instances in a fluid, the harder that force of resistance becomes. So as the ball gains speed from gravity, it loses some speed due to deceleration from resistance force. . . which increases as the ball falls and hence changes the trajectory from parabolic.

Projectiles on the Earth follow essentially parabolic trajectories. *Key Observation 3*

In Chapter **??** we'll begin to develop a graphical way to approach things that move. We're going to learn to draw Feynman Diagrams in Spacetime. Not your father's physics book, this one.

3.8 The Beginning of Physics

When Galileo reached the conclusion that all objects fall to Earth with the same acceleration, he was going against what he—and Aristotle, and everyone else—actually observed. An iron cannonball *would* be observed to reach the ground faster than a wooden one and as you saw in your crumpled paper experiment from page 112, the same object would be observed to fall at different rates depending on its shape. So how in the world could he insist that there is a uniform rule, in spite of these undeniably different outcomes?

Today you'd say "wind resistance" as an explanation of any differences observed of falling objects and you would be right. But for a 17th century thinker, this was a stretch. Galileo said that it's not sufficient to just observe and describe what happens because effects like friction are hiding what Nature *really wants* to do! And uncovering those hidden rules is his science...and ours.

3.8.1 The Red Pill, or the Blue Pill?

This is very un-Aristotle. First of all, Aristotle would never have countenanced doing experiments. You can look at nature, but don't touch. But Galileo was all about constructing experiments and making quantitative measurements of what he observed. This was new and recognizably "modern."

But there's more. To him Nature's rules are hidden to us. We can get close to them by reducing marginal effects like friction but then we have to extrapolate from what we see in our rough and ready laboratory to the hidden rules of a more perfect world. The pendulum bob gets close when it swings back, but the actual rule driving the motion would instruct the pendulum bob to come all the way back to the original height. That's what's real.[26]

[26] It's not for us to get too deeply involved into the question of "what's real." Is what happens to us what's real? Or is the underlying, mathematical regularity what's actually Real. Exactly the premise of *The Matrix* and other science fiction stories.

I cannot over-emphasize how important this is. This is very much Plato's view of nature, not Aristotle's. For Plato, The Real was perfect and with our poor, corrupt visual tools we can only perceive inferior copies of the real things—Platonic Ideals. While there is much that's wrong with Plato's philosophy, the uncovering of nature's hidden order—free of imperfection—is the goal of modern science.[27]

[27] Ask Mr Google about Plato's Cave.

The Father of Physics

Nobody had ever done what Galileo did in all of history. Let's summarize his strategies:

First, Galileo chose not to explain nature (motion in his case) on the basis of logical argument from within pre-conditioned philosophy (neither Aristotle's nor the Church's): Galileo confronted nature without precondition: he assumed that Authority did not dictate how Nature is.

Second, rather than sit passively and observe, Galileo created artificial circumstances designed to explore particular questions: he assumed that Nature can be characterized by doing experiments.

Third, rather than report results as a narrative, Galileo made quantitative measurements under the assumption that arithmetical and geometrical constructions were descriptive of nature's behavior: he assumed that Nature is mathematical.

Nature appears to behave according to mathematics. *Key Concept 6*

Why is Nature inherently mathematical? *Key Question 7*

Those three strategies alone would make Galileo the first, great scientist. But he went further, and his Physics Paternity comes from his Platonism.

Fourth, Galileo chose to interpret nature as consisting of rules that can only be discovered by going beyond the rough regularities in our observations. Nature is best assumed to be simple and mathematical, but unfortunately all that we can observe is complicated by extraneous effects and the true nature...um, of nature is hidden, just out of grasp. Strip away complications like friction as best you can and through mathematical modeling and extrapolation we can uncover Nature's hidden rules.[28]

[28] By now we can go further and add into our models effects of friction and air resistance, for example. This is one of the important features of models as I described in Chapter 2. But it's not perfect, just better.

Without this **fourth** strategy, physics would be impossible.

The rules of Nature are often hidden from experiment and must be inferred by a combination of theoretical modeling and experimental confirmation. *Key Concept 7*

Taken together, these four strategies form the modern-looking nature of physics.

Chapter 4

Diagrammatica: Space and Time Diagrams

Motion

Diagrammatica chapters will be special. Many of the explanations in QS&BBare in a diagrammatic form and will require some concentration. In order to give them the detail that they deserve and yet not clog up a regular chapter, I'll punctuate the narrative with these special interludes, all about diagrams that partner with the previous chapter. I'll be as straightforward as possible. Tedious, even. To get you into a Diagrammatic mood.

In Chapter 3 we dealt with the motion of a single thing: our car headed for northern Michigan, the bicyclist or runner, something falling from the Leaning Tower. While these are all fine examples of motion, we're going to usually be concerned with things that happen between two or more moving objects. Often, their collisions will be our focus.

A Space or Spacetime Diagram for two separate trajectories and can become a diagram for two interacting things by just overlaying them so that their coordinates match up in space and in time. So if one of our runners were to bump into another, we'd have one of those dramatic TV moments and we could drawn their Space and Spacetime Diagrams on one piece of paper. Since no two runners can be in the same place at the same time, where they bump you'd expect that their trajectories in would each show an abrupt change of direction. That's what we'll be about—characterizing the space, time, and momentum features of things that go bump in the night. Or day. Or during the Big Bang.

This won't hurt if you take it line by line. In these Diagrammatica chapters, you'll need your pencil. I'll wait...

4.1 Space and Spacetime Diagrams

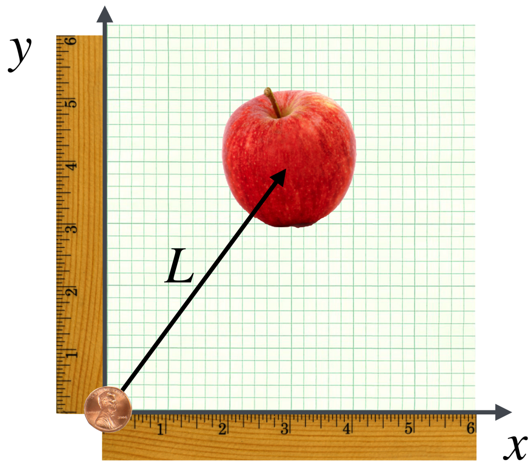


Figure 4.1: An apple on my desk at $(x, y) = (3, 4)$ inches. The penny is at the origin.

Let's go back to the drawing board. In coming chapters we will talk about FeynmanDiagrams as they pertain to billiard ball collisions and elementary particle collisions, but I can set the stage for some of the concepts here. We'll constantly deal with two kinds of diagrams which I'll call "Space Diagrams"and "Spacetime Diagrams." The significance of the name "Spacetime" will be apparent later, butfor now we'll look at it in just a nuts and bolts manner.

As a physicist, I'm obligated by law to think mathematically, so my desk has a grid on it with rulers specifying where everything is (doesn't yours?). Figure 4.1 shows this coordinate system where you can see that I have an apple on my desk at $(x, y) = (3, 4)$: a healthy Space Diagram. Further, where I oriented my rulers is completely arbitrary and I chose to put my space origin at the location of a penny that was there.

4.1.1 Space Diagrams

The idea of a Space Diagram is simple and familiar since we've already looked at a map in Fig. 3.2. That's a drawing in Space that you make all the time. For a map the "space" is the surface of the Earth over some patch on which you're traveling as shown on which you've implicitly embedded time in your map. Each blue speck of ink (a pixel!) on the map represents your position *as time increases.* The third space dimension (up and down) isn't interesting for highways. (And besides that, Michigan is flat!) But in principle our 2-D map could be supplemented with "up" if we took a helicopter. Then we'd need that third dimension.

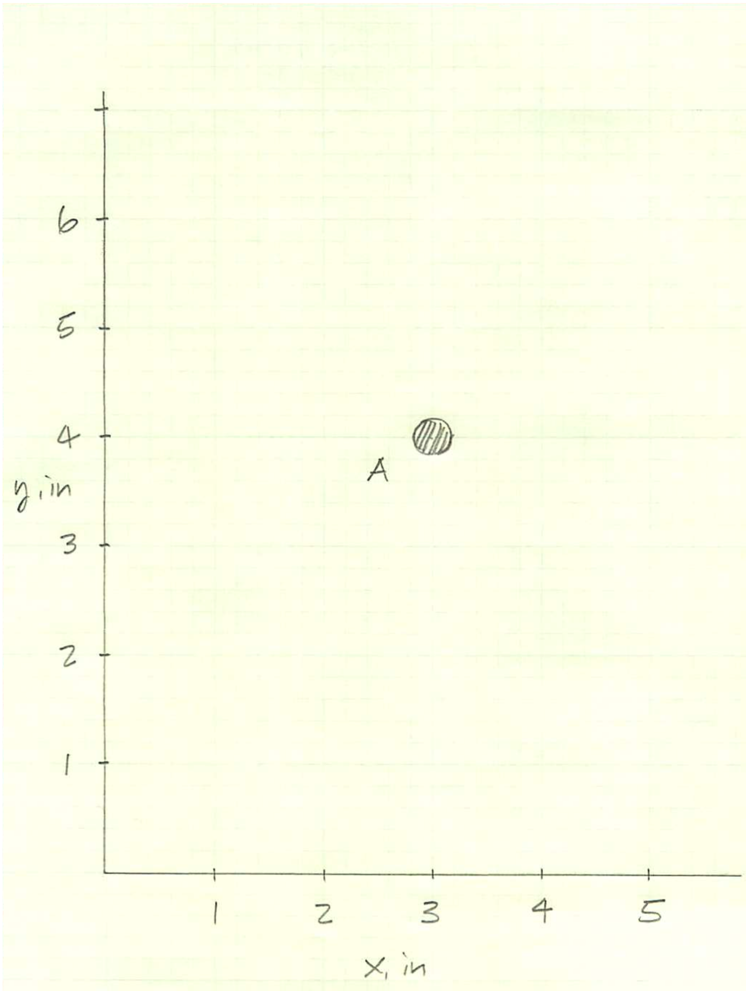


Figure 4.2: The space diagram of the apple, where now we imagine it to have no size. Just a point at A.

You Do It 4.1. Apple Distance

or copy the solution

Calculate the distance from the origin to the apple at point A. Remember the Greek guy's trigangle formula? Then draw a vector from the penny to the apple on Fig 4.2 and convince yourself that its about the right length by looking at the grid in Fig. 4.1

Definition: **Spacetime.**
The four-dimensional grid that incorporates both space and time in a single "fabric."

4.1.2 Spacetime Diagrams

Now the other kind of diagram: In the particle world we're obligated to operate inside a universe of space *and* time. As we'll see when we talk about Relativity, the distinction between them is actually not as much as you'd think since they are on equal footing as coordinates in a fabric we affectionately call: *Spacetime*. So we need to follow trajectories through the three dimensions of space and the additional dimension of time, or in 4 dimensions.

In a Spacetime Diagram we explicitly call out Time as one of the axes. Now I've looked everywhere, but paper is only sold in two dimensions so we have to improvise since we now use up one axis for Time, have only one left for Space. You'll see, it works.

In a **Space Diagram**, the forward progress of time is implied—it's a parameter that's not explicitly drawn. The dimensions of the paper are basically the traditional x and y (or East and North).

In a **Spactime Diagram**, the forward (and backward) progress in time is indicated explicitly on one of the axes. Traditional space is represented in the other axis.

A Spacetime Diagram shows a space trajectory on the vertical axis and the time trajectory on the horizontal axis. *Key Concept 8*

Let's look at my desktop. The y value of the apple's space-coordinate is "4" and the x coordinate is "3." It's just sitting there at that point in space and so its Space Diagram position is just a point at $(x, y) = (3,4)$ as shown in Fig. 4.2. Don't touch it. Notice that a little while later, its spatial coordinates are still (3,4). Is it sitting still? In Space, yes, but what's happened in *time*? It progressed into the future, so it moved in Spacetime and we can draw that trajectory.

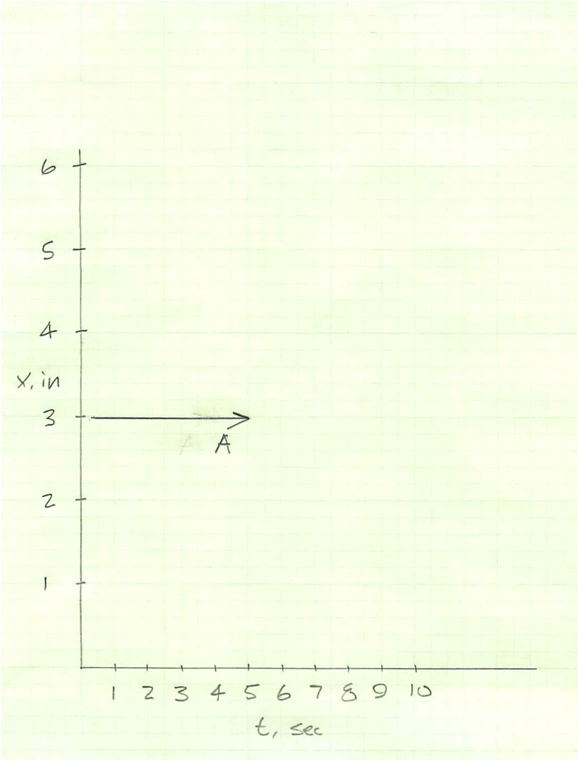


Figure 4.3: The spacetime diagram of the apple sitting still for 5 seconds.

Pencil 4.1. ✎

To go from a Space picture to a Spacetime picture, we have to jettison one of the space axes (remember, 2D paper). Here I'll choose the one we keep to be x (it could have been y). We labeled the *where* of the apple to be point A in a coordinate system with its origin located on the penny. As it sits there, it continues to be located at point A in space.

Moving in Time, But Still

The origin of our time axis can be chosen to fit the situation: look at your watch...we'll say that $t = 0$ seconds...*now*. So in order to analyze the history of a trajectory through space and time, then we have just conveniently defined $x = 0$ and $t = 0$ origins.

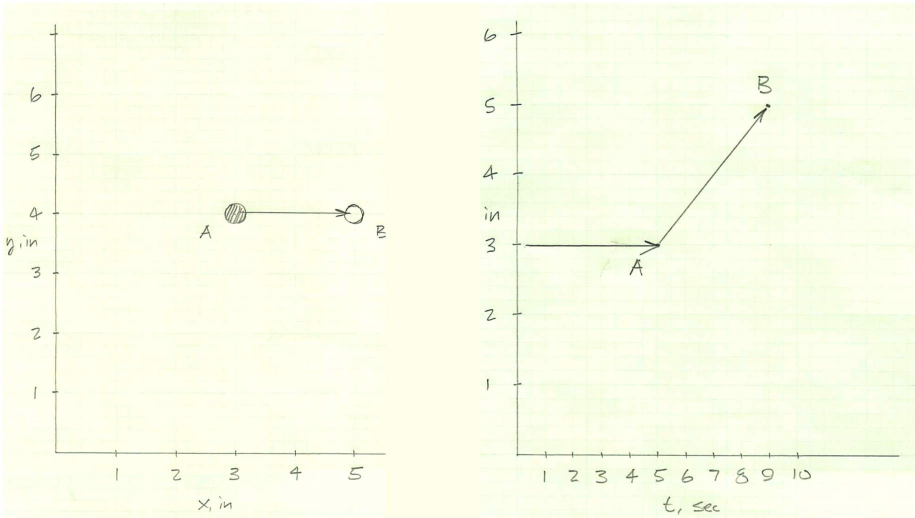


Figure 4.4: In the left-hand diagram (a), the apple has moved from point A to point Z in space. In the right-hand diagram (b), the Spacetime Diagram shows that the apple suddenly (when it was flicked) moved in space (the x dimension change) and in time.

Back to our apple with pencil in hand. When we last left it, its *where* coordinate was at A, $x = 3$, and it moves along its *when* coordinate into the future.[1] We start our clock and wait a bit and so we've progressed in the direction of the "time axis" by a few seconds. Oops, the apple is still sitting at A but another few seconds has gone by and so we've "moved" a few more seconds along the time axis. Let's say that 5 seconds have elapsed since we first started taking notice...that we've progressed from $t = 0$ to $= 5$ s. So we can draw the trajectory of the apple in Spacetime for this trip.

Like before I'll draw space coordinates up and time coordinates horizontally.[2] The Spacetime trajectory of the apple is then a line always at the space coordinate $x = 3$ which extends horizontally—in time—toward $t = 5$. Figure 4.3 is the Feynman Diagram for this silly, but subtle situation. That's your first Feynman Diagram, the one of an apple patiently, sitting on my desk. Let's kick it.

[1] We can move backwards in space. Can we move backwards in time? Stay tuned!

[2] This is the opposite of what's done in professional physics, but I want to keep the notion of a slope in this space as obviously speed.

Moving in Time, But...um...Moving

Sitting still in space isn't so interesting. Let's make our apple roll: at that 5th second flick it with your finger so that it rolls straight along the $+x$ axis for 4 more seconds during which time it moves from $x = 3$ to $x = 5$. And, for the purposes of simplicity, let's assume that it rolls with a constant speed. The Space diagram is pretty simple, right? It's just a horizontal line from the original (x, y) position to its new x value. Figure 4.4a shows just that. This is just a regular map showing a trip from point A to point Z.

The Spacetime Diagram is also simple, but we have to think about it.

Wait. *"Simple," but we have to think?*

Glad you asked. *Jeez. I'm not glad you asked that one! Do it with your hands and you'll see that it's simple, but that, yes, you had to think.*

The constant speed tells us that the rate at which x increases is the same for each increment of time, so our diagram is the straight, but sloped line shown in Fig 4.4b. We can read the speed of the apple directly from the slope which is

$$v = \frac{\text{distance in space traveled, 2 units}}{\text{time it took, 4 units}} = 1/2$$

(in arbitrary units). Of course, if the apple were accelerating, the trajectory in the Spacetime Diagram would be curved.

A Collision Between Two Apples

Here, we flicked the apple, from off-stage so to speak. One actor. But now we'll do the same thing, but actually keep track of two objects, the hitter and the, um, hitee. Let's draw the Space Diagram and the Feynman Diagram for this *collision*. We'll knock it with another, identical apple.

Your pencil is sill out, right?

Remember that the original apple ("T," for Taylor Apple[3]) was happily sitting still at position $(x, y) = (3, 4)$ inches. We'll roll a second apple at it ("B," for Bismark Apple[4]) moving from left to right at the same y position, $y = 4$. Since they both are at the same vertical position, a collision is immanent! The situation is shown in the Space Diagrams in Fig. 4.5, the left figure is for the T apple and the right is for the B apple. Obviously if the objects are identical, then B stops dead and the A shoots out with the same motion as the B had originally. (You've done this with pool balls, right?)

Now lets consider the Feynman Diagram for the same situation. We have to pick a space coordinate to plot on the vertical axis and I'll pick the x direction since all of the motion happens in x. T is sitting still until the collision happens and so we'll draw that in on Fig. 4.6 as the same sort of "sitting still" horizontal line at $x = 3$ as we did in Fig. 4.4. We'll say that the collision happens at $t = 5$ s at which time B hits and stops dead and *its* motion becomes stationary.

B gains distance as time increases, so on our Feynman Diagram, they are slanted lines. Figure 4.6 shows B advancing in the dashed trajectory by gaining distance from $x \sim 1$ to the collision point at $x = 3$ at the collision time of 5 seconds. The target's speed after the collision is the same as the beam's, so it shoots out

[3] http://www.orangepippin.com/apples/taylor

[4] http://www.orangepippin.com/apples/bismarck

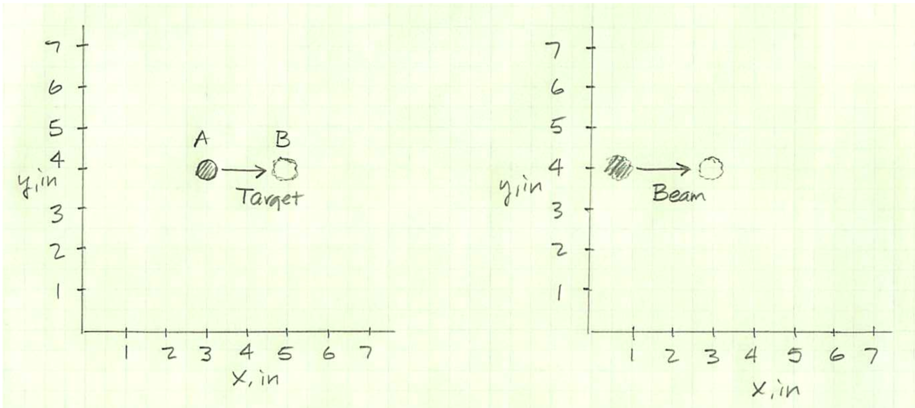


Figure 4.5: Space Diagrams for the before and after picture of the simple collision of the red projectile (beam) puck and the stationary (target) blue puck. The dimensions are roughly those of an air hockey table and the collision all occurs in the x dimension.

after being hit, so it has the same slope in the Spacetime Diagram. There we have it. Our first Feynman Diagram for a two-body, elastic collision.

Picturing trajectories like this will be useful in an operational sense, but we'll see that Feynman Diagrams are also going to be a fundamental tool in EPP.

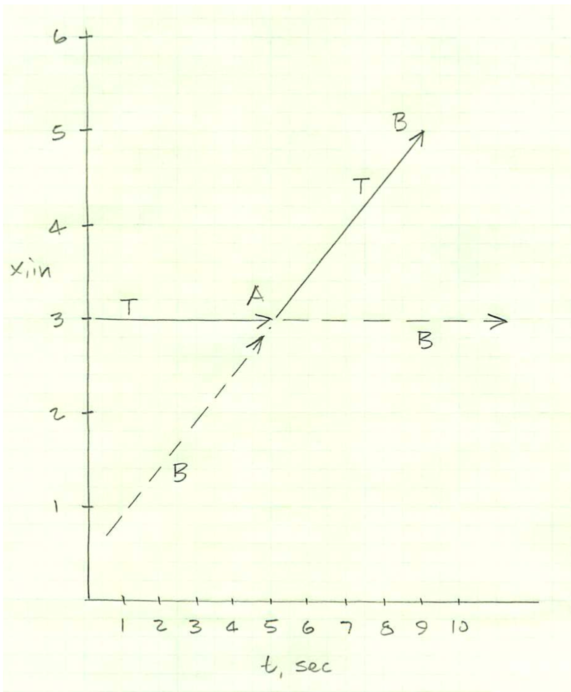
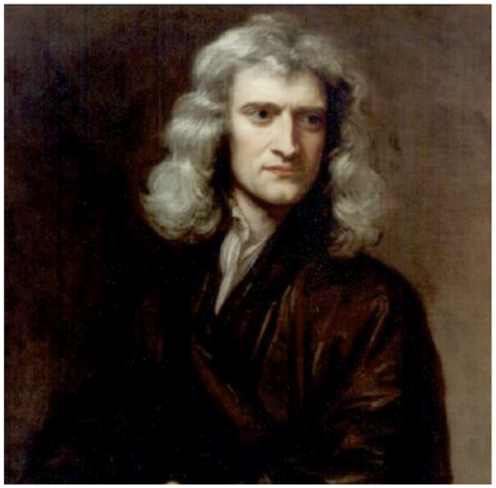


Figure 4.6: The Feynman Diagram for the same collision. Notice that the stationary target eraser is moving in time (horizontal axis) and that the slope of the beam eraser indicates that it is moving in the positive x and positive t directions. The $t = 0$ origin is arbitrary and the collision happens at $t = t_C$.

Chapter 5

Momentum and Force

The Big Mo.



Isaac Newton when at his political prime at his scientific prime. Sir Godfrey Kneller, 1689

Isaac Newton, 1642-1727

““I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.” ”*Correspondence.*

Imagine a three year old boy, fatherless and abandoned by his mother to a grandmother he feared and an unwelcoming grandfather. Imagine that this boy is solitary by nature and surrounded by half-sibling girls until late in his teenage years. Now try to project that child into a well-rounded and comfortable future. Hard to do, right? That's the beginning of the most extraordinary mind ever.

5.1 Goals of this chapter:

- Understand:
  - How to calculate a force required to produce a particular acceleration in one dimension.
  - How to calculate the average change in momentum induced by the application of a force through a given time.
  - How to calculate the centripetal force and acceleration for an object moving at a constant, circular speed.
- Appreciate:
  - Forces create accelerations.
  - That if a momentum vector is changing in direction or magnitude, a force had to be involved.
  - That an object moving in a curved path must have had a force applied to it perpendicular to the trajectory.
- Be familiar with:
  - Newton's third law
  - Newton's Life

5.2 Introduction

We're all about modern particle physics and the collisions we create in order to see into the deepest depths of reality. But in order to get there we need forces to accelerate our beams to high momenta and a language to describe their energies. In addition, we learned long ago that the most intriguing feature of the particles we produce is: mass. All of these concepts have their roots in Newton's notebooks: Force, acceleration, momentum, and mass.

But even though our subject is very contemporary, we're stuck with the definitions and terminology from the 1600's when our heroes emerged: Galileo, Kepler, Descartes, Huygens, Leibnitz, and Newton. While some of those fellows are important for their brilliant intuition (Galileo, Kepler, Descartes), only Isaac Newton, Gottfried Leibnitz and Christiaan Huygens wrote down mathematical relations which form the language and, yes the Metaphors of our modern models of how things move. Typically, we divide mechanics into two parts:

- *Kinematics* is the description of the motion of objects without regard to what caused them to move. If a ball is accelerating by some amount, the rules of kinematics will tell you how far it will go in a

The study of motion without regard to cause.

given time, how fast it's going after a given distance: time, distance, speed, and acceleration are the parameters of Kinematics. When you estimate how long it will take you to arrive at a destination, you're doing kinematics. We tend to attribute the important ideas of Kinematics to Galileo, but he didn't have algebra and the actual mathematics of the kinematical rules came later with Newton and others. What we did in Chapter 3 is sufficient for our purposes.

- *Dynamics* is the study of forces—their causes and their consequences. According to Newton, forces create accelerations in objects by pushing or pulling. That's how the ball in the previous paragraph got its acceleration—something pushed on it. This is a big subject, but we will only consider motions in one dimension, except for circular motion. Here's the rule of thumb for dynamics:

Definition: Dynamics.
The study of forces which cause accelerations.

A force applied to a body will cause it to accelerate. *Key Concept 9*

It's often said that Newton was born the same year that Galileo died, but that's not quite correct. Britiain waited until 1752 to convert to the Gregorian Calendar which required an 11 day adjustment. Since Newton was born on Christmas day in Britain, these famous events are not quite the same year.

5.2.1 A Little Bit of Newton

The biggest scientific life of all, is Isaac Newton's. His childhood was a mixture of pain and some accidental fortunate associations. His father was a farmer in Woolsthorpe,[1] not far from Nottingham and a little more than 100 miles north of London. Isaac senior died before tiny, premature Isaac was born. When he was three years old, his mother, Hannah—a semi-literate woman for whom the farm and manor was a big job—married the 63 year old Rector of North Witham who wanted nothing to do with a frail toddler who was then left in the care of his maternal grandparents and female cousins.[2] Seven years later, Hannah, a widow yet again, returned with two young children in tow. While she had been away, young Isaac had been sent to school, a privilege that might not have happened, had his father been alive.

When he was 12 years old, he was sent to a free grammar school in Grantham where the emphasis was Latin ("grammar," after all), which was the language of intellectuals and in which he wrote his great works. He lived with the apothecary, another lucky break as that family indulged his precocious abilities with tools and crafts. In what must have been one of the most frivolous activities of not just his adolescent, but entire odd life, one night he constructed dozens of kites with firecrackers, which he flew over the town at night "...wonderfully affrighting all of the neighboring inhabitants for some time, and causing not a little discourse on market days..."

He was a loner then, and for most of his adult life. "His school fellows generally were not very affectionate toward him. He was commonly too cunning for them in everything. He who has most understanding is least regarded." When he was 17 Hannah brought him home and tried to turn him into a farmer, but

[1] Of course, an historic treasure in Britain today: http://www.nationaltrust.org.uk/woolsthorpe-manor/

[2] When Isaac was 19 years old, he wrote a list of his sins—he always kept voluminous private notebooks—among which he listed, "Threatening my father and mother Smith to burn them and the house over them."

Figure 5.1: Isaac Newton's childhood home where he first conceived of gravity, optics, and calculus. (Copyright GP Williams and licensed for reuse under this Creative Commons Licence)

he was a disaster. She gave up and sent him back to Grantham to prepare him for university...as the only outlet for his increasingly apparent unusual mind.

He was sent to Trinity College at Cambridge University where he was enrolled as a "Sizar" which was essentially the role of servant to an upperclassman. He was 18 years old, considerably older than most of the students and as a studious person, different from the mostly carefree student body. He made a single friend, a most unlikely event for the difficult Newton. He bumped into John Wickins while he was alone on a bench and they became roommates for 20 years, until Wickins married and left. Wickins was Newton's assistant as Newton began his life of changing the world.

Figure 5.2: Newton's faculty rooms at Trinity College, Cambridge. (http://heirloomheritagetours.com/blog-2/isaac-newtons-cambridge-trinity-college/)

The curriculum at Cambridge was terribly old-fashioned: Aristotle, from top to bottom. But Rene Descartes from the Continent was all the rage and read and discussed secretly by the students. To Descartes, the world was mathematical in a manner beyond what Galileo might have imagined. It was he who blended geometry together with the brand new algebra so that equations might be presented as curves and curves, as equations. Of course we call our axes, "Cartesian" after their inventor. Descartes was also a mechanist: meaning the motions of all things were caused by mechanical interactions of matter. Nothing spiritual, nothing occult. Motion was inserted into the cosmos by God originally, who then apparently abandoned His creation in order to pursue other interests? In any case, Descartes believed that this original motion persisted[3] as the universe formed and that God did not drop in and adjust things. The world was predictable according to Laws (yes, capital L for Descartes) and he proposed to start with the Laws and draw observable conclusions from them. Very top-down, was Descartes. He was all about Why. Newton weaned us from Why to How. Even the planets moved by being carried by vortices of invisible balls. Mechanical modeling was his goal, and analytical mathematics was his tool.

The young, inquisitive Newton ate Descartes up. Here was an escape from Aristotle, but a systematic and mathematical way out, and that was right up Newton's alley. But Descartes' conclusions were problematic for Newton and much of what he wrote later was in reaction to his disagreement with Descartes.

But the Frenchman's[4] mathematics stuck and the neoPlatonists at Cambridge carried Descartes' mathematics program forward and the leader of that movement was Isaac Barrow, who was the first Lucasian Professor of Mathematics. Isaac Newton was the second.[5] He became a student of Barrow's and eventually, his benefactor. But disaster struck in London—in 1664 the Bubonic Plague arrived and before it was over, 20% of the population was dead. By 1665 it hit Cambridge forcing the university to close and the 23 year old Newton went back to Woolsthorpe where he remained to himself for more than a year. There, he consumed all mathematics known at the time and went beyond. While at the farm, he basically invented calculus, had his first ideas about gravity (the famous apple was to have fallen in his presence during this period), and reinvented optics.

[3] This is very close to one of the so-called conservation laws that are so important. We'll see that next chapter.

[4] French, indeed. But Descartes saw what happened to Galileo and relocated to the Netherlands to avoid possible prosecution by the French Catholic Church. While we think of Descartes as the father of analytical mathematics, academia thinks of him as the Father of Western Philosophy. He of "I think, therefore I am" fame.

[5] Other famous Lucasian Professors: George Biddell Airy, Charles Babbage, Paul Dirac (of whom we will fawn over later), and Stephen Hawking

> In the beginning of the year 1665 I found the Method of approximating series & the Rule for reducing any dignity of any Binomial into such a series. The same year in May I found the method of Tangents..., & in November had the direct method of fluxions & the next year in January had the Theory of Colors & in May following I had entrance into the inverse method of fluxions. And the same year I began to think of gravity extending to the orb of the Moon & (having found out how to estimate the force with which [a] globe revolving with in a sphere presses the surface of the sphere) from Kepler's rule...I deducted that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: & thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth, & found them answer pretty nearly. All this was in the two plague years of 1665 & 1666. For in those days, I was in the prime of my age for invention & minded Mathematicks & Philosophy more than at any time since.

When he returned to Cambridge he rose quickly, his mathematical skills having surpassed those of all around him. In 1667 he was elected as a Fellow (like an assistant professor) with a salary. In 1669, Barrow resigned the Lucasian Chair insisting on Newton appointment to the highest post in the college. ("Mr Newton, a fellow of our College, and very young ... but of an extraordinary genius and proficiency in these things.") One of the only duties of the Lucasian Chair was to offer a single course a year, which he dutifully fulfilled, but usually lecturing to a literally empty classroom as nobody could understand him.

Among his discoveries during the plague years was that sunlight was composed of all colors, which was in conflict with the standard idea that white light was a color of its own and that the colors we see are mixtures of white and dark. He passed light through a prism and found that it spread out into a continuous spectrum of colors. This led him to experimenting with light passing through glass and eventually to grinding his own lenses for telescopes. By this time telescopes were beginning to be unwieldy in their length and the effects of light's spreading of color in the lenses ("chromatic aberration") was limiting precise viewing. He eventually changed the design completely. By using mirrors rather than lenses, he could create images of higher quality and with higher magnification in a compact form. His original 6 inch

"Newtonian reflecting telescope" would magnify by 40, which would have required a less precise conventional telescope six feet long to achieve.

The telescope and his explanation was his first entry into membership in the Royal Academy of Sciences. His explanation of how it worked led to the explication of his theory of light, which the Curator of Experiments, Robert Hooke, thought was stolen from his (incorrect) ideas. They became bitter enemies for life, one of a number of such vicious rivalries that Newton suffered through his whole life. So galling was this dispute—and the criticism that his theory of colors attracted from all over Europe—that Newton went into nearly complete isolation, vowing to keep the products of his research secret, rather than ever again suffer such public antagonism. He communicated almost exclusive through voluminous correspondence, much of which still exists. Newton did not suffer fools well, and even legitimate dispute would send him into a towering rage, or stony silence. His response to Hooke was to write a book on optics...which he inadvertently destroyed when a fire from an alchemy experiment that went out of control.[6]

What we all know as Newton's enduring scientific work came as the result of a wager and we'll pick up that story when we study his law of Gravitation and the first-ever attempt at a scientific cosmology in Chapter **??**.

Figure 5.3: Newton's second model telescope. From the Royal Society.

[6] Newton was an accomplished alchemist, probably damaging himself with the noxious chemicals that he inhaled and tasted. He believed that God had hidden first-knowledge of the physical universe to the Ancients, and that among those ideas were alchemical. His lodging was near to his laboratory on the Trinity campus and a huge furnace fire was never put out as he single-mindedly attacked chemistry. He also was an unusual religious fanatic. He learned Hebrew and translated the Bible from original texts and became convinced that fourth-century changes made the false claim of the divinity of Jesus. Rather, he should have been treated as a prophet on par with others. He was not a Christian, but a believer in a God of Nature. He wrote way more about alchemy and his fanatic religious histories than about science. These works are still being studied today.

Box 5.1 Newton and the Book

Newton was in his 40's when he basically sequestered himself in his rooms working on his alchemy and intensely weird religious researches. In 1684 three of his colleagues (the famous London architect Christopher Wren, the scientist Edmund Halley[7], and his arch nemesis Robert Hooke) were trying hard to figure out the shape of an orbit if the force of gravity varied like the inverse square of the distance from the Sun. Hooke claimed in his obnoxious way that he knew the answer, but he would not produce a calculation—because he couldn't since he had no mathematical training. Hooke's instincts often guided his brilliance as an experimenter. But, a little of Robert Hooke went a long way and Wren and Halley got tired of listening to him so they deputized Halley to go ask Newton. So he did. He showed up unannounced at Newton's messy room and asked him. Immediately came the recluse's famous response: "An ellipse." "Why?" asked Halley. "Because I have calculated it." But, typical of the paranoid Newton, he'd not told anyone.

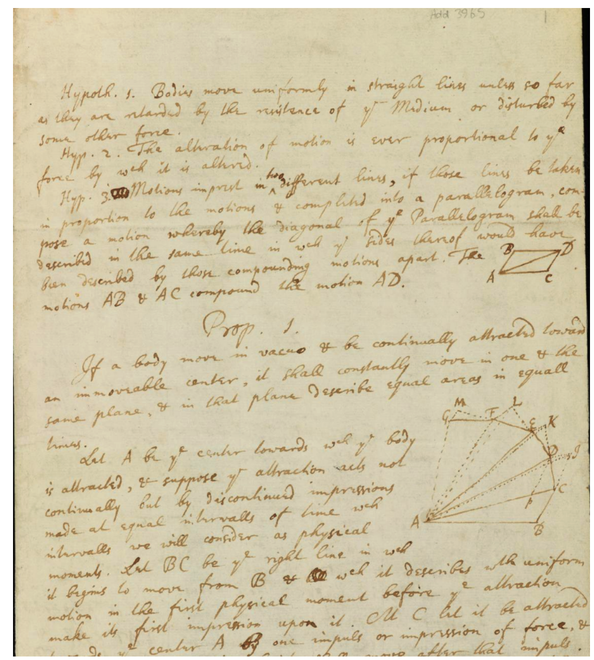
He'd worked out the mathematical rules for the motion of the planets. . . and kept it a secret! Well, the small problem was that while Halley waited, Newton could not find his calculation. A little while later, Halley received 9 pages from Newton that showed: if the force on a planet varies like the inverse square of the distance from the center, then the orbit's shape must be a conic (a parabola, ellipse, circle, or hyperbola). And, he showed

[7] Halley was quite a guy and is worth reading about. `http://en.wikipedia.org/wiki/Edmond_Halley`

that if the orbit is an ellipse, that the force of attraction must be an inverse-square. This pamphlet became known as , *De Motu Corporum in Gyrum* (*On the Motion of Revolving Bodie*s). *De Motu*, as it's known, was a summary of the first book of his magnificent work. Figure 5.4 shows a page of De Motu in Newton's hand that he later prepared for his correspondent-friend, John Locke. This electrified Cambridge and London and set Hooke's teeth on edge as he'd guessed some of the same conclusions and again insisted that Newton had stolen his ideas, this time on gravitation.

Halley realized what Newton had done and implored him repeatedly to write it all. Newton finally agreed and went into one of the historically most intense periods of concentration ever embarked on by anyone. For two years he worked night and day, forgetting to eat, wandering around Cambridge without regard to his surroundings. Thousands of pages of manuscript littering his quarters along with days' worth of uneaten food. Two years! Eventually he emerged with the first book of what was to be three volumes of *Philosophiæ Naturalis Principia Mathematica*, or the *Mathematical Principles of Natural Philosophy* affectionately known ever after as "The Principia." It was all there in Latin. His laws of motion and gravitation, but also of fluids and the strengths of materials. He'd pestered scientists and astronomers from around Britain for data on the planets and the tides. He'd made measurements of motion in his own lab. He let his alchemy furnace go out forever as he worked solely on his system of the world. The arguments were mathematical and constituted the first workable system of nature. He continued to hide his calculus, preferring to speak in terms of limits and extrapolations using geometrical constructions, surely backed up by his own private calculus based calculations. Principia went through three editions after the original 1686 start, often with him revising his last chapter, which was more philosophical, but also with successive furious deletions of the names of rivals.

Hooke had persuaded the Royal Society to act as the publisher of Principia and Newton dedicated it so. But the coffers of the Society were dry when it came time to print as they had used up their entire accounts in a lavishly illustrated two volume History of Fishes. So Halley took a deep breath and paid for the initial publication himself. This of course led to his active interest in encouraging Newton to push the book off at booksellers and libraries himself. Never was there a more generous gift to science than Halley's unselfish gesture. And for a book that only a few people in the world could read, but a book that quite possibly initiated the Enlightenment and people's relationship to our universe.


Add 3965

Hypoth. 1. Bodies move uniformly in straight lines unless so far as they are retarded by the resistance of ye Medium or disturbed by some other force.

Hyp. 2. The alteration of motion is ever proportional to ye force by wch it is altered.

Hyp. 3. Motions imprest in two different lines, if those lines be taken in proportion to the motions & completed into a parallelogram, compose a motion whereby the diagonal of ye Parallelogram shall be described in the same time in wch ye sides thereof would have been described by those compounding motions apart. The motions AB & AC compound the motion AD.

Prop. 1.

If a body move in vacuo & be continually attracted toward an immoveable center, it shall constantly move in one & the same plane, & in that plane describe equal areas in equall times.

Let A be ye center towards wch ye body is attracted, & suppose ye attraction acts not continually but by discontinued impressions made at equal intervalls of time wch intervalls we will consider as physical moments. Let BC be ye right line in wch it begins to move from B & wch it describes with uniform motion in the first physical moment before ye attraction make its first impression upon it. At C let it be attracted towards ye center A by one impuls or impression of force, &

Figure 5.4: A portion of De Motu in Newton's hand in 1684 written for John Locke who was an early reader of Principia.

5.3 Getting Going

We don't need to deeply study Newton's laws for our purposes in Particle Physics like you might in a general physics course, but his whole system is built on two ideas that do matter to us: they are the idea of **mass** and the concept of **momentum**. Let's move.

In Chapter 3 we discussed the rules that govern any object's motion, whether it moves at a constant speed or accelerates. What we conveniently avoided was how that motion is caused—on that Galileo had little to say. Of course Aristotle had something to say about everything and he insisted that motion[8] is not for free, that one always needs to apply a force to keep something moving, or to start it moving from rest.

[8] Remember, unnatural motion requires a push, natural motion just happens. For him.

One of the many ways that Isaac Newton got into the textbooks was to say to Aristotle: "no." Constant motion *is* free. It's only accelerated motion that requires payment in the form of a force. Further, while Aristotle simply declared what his rules were, Newton built the first-ever mathematical model describing all motion. Remarkably, his model has functioned for four centuries and still forms the basis of mechanical and civil engineering projects.

Definition: state.
The circumstance of an object's position and it's momentum define its "state."

Here's what he said: in order to change the "state" of motion of any object requires the application of a force. To start something moving from rest? Apply a force. To speed up or slow down something already moving? Apply a force! To cause something to deviate from a straight line? Yes. Another force. To keep something moving at a constant speed? No force required, thank you. So whenever there's a change of velocity,[9] a force is at work, so forces are responsible for acceleration.

[9] Hold that thought. The difference between speed and velocity is crucial here. More in a bit.

5.3.1 Impulse

To get something up to speed, you must whack it or shove it—either a sharp collision or a steady push increases the speed of an object. Push harder? More speed. Push longer? Again, more speed. And as you know from any sport involving a collision, something that's moving fast can in turn exert a bigger force than something that's moving slowly.

So let's codify that everyday notion into a formula. Let's imagine a force, F that pushes during some time interval, Δt. A whack means that Δt is small (like a golf club hitting a ball) while a steady shove (like a rugby scrum) means that the force is slowly applied so Δt is larger. Here's a trial formula to reflect the fact that either (or both) circumstance changes the state of motion of an object:

Definition: $\propto$.
"proportional to" or sometimes "goes like" as in "this function goes like the square"

$$F\Delta t \propto \Delta v. \tag{5.1}$$

The application of a force for a time interval means that the speed changes in proportion. Increase F, Δt, or both and the speed goes up. The quantity on the left side is called the Impulse in physics. It's the sports-quantity. Any game involving a ball involves impulse. The quantity on the right implies that the speed changed and of course if the speed changed, then the object accelerated.

But now let me ask you: Suppose I apply a force of 100 pounds for 60 seconds to a Volkswagen and and you apply a force of 100 pounds for 60 seconds to a little red wagon. Will the resulting Δv be the same for both vehicles? Of course not. The little red wagon will gain more speed than the Volkswagen (regardless of what color it is). So Eq. 5.1 is not the whole story. What's missing is the reluctance that any object has to being accelerated, which has a name: inertia.

Definition: Interia.
Inertia is the reluctance that an object has to being accelerated.

5.3.2 Newton's Mass

Mass is a toughy. Here's how he defined it in the *Principia*:

> Mass is the ...“quantity of matter...arising from its density and bulk conjointly.”

There you go. Useful? No? What he seems to be saying is that mass is the product of density and volume. But, he doesn't tell what density is which is why it's not helpful. When people invent whole sciences, they also need invent a language! That his words don't quite work shows that his concepts and his mathematics are a bit ahead of him. In any case, you do perfectly well—at least in solving homework problems or building bridges—to accept the idea that mass is the amount of the “stuff” in an object and that it's also the quantitative measure of an object's reluctance to be coaxed into changing its motion.

Here begins our love-hate relationship with mass.

Mass is an object's resistance to being accelerated. *Key Concept 10*

What is the nature of Mass? *Key Question 8*

Figure 5.5: In 1971 Alan Shepard smacked a golf ball on the moon with a makeshift 6 iron. He'd need to apply the same force up there that he would on Earth in order to achieve the same acceleration—but on Earth there would be air resistance, so it would not have gone as far.

At the deep level of elementary particles, mass confuses us, perhaps in a different way from how it confuses college freshmen. We think that mass may actually not be an actual property of object, but rather a result of an object's interaction with a spooky field that sprang into existence just after the birth of the Universe. Now, in the 21st century, we've got a whole new set of neuroses about this subject, as understanding it occupies almost the entire Particle Physics community. So, Mass has been a problematic subject since its beginning in Newton's hands.

5.4 The "Quantity of Motion"

We just developed a sense that our hand-built, car-pushing formula, Eq. 5.1 has to depend on speed and mass and so we'll just add it in on the right-hand side to get:

Pencil 5.1.

$$F\Delta t = m\Delta v \tag{5.2}$$

Definition: **impulse .**
is the quantity which is the force applied to an object throughout at time span, Δt. It's equal to the change of the mass times the velocity experienced by that object.

Being more explicit, impulse which when spelled out is

$$F\Delta t = mv - mv_0 = \Delta(mv). \tag{5.3}$$

So since the mass of the Volkswagen is much bigger than the little red wagon, the same force applied through the same time results in a smaller speed change for the former, rather than the latter. Think of it this way: the numerical product of $F\Delta t$ (which was the same for your push as it was for my push) is shared by m and Δv so more m leaves smaller Δv and of course, a smaller m means more is left for Δv. This collecting m and v together proves to be useful.

Everyone knew that Aristotle's ideas about motion were silly—what's pushing on a projectile? He danced around this and said: the air rushing around from the front to push the object from the back. What if it's an arrow that slices through the air going forward? If you shoot it tail first are we to believe that now the air pushes on the arrowhead that before sliced through it? Seriously? That's all Aristotle's got.

5.4.1 Momentum

Newton's second good idea is the concept of "momentum" which he called the "quantity of motion"—a nice description, I think. The idea that a moving object possessed *something* —some quality—was pretty hard to ignore. But, nobody could figure out how to describe it for 2000 years before him. Aristotle just denied it: "No," a moving object doesn't possess any quality. Galileo vaguely said "yes," there is something "in" a moving body that he called *impetio*. Kepler seemed to say "yes." Descartes definitely said "yes." Newton agreed with his 17th century predecessors but made the idea useful.

What he concluded was that the "quantity of motion" is **momentum**. Keeping with tradition by using the symbol "p" as its nickname, momentum is:

$$p = mv. \tag{5.4}$$

Definition: **Momentum.**
$\mathbf{p} = m\mathbf{v}$
Momentum is proportional to speed and mass. This is specifically "linear momentum." It is a vector because $\mathbf{v}$ is a vector.

We can continue the manipulation of Eq. 5.3 and restate it one more time in terms of momentum:

$$F\Delta t = \Delta p. \tag{5.5}$$

We're going to find that momentum is most important in our particle physics story—much more so than Force, velocity, or acceleration. We'll use it over and over in different guises.

Equation: **Momentum.**
$p = mv$

Box 5.2 The relationship for all sports using a ball.

Equation 5.5 works in three ways: either you know the force and you use the formula to calculate the change in momentum in a given time. Or, you know the change in momentum and you use it to calculate the total force in a given time, or you adjust the time for a given force.

Now you've gotten the formula that governs all sports involving whacking one thing with another... like baseball or tennis—or football. Think about what you almost always want to do: you want to make the ball go faster after you hit it. That means, you want the change in the momentum to be the highest possible. So, Equation 5.5 tells you how: you hit the ball as hard as you can (that's a large F) and you get "good contact" (which means you hit the part of the bat or racquet or club where you can touch the ball as long as possible... which is the largest Δt).

This also explains how airbags and bumper-crumple zones in automobiles work. There, you know what the change in momentum is... it's

$$m \times (v_{\text{after}} - v_{\text{before}})$$

That's the *change* where the final velocity is zero (the car stops). The initial velocity is fixed and so v_{before} has to be divided up between the force and the time in $F\Delta t$ where the force is applied, say to your bumper. High force is not good for the occupants inside the car, so this leads to the design goal of spreading out the time—large Δt so that the force will be smaller. This is the same reason that you bend your legs when you jump off a table and hit the floor.

Momentum Is a Vector

Because velocity is a vector, momentum is also. In the next chapter when we consider collisions the direction-part of the momentum vector will play a crucial role. For that matter, since momentum and force are vectors the actual general statement about Impulse is:

$$\mathbf{F}\Delta t = \Delta\mathbf{p} \tag{5.6}$$

5.5 Newton's Famous Three laws

Newton's Momentum and Mass are at the heart of his three laws of motion. Let's go through them in words, and then one of them in more detail in symbols.

Newton's 1st law of Motion says that anything that's moving at a constant speed (which could be zero) will continue in that way unless a force acts on it. That's a statement about inertia—resistance to acceleration. (He inherited this from Galileo, but gave it a quantitative meaning.)
Newton's 2nd law of Motion says that momentum is changed when a force acts on an object for a duration of time. Or, you might have learned it as a defining statement about "force" namely:

Force is equal to the rate of change of momentum. *Key Concept 11*

Newton's 3rd law of Motion is subtle. It says that if you push on something—anything and with any amount of force—that object will push back with exactly the same force. We'll think harder about the 3rd law when we talk about collisions.

Your immediate reaction might naturally be to wonder how anything moves at all, since there appears to be a balance from this rule. What matters for an object to feel an imbalanced force *on it*. That it exerts a force *on something else* doesn't affect its own motion. So, if a donkey pulls on a wagon, the wagon pulls back...but the force *on the wagon* is itself imbalanced and so it moves.

5.5.1 Newton's Second law

Most of those problems you might have worked in a physics class are related to the 2nd law, which is all about momentum and how to change it from from the vector version of Eq. 5.3.1. A simple arrangement of that impulse equation, yields the real mathematical definition of Newton's 2nd law:

Pencil 5.2. ✎

$$\mathbf{F} = \frac{\Delta\mathbf{p}}{\Delta t} \tag{5.7}$$

Maybe you've maybe seen this equation (on a tee shirt?) but written in a different way. Inserting back the definition of momentum, $p = mv$ (in one dimension, so we'll drop the vector notation):

$$F = \frac{\Delta p}{\Delta t} = \frac{\Delta(mv)}{\Delta t}. \tag{5.8}$$

Let's assume that the mass of the object doesn't change through the push, so we pull it outside of the Δ and get

$$F = \frac{m\Delta v}{\Delta t} \tag{5.9}$$

and, what's $\Delta v/\Delta t$? That's nothing more than acceleration, so

$$F = ma \tag{5.10}$$

which is the tee shirt version of the famous Newton's Second law of motion.[10]

[10] The more complete statement of Newton's Second law is Eq. 5.8.

Notice that if you isolate the acceleration, you get the force divided by the mass. Therein lies the "inertia" nature of mass since

$$a = \frac{F}{m} \tag{5.11}$$

explicitly showing the inverse relationship: the larger the mass (so that $1/m$ is a small number) the harder it is to accelerate for a given force (the Volkswagen). Conversely, if that force is now applied to a lighter object (so that $1/m$ is a larger number) then the acceleration will be more (that's the wagon).

Equation: Newton's Second law.
$F = ma$ (for constant mass)

If you think about it, there are three ways for **p** to change:

1. As we've just seen, if the speed changes, the velocity vector changes, so then the momentum vector changes. That's easy.
2. If the *mass* changes, then the momentum changes. That's easy mathematically, but huh?
3. If the velocity vector changes, but the speed stays the same? Ah. That's interesting and we'll look at that carefully in Section 5.11.

#1 is obvious. If you change the speed, you will change the velocity. #2 is less obvious but think about pinching the opening of a balloon that you've blown up. The balloon's mass consists of the stretchy material of the balloon plus the mass of the air inside. Held out in front of you there is no net force and it's at rest in your hand. If you open the pinch, then the air rushes out *fast* and so the mass of the air inside the balloon rapidly decreases. The consequence of that decrease in mass, a Δm in a time Δt by itself results in a change in the momentum of the balloon! That's how rockets and jet turbines work.

Item #2 at the left requires some thought. Remember what "change" means: something at the end minus something at the beginning. So, the mass in a balloon is decreasing, so $\Delta m = m - m_0$ which is a negative quantity. in fact, the mass escapes out the nozzle and the momentum change is in the opposite direction (the negative sign of Δm becomes a directional sign of the opposite of the direction that the mass goes.

Weight

We tend to mix up the units of weight and mass at the grocery store, for example. A gram is a unit of mass, while an ounce is a unit of force. A kilogram is just a 1000 grams and still a measure of mass and a pound is 16 oz and still a measure of weight. But we get away with using both systems since we tend to buy things and compare them on Earth. If we had a Mars colony, well then there would be trouble. That 5 Earth-pound bag of Gold Bond flour at Kroger would be 13 Mars-pounds. But in each location, it would still be a 2.27 kg bag of flour. How can that be?

Table 5.1: Units for quantities used in weight calculations.

	English	Metric	Conversion
acceleration	ft/s^2 g = 32.2 ft/s^2	m/s^2 g = 9.8 m/s^2	1 ft/s^2 = 0.305 m/s^2
mass on Earth:	slugs a mass of 1 slug → weight of 32.2 pounds	kilograms (kg) a mass of 1 kg → weight of 9.8 N	1 slug = 14.59 kg
force	pounds (lb)	Newtons (N)	1 lb = 4.45 N

Weight is the force that the Earth exerts on objects on its surface while as we've seen, mass is the amount of inertia that an object possesses. The inertia is determined by measuring the force that it would take to accelerate the object to a given amount. What Galileo showed was that the acceleration due to gravity on the surface of the Earth is a particular value—he presumed was a constant. If the ground suddenly disappeared, then everything with mass would start to fall with that acceleration, g. But happily the ground pushes back and things are stable on the surface. So from Newton's Second law, when we have a mass and we have an acceleration, we can calculate a force and we define that particular force of attraction by the Earth the weight. Let's call it w and we can write it out:[11]

[11] Don't be confused. This is just Newton's Second law, but when the acceleration is the particular value of g, the force is the particular kind of force we call "weight."

$$w = mg \tag{5.12}$$

We can measure the weight by making use of the fact that the Earth pushes back with the same value as the weight... when you think of weight, probably a spring is doing the pushing-back. That's your bathroom scale which is calibrated in the U.S. to read that push in pounds. We'll see how the Earth does this in a bit when we get to Newton's other law, that of Gravitation.

Unfortunately in the English system, the unit of mass is “ slugs.” So we can collect our units appropriate to Newton's Second law in Table 5.1. In the last column, the standard abbreviations are shown as well.

Example 5.1

Weight

Question : If I weigh 200 pounds on Earth, what is my mass in slugs? In kilograms?

Solution:

$$\begin{aligned} w &= mg \\ m &= \frac{w}{g} \\ &= \frac{200}{32.2} \\ m &= 6.2 \text{ slugs} \end{aligned}$$

To convert to kilograms, we can do conversions within the correct quantities using Table 5.1. So let's do w(English) → w(metric):

$$w(\text{metric, in N}) = w(\text{English, in lb})\,\frac{4.45\ \text{N}}{1\ \text{lb}}$$
$$w(\text{metric}) = 200 * 4.45 = 889.6\ \text{N}$$

Now calculate the mass in kilograms like before:

$$\begin{aligned} w &= mg \\ m &= \frac{w}{g} \\ &= \frac{889.6\ \text{N}}{9.8\ \text{m/s}^2} \\ m &= 90.8\ \text{kg} \end{aligned}$$

We can check with Fig. 2.4 and see that we got the right answer.

Example 5.2

Jumping from a Step

Question : Suppose I'm 200 pounds, or 90.7 kg and I jump from a step which is 1 meter high, about a yard. Now, with my artificial knees, I probably shouldn't do that, but were any of us to do so, we'd automatically flex our knees on landing and here's why.

Solution:

Equation 5.3.1 tells the story. On the right hand side is the change in momentum. Now, just before my feet hit the floor, I'm traveling downward with the greatest velocity possible. The *change* in momentum is

$$\Delta p = p - p_0 \tag{5.13}$$

where p is the final momentum, and p_0 is the initial momentum. Now, the final momentum is easy. It's m times the final velocity, which is "stopped" or 0. So,

$$\Delta p = 0 - mv_0$$

where v_0 is the initial velocity. Although we didn't talk about it, if I start from rest, then the velocity of an object falling under gravity is $v = \sqrt{2gx} = \sqrt{2 \cdot 9.8 \cdot 1} = 4.4$ m/s.

The final momentum is then $p_0 = mv = 90.7 \cdot 4.4 = 402$ kg m/s. Now, all of this momentum has to be taken up by the combination of F (which is applied through my legs to my poor knees) and Δt which is the amount of time that that force is applied. If I land stiff-legged, say my impact happens in 0.1 seconds, then the force applied to my knees would be

$$\begin{aligned} F\Delta t &= \Delta p \qquad (5.14) \\ F &= \frac{402}{0.1} = 4020 \text{ Newtons} \end{aligned}$$

The metric system of force is "Newtons" and here 4000 N is about 900 lbs.

If, I can spread out my shock-absorption by bending my knees...to maybe as much as a second, then I would relieve the force transmitted by a factor of 10!

Example 5.3

Biking at a constant acceleration

Question : In Chapter 3 we described the bicyclist's constant acceleration of 2 m/s^2 and the subsequent increase in speed in time and consequent quadratic increase in distance covered. If I weigh 200 pounds, we found in Ex. 5.1 that my mass is 90.7 kg, how much force do I have to apply to the ground through the pedals and the tires in order to keep up that constant acceleration? What fraction of my weight is this force?

Solution:

This is a simple application of the popular form of Newton's Second law, Eq. 5.10.

$$\begin{aligned} F &= ma \\ &= (90.7)(2) \\ F &= 181.4 \text{ N} \end{aligned}$$

To find my weight, we again can use the same formula with an important difference (we'll call my weight W) and I'll approximate the acceleration due to gravity, which is $g = 9.8$ m/s^2, as $g \approx 10$ m/s^2

$$\begin{aligned} W &= ga \\ &= (90.7)(10) \\ W &= 907 \text{ N} \end{aligned}$$

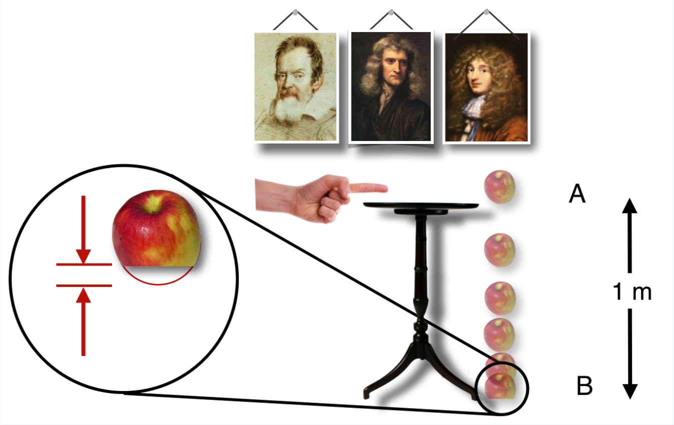
So the force that my legs would have to continuously apply to the pedals, and in turn to the ground through the friction between the tires and the road is about 20% ($\approx 180/900$) of my weight.

How much sustained force is this? Well, suppose we have a stationary bike hooked up to a pulley and a bag with five bowling balls. The force required to keep that bag 'O balls aloft—forever—is the amount of force that I'd have to sustain—forever—to maintain that acceleration. Now is that sensible? Figure 3.8 shows that after 10 seconds of this acceleration I'd be traveling at 20 m/s, which is about 45 mph. So obviously, that's too fast to imagine pedaling a bicycle for 10 seconds. Rather, if it were possible for me to exert 181 N, after about a couple of seconds, I'd be moving around 10 mph and surely at that point I'd stop trying to accelerate and apply just enough force to maintain that speed.

Example 5.1

Apples falling again

Question:



In Example **??** you calculated that the speed that an apple would attain if it was dropped 1 meter would be 4.4 m/s. You did, right? Look at the figure at the left for our new situation. The apple at A is dropped onto the carpet and bruises flat at B, slowing it down to a stop. The carpet applies a force to it which would be pointing up. (You can see the damage in the inset.) If it takes 0.090 s for the apple to stop, what is the average force that the carpet applies to bring it to rest? The mass of an apple here is 0.1 kg.

Solution: This is another application of Newton's Second law where we have:

$$F = \frac{\Delta p}{\Delta t} = \frac{mv}{\Delta t}$$

The change in velocity is of course the velocity that the apple has just before it hits since it's dropped from rest. Putting the numbers:

$$F = \frac{(0.1)(4.4)}{0.09}$$
$$F = 4.9 \text{ N}$$

which is about half of the force of gravity. The apple would probably not bruise.

Now you try it. Drop it on the carpet.

You Do It 5.1. Apple Force

or copy the solution

If the apple in the above example is dropped on a hard floor, then it will stop more abruptly. Llet's pretend that it takes only 0.005 seconds to come to rest. What is the average force that the carpet applies to the apple to make bring it to rest? Would it bruise more or less than on the carpet?

5.6 Circular Motion

Item #3 back on page 149 of how a change in momentum can occur is subtle, but you use it and experience it every day. Suppose you're a passenger in a car going around a curve. When you enter the curve, you're moving north. When you emerge from the curve, you're pointing west. Watch the speedometer and make sure that your driver stays at the same speed through the whole path. So did your speed change?[12] No!

[12] a "Who's buried in Grant's Tomb" question

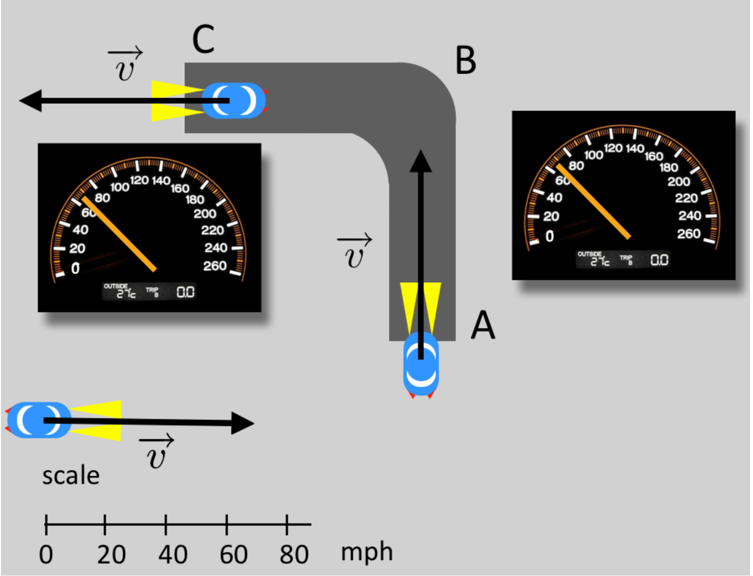


Figure 5.6: At the beginning the car is headed north at point A with a velocity vector shown and the speedometer pictured at that point. The car curves to the left at B and at point C the car's speedometer still looks identical to point A, but the velocity vector is pointing to the west. The speed is constant, but the velocity is different. The scale at the bottom left shows that the length of the velocity vector corresponds to just about 65 mph.

Did your velocity change? Yes! Because, the *direction* of your speed changed. This is shown in Fig. 5.6. What do you *feel* during the curve?

Let's recite the series of events that follow from going around this curve: You stay in the car, so you are being "forced" to deviate from straight-line motion. A variety of mechanisms cause that to happen: your seatbelt, the friction of your pants and the seat, the door pushing on your shoulder. All of these apply a force in a direction to the *inside* of the circle that your car is moving along.

- Your speed didn't change.
- But, your direction changed, so your velocity changed.
- If your velocity changed, and your mass remained the same your momentum changed (#1)
- If your momentum changed, there was a force applied on you (Second law).

These various forces on you all cause you to go in the same circle as the car (Second law) while you are trying desperately to continue to go straight (First law!). This force that causes motion to deviate from a straight line is called "centripetal force" and this is another genius idea of Newton's.

Definition: centripetal force .
is a force that causes an object to go in a curved path. It points to the center of the curve.

The essence of circular motion can be visualized in Fig. 5.7 where a figure is twirling a ball attached to a rope in a circle. Let's ignore gravity for a second and concentrate on the motion in the plane of the rope and ball...and his fist. You know by now what would happen if he let go of the rope. Without the

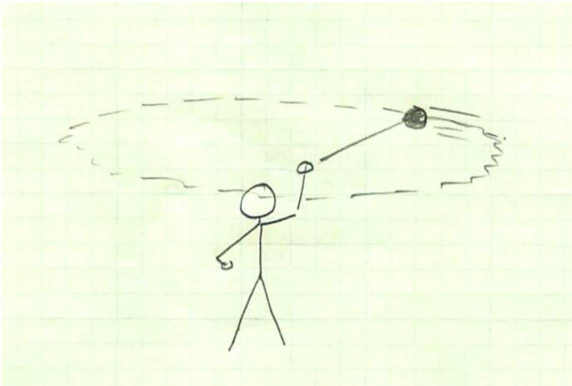
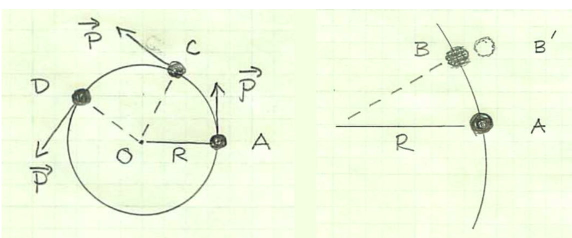


Figure 5.7: Twirling a ball in a circle with a tether.



(a) (b)

Figure 5.8: Looking down from the top of a ball's path.

rope pulling in towards the center, there is no horizontal force and according to Newton's First law, the ball would go straight. So the rope is causing the ball to deviate from a straight line.

Figure 5.8a shows the view of the ball over the head of the figure. The rope, R, is shown and the location of the ball is represented at three spots around the circular trajectory, A, C, and D. This is meant to be uniform motion, which here means that the speed is a constant...like the speedometer in the opening discussion. But of course the velocity is changing, by virtue of the changing direction. But what's represented in this figure is now the momentum, **p**, or here, $\vec{p}$, which is the mass m times the velocity, **v**. Since the speed is constant, the momentum vector has a constant length, but because the motion's direction is around a circle ("not straight"!), the momentum vector is tangent to the circle at all points around the path. Newton's brilliance was to explain this using his three laws.

Figure 5.8b shows a segment of the circle as the ball passes point A in Fig. 5.8a. Newton reasoned that the ball would "like" to go straight, to point B' but that the rope tugs it back to point B. So the ball goes a little, gets tugged back, goes a little further, gets tugged back, and so on. These little tugs were in his mind acting all around the circle, which in the limit of being infinitesimally spaced create a continuous, circular trajectory. This notion of "infinitesimal" was kin to the habit of mind he was developing in the invention of calculus.

Pencil 5.3.

But let's carry this further. Since the momentum at, say A is different from the momentum at, say C (because the direction is different), even though the magnitudes are the same, there is still a changing momentum, a non-zero $\Delta\vec{p}$. If there's a change in momentum of any kind, there's a force:

$$\vec{F} = \frac{\Delta\vec{p}}{\Delta t} \tag{5.15}$$

Let's see what he found. Figure 5.10 shows two strategically placed points on the circle "here" and "there" and the corresponding momenta of the ball associated with each point, $\vec{p}_{\text{here}}$ and $\vec{p}_{\text{there}}$. So,

$$\vec{F} = \frac{\vec{p}_{\text{there}} - \vec{p}_{\text{here}}}{\Delta t} \tag{5.16}$$

Let's get a feel for what this means by actually manipulating the momentum vectors and look at what the numerator gives us in Eq. 5.16. We'll remove them from the diagram and take their difference according to the rules of vector subtraction...by adding. Figure 5.9 shows the process.

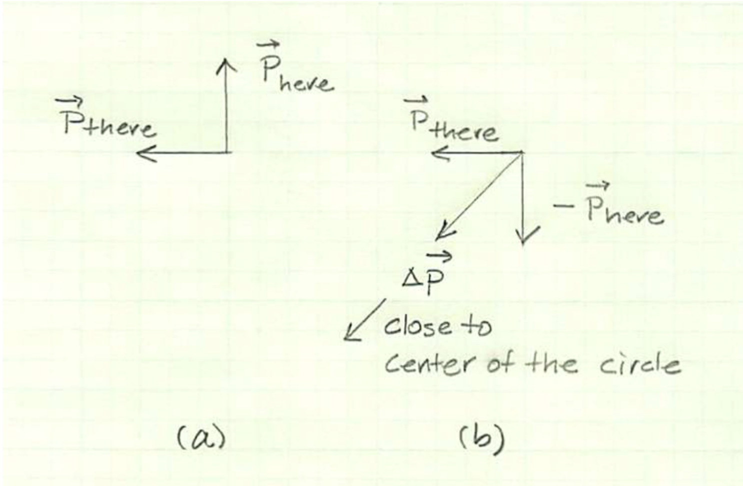


Figure 5.9: (a) shows the vectors brought tail-to-tail with their original orientation unchanged. (b) shows them with $\vec{p}_{\text{here}}$'s direction reversed.

In the left side of the figure, the momentum vectors have just been redrawn. We need to form the difference as in Eq. 5.16 and we remember that the difference of two vectors can be:

$$\vec{F} = \vec{p}_{\text{there}} - \vec{p}_{\text{here}} = \vec{p}_{\text{there}} + (-\vec{p}_{\text{here}}). \tag{5.17}$$

So by reversing the direction of $\vec{p}_{\text{here}}$, we can just add it to $\vec{p}_{\text{there}}$ and get the required difference from Eq. 5.16. This is shown in the right side of Fig. 5.9. That difference is labeled $\Delta\vec{p}$ and its direction is very close to the center of the circle! This was his brilliance! If the "here" and "there" points were closer and closer to one another, then the difference would point closer and closer to the center.

So like we knew all along, the rope is what causes the ball to go in a circle, the combination of Newton's First law with his Second law, and the crucial recognition that momentum is a vector, leads to the demonstration of the force towards the center is responsible for the change of momentum.

This force is that special "centripetal force" that we encountered going around the curve in the car above. All non-straight motions are caused by an centripetal force. If the trajectory is circular, it's easy to see that it points to the center of the observed circle. If the trajectory is uniformly curvy, at each point there can be an instantaneous "circle" and the force would point towards it.

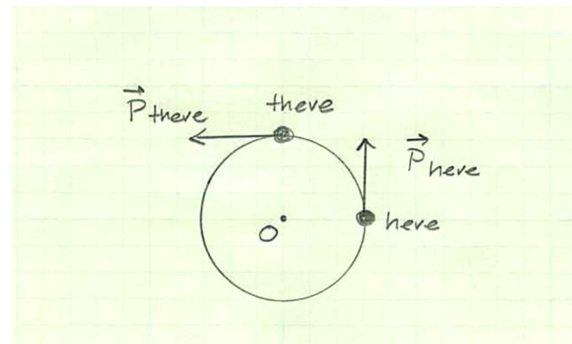


Figure 5.10: Two momenta shown for two stratetically placed points around the circular path.

5.6.1 Centripetal Force and Centripetal Acceleration

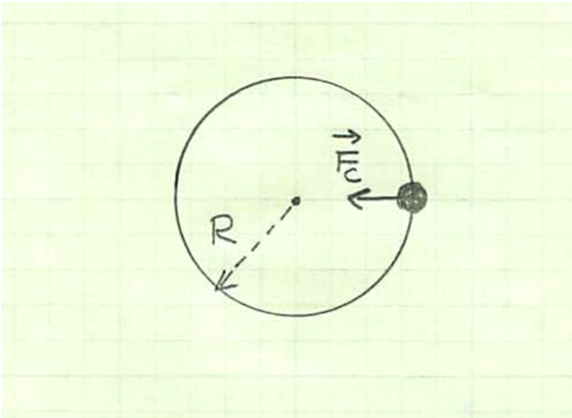


Figure 5.11: The centripetal force shown for a circular path of radius R.

If this understanding of circular motion weren't enough, he went a step further in his paranoid sort of way. He actually found a relationship for what the centripetal force would be and did it both using his new calculus and in a strictly geometrical fashion. The latter he published in *Principia*, and like other such derivations, kept the calculus version to himself. Why? He feared being scooped. Calculus was his private tool for a long time.

Without going into those details, I'll just report the results. The centripetal force is special, and I'll call it F_C:

$$\vec{F}_C = m\vec{a}_C$$
$$\vec{F}_C = m\frac{v^2}{R} \tag{5.18}$$

Here, a "centripetal acceleration" is also assigned and is related to the distance from the object to the center as shown in Fig. 5.11.

$$a_C = \frac{v^2}{R} \tag{5.19}$$

Equation: Centripetal force..
$F_C = m\frac{v^2}{R}$ (mass times centripetal acceleration)

An object traveling on curve requires a force directed to the center. *Key Concept 12*

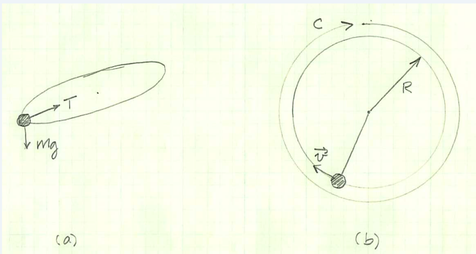
There are two ways to use this concept: If you want something to move in a curved path at a particular speed, you can calculate and apply a precise (centripetal) force—tug it—to make that happen. If you see that an object is *not moving in a straight line*, then you must be able to identify a centripetal force being applied to it! Sometimes identification of such a force is tricky. For example, for our car, what actually causes the car itself to go around a circular curve? Obviously that force is the force of friction between the road surface and the tires of the car. Reduce the stickiness of the road (ice, snow, rain?) and that force of friction is reduced and the force that's possible is reduced, sometimes considerably. You instinctively know this, so you drive slower (reducing v in the numerator of Eq. 5.18 to match the F_C that *can* be produced given the conditions.


Figure 5.12: Cynthia Watt who won the 2015 Big Ten Outdoor Track & Field Championships in East Lansing, Mich. (Photo by Matt Mitchell)

Example 5.1

Hammer Throw.

Question:



The Hammer Throw is an old track and field event. For men, a 16 lb ball (7.3 kg) is attached to a chain that's approximately 4 ft long (1.22 m) and whirled around a circle and let go. Olympic-class hammer throwers spin their bodies incredibly fast—in their last "wind" before release they are spinning less than a second per revolution. Let's call it 0.3 seconds. Figure 5.12 shows a collegiate hammer champion at work.

1. Calculate how fast the ball is moving at that rotational rate.
2. Using that speed, what is the force that their arms must exert in order to keep the weight moving in a circle?
3. What fraction of the weight of the hammer is that force?

Solution:

1. In order to calculate how fast the hammer is traveling around its arc, we have what we need to know: We know how long it takes to make a complete revolution and we know how far it goes in one revolution is the circumference, C, of that circular path. Figure 5.1 shows the forces and the distances for our situation.

$$C = 2\pi R = (2)(\pi)(1.22) = 7.67 \text{ m}$$

So the speed is:

$$v = \frac{C}{t} = 7.670.3 = 25.6 \text{ m/s}$$

(This is about what the measured "escape velocity" is for world-class throwers, who can toss the hammer more than 80 m. Mr Google will quickly tell you that this is about 60 mph.)

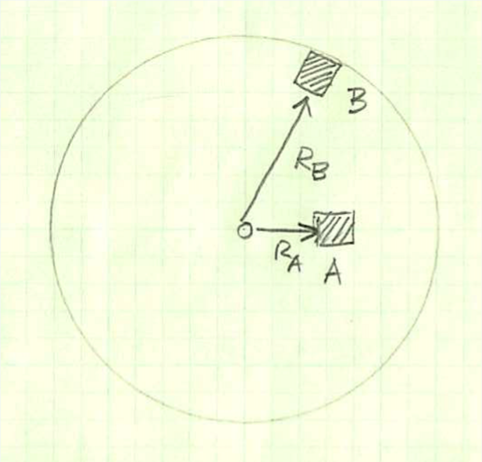
2. The force on the thrower's arms would be the centripetal force of the hammer, $F_C = m\frac{v^2}{R} = (7.3)\frac{25.6^2}{1.22} = 3900$ N which is about 880 pounds! The womens' hammer is 4 kg, so the force that they would experience for the same speeds would be about 460 pounds.

3. The weight of the men's hammer is $W = mg = (7.67)(9.8) = 75$ N which is 1/50th of the centripetal force: $\frac{3900}{75} = 52$.

Example 5.2

Playground merry-go-round.

Question:



The figure shows an upper view of a merry-go-round with two children at two different distances from the center. What is the force of friction required to hold child A on board? Is the force of friction required less or more to hold that same child on at B? $R_A = 3$ m and $R_B = 5$ m. The merry-go-round makes one complete revolution in 10 seconds and the child weighs 50 pounds, so 22.7 kg.

Solution: In order to know the force of friction required, we need to know the speed, which we get just like we did in the hammer throw example.

$$v_A = \frac{C_A}{t} = \frac{2\pi R_A}{t}\frac{(2)(\pi)(3)}{10} = 1.9 \text{ m/s}$$

The force is then $F_A = m\frac{v_A^2}{R_A} = (22.7)\frac{1.9^2}{3} = 27.3$ N which is about 6 pounds of force. Maybe sticky tennis shoes?
If the child stands all the way at the rim, is it harder or easier to stay on? A different, but related question: is that child moving faster or slower than the child who's closer to the center? You've all done it and you know that it can be very hard sometimes at the rim of such a playground device. So the faster it goes, the higher is the force required. If the force is just a constant (like friction), then we can derive a relationship that will tell us the force as a function of radius. Stay with me here. Your pencil is out, right?

$$F_C = m\frac{v^2}{R} = m\frac{(2\pi R)^2}{R} = m(4\pi^2)\frac{R^2}{R} = 4\pi^2 mR$$

...the further out you venture towards the edge, the higher is the force you need to apply to stay moving in a circle, linearly with your distance from the center. By the way, have you seen the multitude of YouTube videos of idiots speeding up a merry-go-round with a motorcycle laying on its side with the tire powering the rim? This is Darwin at work.

Example 5.3

Racing.

Question:

The first turn at the Indianapolis 500 raceway has a radius of curvature of about 800 feet (about 244 meters) as you can see in the picture. Racing tires are "flats" and have maximum rubber on the road surface for the most friction.

1. If the force at which sliding would start to happen is 1500 lb (6700 N), what is the maximum speed that a driver can achieve? An representative weight of an indy car is about 1600 pounds, or about 725 kg of mass and 7100 N.
2. How many factors of g does this force represent?

Here the force is fixed—it's determined by the road surface, tires, and weight of the car—and we need to know the speed. A little bit of speed increases the force quickly, since it's proportional to v^2. A race is all about speed, so a lot of engineering goes into keeping the car firmly on the track.

Solution: From the centripetal force equation, we can solve for the velocity:

$$F_C = m\frac{v^2}{R} \rightarrow v^2 = \frac{RF_C}{m} = \frac{(244)(6700)}{725} = 2250 \text{ So the speed is } v = \sqrt{2250} = 47 \text{ m/s ...which is about 103 mph.}$$

The number of "g's" that the driver would feel is the centripetal acceleration divided by g, g's $= \frac{1}{g}\frac{v^2}{R} = \frac{2250}{(9.8)(244)} = 0.94$. In fact, an Indy car can take that curve above 200 mph. Why is this different? First, the tracks are banked at about 9°; next there are front and rear air-foils that create a down-force, so hundreds of pounds of more "stick" between tires and road (the driver can adjust this); and finally, drivers "aim" at the turn more strategically than just going in a circle.

Drivers pull as much as 3-4 g in each turn! They must fight this force in order keep his or her neck straight four times around the track, for 500 miles: 800 times that the driver has to exert this strenuous resistance. Now you know why this race is referred to as "grueling."

The big use of centripetal force will re-emerge in the great triumph of Newton's, namely his law of Gravitation which we'll encounter in Chapter **??**. This really made his bones. But next, let's bang things together.

Chapter 6

Collisions

Smashing Things Together

Christiaan Huygens by Caspar Netscher, 1671.

Christiaan Huygens, 1629-1695

"There are many degrees of Probable, some nearer Truth than others, in the determining of which lies the chief exercise of our Judgment." *Cosmotheoros (1695)*

Isaac Newton wasn't the only smart guy around. Although he had respect for only a few contemporaries, Christiaan Huygens, a Dutch gentleman of means—and oh, by the way, a genius astronomer, inventor, and mathematician of such esteem that he was a elected foreign member of the British Royal Society—was one of them. The other, much to his consternation, was his rival for the invention of calculus, Gottfried Wilhelm von Leibniz. It's interesting that neither of these two had academic day jobs. Huygens did what he pleased and Leibniz was a diplomat for the House of Hanover for much of his life.

6.1 Goals of this chapter:

- Understand:
  - The meaning of Momentum Conservation
  - How to use the momentum conservation equation in one dimension
  - How to draw the Feynman Diagram for collisions of two objects
- Appreciate:
  - How momentum conservation works graphically in two dimensional collisions
- Be familiar with:
  - The history of understanding collisions
  - Huygens' life

Figure 6.1: The surface of Saturn's moon, Titan as captured by the ESA space probe, Huygens.

Figure 6.2: Read about Christiaan Huygens in the The MacTutor History of Mathematics archive.

6.2 A Little Bit of Huygens

Christiaan Huygens grew up in a privileged household in the Hague where his father, Constantijn Huygens, was a diplomat and an advisor to two princes of Orange. The Huygens' home was unusual. Guests included Descartes and Rembrandt, whom Constantijn helped support. Constantijn was friends with Galileo, a poet and musician and was even knighted in both Britain and France and so perhaps it was logical that he would provide for Christiaan's home-schooled.

When ready, he was sent to Leiden University to study law two years after Newton was born, whereupon Christiaan largely discovered professional mathematics. A pattern we've seen before, although unlike Galileo's case, Christiaan's mathematical education had included encouragement from Descartes when he as a child. So his father was much more understanding...and the need for a "job" was never an issue in this family.

Huygens became interested in astronomy and learned to grind and polish lenses in a new way which led to the development of a telescope of unparalleled quality. Among his first discoveries the large moon of Saturn, named Titan and then the resolution of Saturn's rings, which to Galileo had been only a confusing, unresolved bulge. The space probe "Huygens" was designed by the European Space Agency and landed on Titan on January 14, 2005. From Christiaan Huygens' first glimpse of Titan, to the photograph of its surface (Figure 6.1) is a nice story.

His astronomy led him to a need for accurate time, whereupon he invented the first pendulum clock (think "Grandfather") —by inventing a pivot that made the pendulum swing in the pattern of a cycloid, rather than strictly circular motion of an unhindered pendulum. Galileo had shown that the period of a pendulum was independent of the amplitude of the swing ("isochronous") but this is only approximately true when the amplitude is small. Huygens showed mathematically and then by construction that if the bob can be made to take the path of a cycloid, that the motion would be isochronous even for large swings...and hence useful in a clock. He also carefully considered the forces on an object in circular motion and the derivation of the centripetal force in Chapter 5 was actually obtained first by Huygens for a circle. However, he was confused by the tendency of an object to move away from the center of a circle and called the force *out* "centrifugal force." Newton also was initially confused by this but figured out that gravity, for example would pull *in* and hence coined the name "centripetal" to contrast it to Huygen's idea.[1] Newton's analysis was general and included circular, elliptical, parabolic, and hyperbolic orbits, so we tend to credit him with the correct understanding of curved motion.

[1] It should be noted that Huygens was also a committed Cartesian and interpreted this centrifugal force as the force of a fluid on an object, the source fo that fluid for an astronomical object being something similar to Descartes' little balls in vortices.

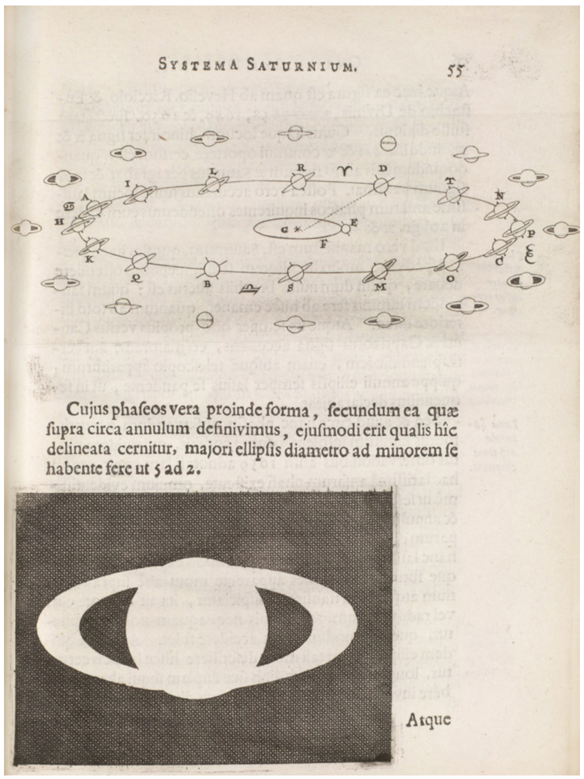


Figure 6.3: huygenssaturn

Huygens traveled widely and spent considerable time in Paris in multiple long stays. He visited Britain many times as well, and when he thought he was dying (he was often in frail health) bequeathed his notes to the British Royal Society. The British Royal Ambassador wrote, "...he fell into a discourse concerning the Royal Society in England which he said was an assembly of the choicest wits in Christendom...he said he chose to deposit those little labours...in their hands sooner than any else..." His scientific circles were very broad and he counted as among his friends, most of the intellectuals of the day, including Boyle, Hooke, Pascal, and indeed, Newton whom he visited shortly after *Principia* was published.[2] Christiaan Huygens died at the age of 66 four years after his visit with Isaac Newton, who referred to him as one of the three "great geometers of our time."

[2] Leibniz met Huygens in Paris and credited him with mentoring his mathematical development as a young man.

For our purposes, it was Huygens' consideration of collisions that is his legacy. Descartes had considered the problem of colliding two objects together and he applied his notion that all motion in the universe is conserved to such problems. But he was confused about just what was conserved and didn't have an appreciation of vectors or momentum. Putting Huygens' work together with Newton's gives us our modern ideas about how the total momentum of a system of colliding objects is preserved and that's our concern in this chapter.

In EPP we're all about collisions. We make huge facilities to do nothing but crash things together and we still use the same ideas and language first invented in the 17th and 18th centuries, albeit fancied up for modern applications.

6.3 Early Ideas About Collisions

One of the treasured concepts for physicists is the idea of the Conservation of some quantity. We'll make use of that idea over and over.

While we have a sophisticated justification for this affection for conserved things, even before Conservation Laws were a notion at all, natural scientists had an intuitive sense that Nature seemed to preserve some qualities. The first such serious assertion became known as the "Conservation of Momentum." Descartes started it all when he declared that the total "amount of motion" is unchanged, just shared among all of the various bodies in the cosmos.

Some of the quantities which appear to be absolutely conserved in Nature are:

- momentum
- energy
- angular momentum
- electric charge

As we'll see, energy and momentum are actually a single quantity which is conserved.

Definition: Conserved Quantity.
A conserved quantity in physics is one that is unchanged during a time interval—typically a "before" and "after" some event. These statements are referred to as "Conservation Laws."

His model of the universe assumed that it was originally kick-started with all of the material bits set into initial motion and all of this primordial motion shared among all objects forever. Add those individual bits of motion up at any time and you get the amount of motion you started with. "Bits of motion" for him meant: speed. This eventually led him to his Big Idea that outer space was filled with various sized balls which were originally rotating together in a great "vortex," and in that way dragging the planets along with them. Those balls constituted his choice as material cause of the orbits of the planets

Wait. *What balls?*

Glad you asked. *I'm sorry? Oh, you mean how did he come up with this idea? As a materialist he was forced to postulate some contact force between objects to make anything move, including the planets. Descartes' philosophy influenced his science: it was top-down. Postulate a cause and then work it out. But don't postulate motion without first setting up the mechanism. His commitment meant: I see the planets moving, so something has to be pushing them. Balls sharing the original primordial motion is as good as anything else.*

While Newton dismantled Descartes' vortices and after considerable massaging the preservation of the total motion was an idea that was still around after Newton was finished. Certainly, the rotations of the planets (and the Moon) about their centers is a modern reflection of the original rotations of the matter that under gravity slowly coalesced into their solid masses. Still orbiting, after all these years.

6.3.1 Modern Ideas About Collisions

Let's think about how Descartes might have come to his conclusion and how he was wrong. Figures 6.4, 6.5, and 6.6 picture three familiar kinds of collisions.

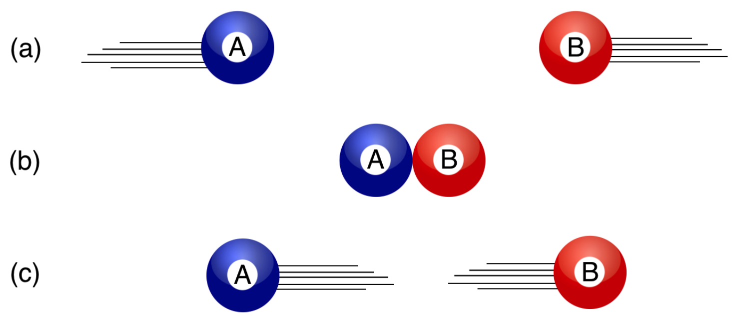


Figure 6.4: Two identical billiard balls are moving towards one another with the same speeds in (a), collide at (b), and recoil from one another at (c).

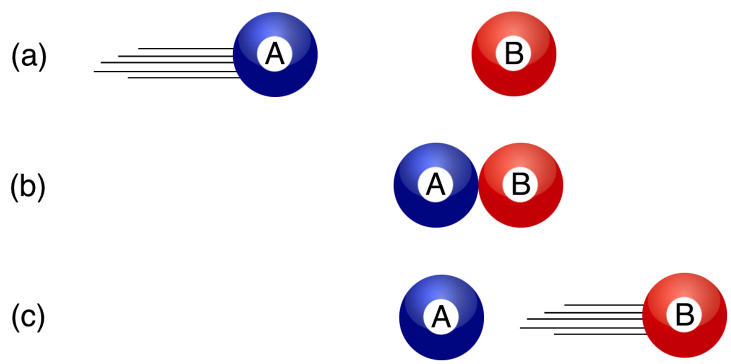


Figure 6.5: One billiard ball B is sitting still and another ball A is headed right for it in (a), they collide at (b), and then B shoots off to the right, leaving A stopped where they hit.

From your own experience you can guess at the outcomes of each. Look at Fig. 6.4: two billiard balls (each with the same mass) are in a head-on collision. B comes from the right at speed v and A from the left, also at speed v. What happens? Descartes would have said that the total speed at before they collide

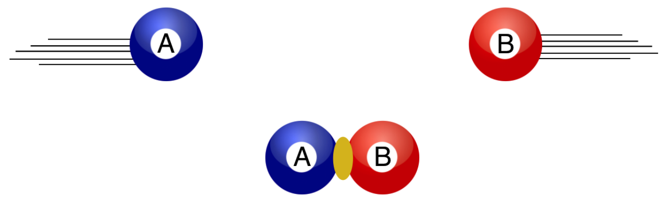


Figure 6.6: Two identical billiard balls are moving towards one another with the same speeds in (a), collide at (b), and stick together (that's gum between them).

is $2v$ and so the total speed after they collide would also be $2v$. And he would be right. In this collision, both balls would recoil from one another and just reverse their original motions.

In Fig. 6.5, now B is sitting still minding its own business and A comes along with speed v and strikes it. This is a familiar "dead ball shot" in billiards and I'll bet you've done it with pool balls or something similar. The outcome is that A stops in its tracks and B shoots out with the same speed that A originally had, so Descartes would still be right.

But now look at Fig. 6.6. It's the same as Fig. 6.4 except now the balls stick together when they collide. The result is that they stop dead—much like two cars crashing together or a running back and fullback colliding at the line of scrimmage.[3]

Here, Descartes would get it wrong: he would have said that the original motion was $2v$, but we know that the final "motion" is 0.

Obviously he didn't appreciate the importance of the *directions* of the *velocities* in the addition to their magnitudes. That is, he didn't know about vectors. In the sticking-together collision we instinctively know that the result is: $v_A - v_B = 0$ so that they stop. *Speed* is not the conserved quantity. Newton's *momentum* was the key, defined as the vector quantity, $\vec{p} = m\vec{v}$, magnitude and direction. He had the beginnings of an appreciation for the direction of momentum, filling a concept-gap that neither Descartes nor Galileo had come to on their own. But Huygens got it right: he understood the idea of vectors.[4]

Collisions were a fascinating study for those working in the 17th century. Everyone understood that friction confused the real picture, so people relied on colliding pendula where these effects were reduced. One has visions of everyone having many "executive toy" contraptions in their workrooms, changing out the bobs and causing clacking collisions with careful measurements of the outcomes. Huygens made use of his home-town canals in Amsterdam as way to collide masses in a controlled way.

Amsterdam's canals provided a near-frictionless racetrack for accelerating particles. He would station a colleague on a canal-boat with a pendulum which would collide with one held by someone on the shore. As the boat went by, nearly frictionless collisions were created and, he was able to get study the collision from the point of view of a "fixed" coordinate system (say, the shore-guy when the boat-guy went by) and

[3] where maybe "dead" is an unfortunate choice of phrase!

[4] Actual vector notation wasn't invented until the late 19th century.

Figure 6.7: A woodcut from Huygens' work illustrating his little 17th century particle accelerator. One person with a pendulum is in a boat on an Amsterdam canal and the other is on the pavement with a pendulum as well. Back and forth they went, with different masses dangling from their ropes. It must have been a sight.

the "moving" coordinate system (the boat-guy). His geometrical explanation was very complicated, but essentially correct. I'll describe it here in modern language.

I want to concentrate on the collision of Fig. 6.5. We'd better be able to predict the outcome of this mundane example to believe Newton's mechanics.[5]

[5] Notice, that Descartes is happy and applauding in the background since the total amount of motion before and after the collision is unchanged. But we already know that his theory is busted.

6.4 Impulse and Momentum Conservation

Let's develop the simple machinery from Newton's ideas. Remember from Eq. 5.6 that the the momentum change of an object is equal to the force that alters its motion times the time through which that force acts.

$$\mathbf{F}\Delta t = \Delta\mathbf{p} \tag{6.1}$$

Think about what happens when object A collides with object B. Let's imagine that A is your left hand and B is your right hand. Now give Huygens a big hand (shall we?) and clap them together. You can make two simple, but important observations from this:

- When they just start to make contact your left hand begins to exert a force on your right hand and your right hand exerts a force on your left. Is there any difference? Does one hand fly away from the other? No...from Newton's 3rd law, these forces are equal (and opposite). So: $\mathbf{F}_{\text{left}} = \mathbf{F}_{\text{right}}$.
- Nothing is perfectly stiff, so there is some elasticity or crumpling or bending and so the total force that each exerts is spread out over the same times as they continue to press against one another. Is the time your left hand is in contact with your right hand any different from the time that your right hand is in contact with your left? No, of course not. So: $\Delta t_{\text{left}} = \Delta t_{\text{right}}$.

So let's look at the impulse experienced by any two colliding objects A and B. Here are those relations:

Pencil 6.1. ✎

$$\mathbf{F}_A\Delta t_A = \Delta\mathbf{p}_A \tag{6.2}$$

$$\mathbf{F}_B\Delta t_B = \Delta\mathbf{p}_B \tag{6.3}$$

So as we learned "by hand," both, $\mathbf{F}_A = \mathbf{F}_B$ and $\Delta t_A = \Delta t_B$ so:

$$\Delta\mathbf{p}_A = \Delta\mathbf{p}_B. \tag{6.4}$$

That is, the change of momentum that A feels is the same as the change in momentum that B feels. Let's pretend that their collision is in one dimension and remember the convention for "change of" and that we'll use the subscript "0" to indicate the initial quantities:

$$\Delta p_A = p_A - p_{A,0}$$
$$\Delta p_B = p_B - p_{B,0}$$

Further, we'll presume that A was initially moving in the positive x direction so that its force and velocity are initially in that direction, and so from that designation, B is moving in all of the opposite directions. We then find:

$$\begin{aligned} \Delta p_A &= -\Delta p_B \\ p_A - p_{A,0} &= -(p_B - p_{B,0}) \qquad \text{rearrange these terms...} \\ p_{A,0} + p_{B,0} &= p_A + p_B \end{aligned} \tag{6.5}$$

Equation: Momentum conservation for two bodies.
$p_{A,0} + p_{B,0} = p_A + p_B$

Equation 6.5 is a really important relation! It says in words that the total momentum (of A plus B) at the beginning, before the collision, is equal to the total momentum of both objects at the end. This is the statement of our first serious "Conservation Law." In this case, Momentum Conservation.

> **Wait.** *You used the L word.*
>
> **Glad you asked.** *You're right. These conservation rules are so important—Nature is really telling us something when we discover such a rule–that it's one of those times when history and custom forces us using the Law designation.*

Definition: Momentum Conservation.
The total momentum at the beginning of any process is equal to the total momentum at the end of that process.

You can take it to the bank:

Momentum is always conserved. *Key Concept 13*

6.5 Examples of Collisions

Armed with the idea of momentum conservation as summarized by Eq. 6.5 we will analyze some example collisions. In EPP this is an essential part of our story since in order to peer inside of our elementary particles we have to probe them by shooting other elementary particles into them.

For simplicity's sake, our collisions will be between particles that have no physical size—"point particles"—and do not deform upon impact.

Wait. *There you go again. You've decided to describe an unrealistic circumstance. How is that helpful?*

Glad you asked. *As is usually the case, when we build a model we do so in a way that emphasizes the dominant physics at the expense of less significant, but realistic effects, that would get in the way of a simple description. In this case it's also a pretty good approximation for many real situations. Superballs are very resilient and behave like almost perfect objects. Also if we shot one electron toward another election, the electrostatic repulsion is responsible for the recoil and it's almost perfect. So this is a pretty good model for us!*

Some of the discrimination among them involves energy, which we'll cover in the next chapter, but I think this will make sense even in an everyday sense.

We'll consider four different examples of collisions in one dimension, and one in two dimensions. I'll do this simple one in this chapter, and the other four in the Diagrammatica 8 chapter that follows. I'll label them like a little formula, or chemical reaction. Here they are:

1. $\boldsymbol{A+B \to A+B}$ An elastic collision of two equal mass objects called A and B.
2. $\boldsymbol{A+b \to A+b}$ The elastic collision of two objects of different masses (A more massive than $\boldsymbol{b}$, a slightly different version of the above).
3. $\boldsymbol{A \to B+C}$ The decay of a large object, A into two different smaller objects, B and $\boldsymbol{C}$.
4. $\boldsymbol{A+B \to C}$ The collision of two different objects, sticking together to make a bigger, third object, $\boldsymbol{C}$.
5. $\boldsymbol{A+A \to b+c}$ The head-on collision of two identical objects that produce two other objects. (This we'll consider after we've been through some quantum mechanics.)

Definition: **Event.**
In a physical process...and event is when "something happens"! The makeup of the systems before and after an event may be very different.

Let's also establish some terminology, some of which is repurposing regular words into specialized physics terms:

Definition: **Initial State, Final State.**
The configuration of a system before a collision (or decay) and after a collision (or decay).

- I'll use the words *collision* and *scatter* interchangeably.
- We'll often call the whole scattering process—objects approaching, colliding, and separating—an *event*.

- The *state* of an object is its particular values of position and momentum. These six quantities are enough to completely characterize an object's motion.
- Let's treat the moment of collision as a special time in the event serving to separate the *initial state* (before) from the *final state* (after). So an event consists of the initial state, the collision, and the final state.
- For our purposes, we'll reserve the phrase *elastic collision* to mean that the two objects that scatter at the beginning are the same two objects after they collide. They have different states, but they are still the same objects in an elastic collision.
- A *decay* is an atomic or nuclear physics idea, but it has everyday analogues. A firecracker that explodes into a few pieces is like the "decay" of a firecracker. My analogy below is more imaginative!

Notice that in Fig. 6.4 and 6.5 I separated out these three states of the event. The initial state is in each (a) and the final state is (c). The collision itself is the intermediate, (b). All three together constitute the story of an event.

And, an aside on units: The units of momentum are associated with the units of $p = mv$, so mass times velocity. So a kilogram-meter-per-second would be a perfectly good unit for a momentum. In the English system, so would slug-ft-per-second, which is fun to say, but it's not usually used.

You Do It 6.1. kgm/s-hour

Rather than kg m/s, what would a unit of momentum be if you read your speedometer (in Canada!) in kilometers per hour? Assume the mass is again in kg.

or copy the solution

But I recognize that the units are unfamiliar and could get in the way of grasping the important stuff, so let's just invent our own unit-less arbitrary measure. If I say that a particular momentum is "5," we'll interpret that to be in arbitrary "momentumunits." Let's look at a variety of different processes with this in mind and repurpose some words:

- **Beam**. Any of our objects that are moving toward a collision, we'll call a beam or beam particle.[6]
- **Target**. A beam is aimed at an object which we'll call it a "target." Just what is the target and what is the beam is a matter of your point of view, but it will be clear in our contexts. You'll see.

A particular collision is when the target is sitting still and the beam hits it—like our billiards or ice hockey example. This situation is called a Fixed Target Collision.
Let's play pool.

Definition: Beam.
When a collision happens between two objects and one of them is aimed at the other, it's called the Beam or the Projectile.

Definition: Target.
When a collision happens between two objects and one of them is struck by the beam, it's called the Target.

[6] We'll use the ideal circumstance of our colliding objects having no extent: that is, everything will be a point-sized "particle." Once an object has a finite size, then hitting on an edge will start rotations of the colliding constituents and that complicates things beyond where we need to be.

6.6 Elastic Scattering: $A + B \to A + B$

In Fig. 6.5 we had the "simplest collision of all" in which two identical things collide and bounce off from one another:

$$A + B \to A + B \tag{6.6}$$

Let's analyze this event in detail, and then...one more time after this! Remember that B is sitting still, minding its own business, when it's struck by A, so the target is B and the beam is A.

Definition: Fixed Target Collision.
A collision in which the beam strikes a target which is at rest.

6.6.1 Two Body Scattering in Everyday Life

Sports are the easiest places to imagine elastic scattering in everyday life. For example, a bat hitting a pitched baseball is such a collision. Here we would probably identify the target as the bat and the beam as the ball. In any case, unless the bat splinters, the initial state is a bat and a ball and so is the final state. Another such collision is a golf club striking a golf ball. This would be a fixed target scattering event since the golf ball is sitting still.[7] Similarly you could think of a football kick-off; the iconic and rarely performed, full sliding mug of beer into your hand at the end of the bar; and of course a car accident in which the cars bounce off from one another (don't couple together).

[7] A much harder—and more fun to watch—version of this game would be one in which the ball was in motion, I guess.

In particle physics, many experiments are performed with a stationary target, like a liquid hydrogen container—which is then a bucket of protons—and a beam that could be any from among the zoo of particles. Currently at the Fermi National Accelerator Laboratory, beams of neutrinos are emerging from such targets.

6.6.2 A Particular Elastic Scattering Example: #1

Let's stick with the venerable pool ball collision example. Ball B is sitting still and the cue ball A strikes it directly in its center so that there is no sideways motion after the collision. Can't get any simpler.

Pencil 6.2.

In our fake "momentumunits" let's invent an example and follow it through:
Here's what we know in this example:

- The initial momentum of the cue ball is 12, $p_0(A) = 12$.[8]
- The initial speed of the B ball is zero, $v_0(B) = 0$.
- The mass of each ball is 6, $m_A = m_B \equiv m = 6$.
- The velocity of the outgoing B ball is 2, $v(B) = 2$

[8] Remember the subscript "0" means "initial."

Questions:

1. What is the initial momentum of the B ball?[9]
2. What is the *total* momentum of the entire initial state?
3. What is the initial speed of A, $v_0(A)$?
4. Using your experience, what is the final momentum of B, $p(B)$?
5. Using your experience, what is the final velocity of b, $v(B)$?

[9] A "who's buried in Grant's Tomb question.

6.6.3 Momentum Conservation

Before looking at momentum conservation, we can deal with the first three questions easily:

1. If the B ball is stationary, then its velocity is 0 and so its momentum is 0, $p_0(B) = 0$.
2. The total momentum of the intial state is the sum of all of the individual momenta of any objects in the initial state. So in this case, $p_0 = p_0(A) + p_0(B) = 12 + 0 = 12$.
3. The momentum is $p = mv$. Since the mass of A is $m_A = 6$ and the momentum is $p_0(A) = 12$, then $v_0(A) = 12/6 = 2$.

Now let's conserve momentum and solve the event. Our particular situation in symbols is:

$$A + B = A + B$$

$$\text{cue ball} + \text{stationary ball, B} \rightarrow \text{now, stationary cue ball} + \text{now, moving ball B}$$

Let's turn this into a momentum conservation equation:

$$\vec{p}_0(\mathrm{A}) + \vec{p}_0(\mathrm{B}) = \vec{p}(\mathrm{A}) + \vec{p}(\mathrm{B}) \tag{6.7}$$

Since we are dealing in only one dimension, we can stop using the vector notation and let the algebraic sign indicate direction (+ to the right and − to the left). So we have the simpler:

$$p_0(\mathrm{A}) + p_0(\mathrm{B}) = p(\mathrm{A}) + p(\mathrm{B}) \tag{6.8}$$

From playing pool, you know what happens, but let's do some momentum-accounting by filling in tables like Table 6.1.

	Before (initial state)	After (final state)
A	12	a
B	0	b
total	12	c

Table 6.1: Momenta in arbitrary units for the billiard ball collision with some blanks to fill in.

In the before column of our momentum-accounting in Table 6.1 we've listed what we know. In the after column, we know that momentum conservation insures that c has got to be the same as the initial state's total, or $c = 12$. Our experience tells us that when we strike B with the A, that the A suddenly stops dead and B jumps forward. Pool balls are especially rigid and, apart from the effects of rolling, their collisions are pretty good examples of elastic scattering. So our *experience* would tell us to put in $a = 0$. So, here's the big question (drumroll): What is $b =$? Of course, it's 12 from momentum conservation and that's the answer to question #4.

The last question asks about the speed of B after the scattering event. Since the mass is 6, $m_B = 6$ and the momentum is $p(B) = 12$, then we can see easily that $v(B) = 12/6 = 2$ and B goes scooting away with the same speed as A had before the collision. So Table 6.2 completes our understanding of this collision, using experience as our guide.

	Before	After
A	(6)(2)	0
B	0	(6)(2)
total	12	12

Table 6.2: Momenta in arbitrary units for the billiard ball collision, but indicating the masses and the velocities in their own arbitrary units.

But wait. There's actually a problem.

We used our experience to determine the final state. But what if we let Newton and Huygens explain this result? Put on your seatbelt.

Since our two balls are identical, their masses are the same and we'll use the single m for that common mass:

$$m_A = m_B \equiv m.$$

Momentum conservation says:

$$\begin{aligned} p_0(A) + p_0(B) &= p(A) + p(B) \\ mv_0(A) + mv_0(B) &= mv(A) + mv(B) \end{aligned}$$

where I've put in for the common definition of mass and unique velocity for each ball.

Since m is the same everywhere, we can cancel them all on the left and right sides. And, for this particular example, B was sitting still at the beginning. . . so. . . what's its velocity at the end?

Of course, we know that B is stationary, $v_0(B) = 0$. So, we get:

$$\begin{aligned} v_0(A) &= v(A) + v(B). \qquad (6.9) \\ v(B) &= v_0(A) - v(A) \end{aligned}$$

Equation 6.9 is Descartes' idea again: the total "motion" of the first object at the beginning is not lost but shared between the motions of all of the objects after the collision. But we're not happy. No, not at all.

Equation 6.9 is a simple, but terrible result! The beam object should stop dead so that $v(B) = v_0(A)$ but that's not what results from momentum conservation alone. Equation 6.9 doesn't *exclude* that result, but our balls don't just *sometimes* behave the normal way, **they always do**! Yet using Newton's and Huygens' mechanics the result we got is wishy-washy. Equation 6.9 allows the final speeds to be anything for either ball as long as they add up to $v_0(A)$. We expected to see specifically $v(B) = v_0(A)$, but we didn't get that.

Wait. *You mean that Newton and Huygens were wrong?*

Glad you asked. *Looks that way, but it's not quite that bad. We need more physics.*

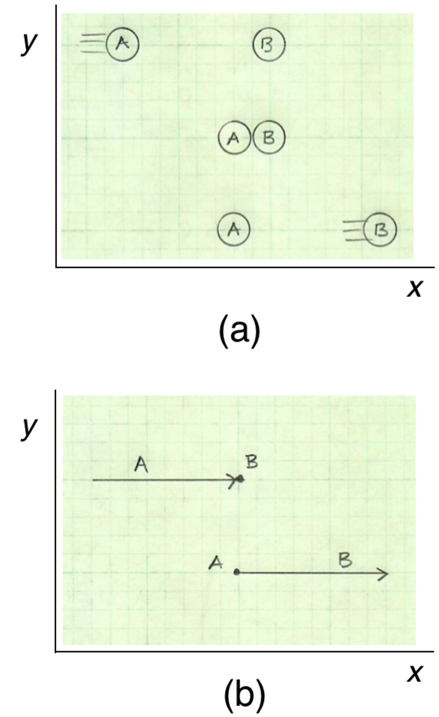
We need to introduce the big subject of kinetic energy which is the subject of Chapter 7.

6.7 Diagrams. Lots of Diagrams

I want us to become familiar with three kinds of diagrams for collisions: space diagrams, spacetime diagrams, and momentum diagrams.

6.7.1 Space Diagram

I think that the Space Diagram comes to mind easiest. Figure 6.8 is an abstraction of the collision in (a) looking down on the table. The bottom figure (b) shows the trajectories of the final state balls. Just like a map...tracing the paths. The dot indicates that a ball is stationary. The coordinates are space distances in both dimensions and time is implied, from the earliest (top) to the latest (bottom) as if snapshots were taken from above the table.



6.7.2 Spacetime Diagram

Figure 6.9 is the Spacetime Diagram for this collision. First, since time is one of the axes, "before" and "after" come for free on one drawing. Second, since this collision happens in one dimension, the vertical x axis represents all of the action. The B ball is just sitting still in space at position x_0 but it's moving in time. Finally, the collision happens at a particular time that's indicated to be t_C. So B's spacetime representation is a horizontal line at x_0 and extending from before t_C until exactly $t = t_C$.

Meanwhile, A is moving with a positive velocity (the x distance is increasing in time in a positive sense) and so its speed is represented as a positively sloped line in spacetime. Until t_C. Then A stops and B continues with the same velocity that A enjoyed before it collided with B.

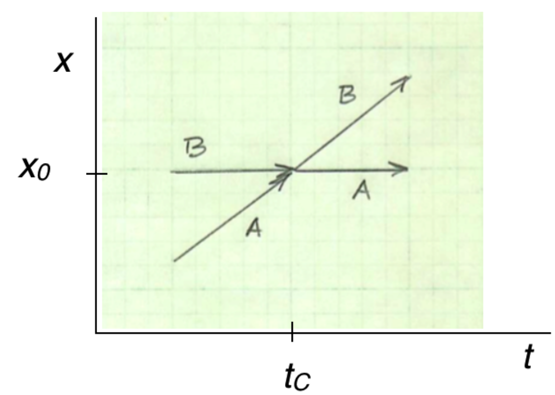


Figure 6.9: collisionspacetime

Momentum Diagram

It's also useful to include another diagram...one that represents the collision in a "mathematical space" that's sort of regular space but an overlay on regular space. The action happens in space, but what's drawn are vectors that represent momenta. Since a momentum vector points in the space (x, y, z) direction of the velocity, we can do this. But now the *length* of a momentum vector will be the value of the mass times

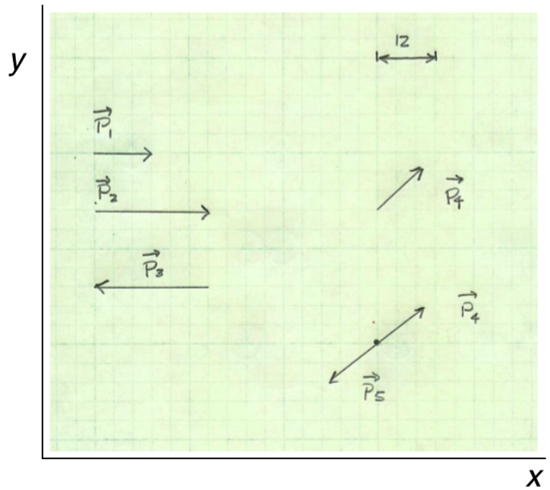


Figure 6.10: momentumcollection

the velocity. Figure 6.10 is a collection of momentum vectors using our fake units. The "space" is regular coordinate space, but the length of the vectors is in momentum units. The key at the top shows how long a momentum vector of 12 would be.

Momentum p_1 is a momentum pointing to the right (positive x direction) and it has value (length) of 12, according to the key. Notice that none of the momenta in the collection are identical. p_2 and p_3 both have lengths of 24, but different directions, so they are different momenta. p_4 is also a length of 12, but points in a direction that's a little of x and a little of y. Just below it is another vector, p_5 which points in the opposite direction from p_4. So we'd say that these momenta are **balanced** and that $p_4 = -p_5$.

In fact, this pair of momentum vectors is precisely what you would expect to see on a momentum diagram for a conserved momentum situation!

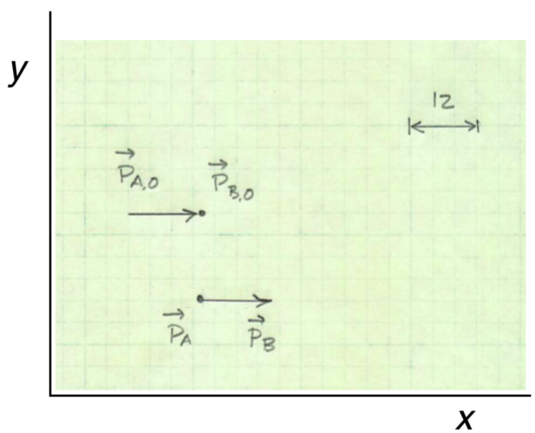


Figure 6.11: simplestonemomentum

So for our simplest collision, we can refer back to the development of the table and draw our momenta to scale. Figure **??** is the momentum diagram for this collision. I've explicitly labeled the momenta as $\vec{p}_{A,0}$, etc. But in order to simplify notation, in the future I might just call such a vector $\vec{A}_0$ or even A_0 if the direction is trivial.

Let's summarize all three diagrams in Fig. 6.12. This Collection O'Diagrams...is repeated over and over in the Diagrammatica 8 chapter where all of the collisions we might see are described.

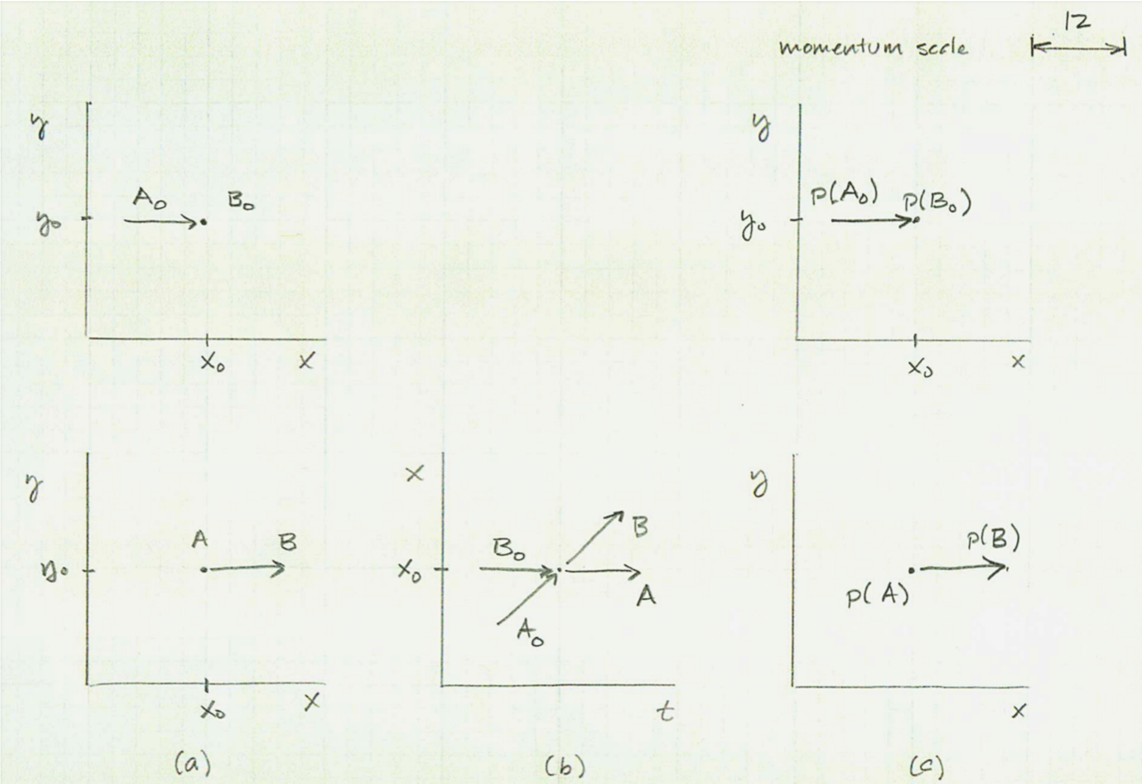


Figure 6.12: One one figure, all three diagrams are shown for our "most simple" collision. The Space Diagram on the left, the Space-time Diagram in the middle, and the Momentum Diagram on the right. In the next Diagrammatica chapter, these diagrams are all developed and cataloged for the various collisions we'll consider.

Box 6.1 Three Important Kinds of Diagrams Summary

We now have three kinds of diagrams that are useful in thinking about any reaction (like a decay) or collision. In fact—any physical process that takes place in time. Two of these we saw in Chapter 3 and Diagrammatica #1, Chapter 4, and momentum conservation motivates the third.

Space Diagrams. This is the map picture, where in a space of distances in both axes (thinking in two dimensions), a trajectory is traced out as a contour. Time is implied as the trajectory takes you from a starting to an ending place with each point in between having an implied time stamp. For our earlier examples, we had only one object that moved...our car in Michigan, the bicycle, and so on. In particle physics, we have multiple objects and often one thing turning into another thing or things. So we should have a space trajectory for each object and it can get messy. For simplicity, we'll represent separate Space Diagrams for the before picture and the after picture...the initial state and the final state...where something will have happened in between. You'll see. For the pool balls example, the Space Diagram is shown in Fig. 6.12a.

Spacetime Diagrams. We've seen examples of Spacetime Diagrams—Feynman Diagrams, we'll call them—and they will figure prominently as we go forward. Here a representative space dimension is pictured along the vertical axis (the circumstances govern which one is chosen) and the horizontal axis represents time. With this arrangement, the slope of any Spacetime trajectory is its velocity, space divided by time. The Spacetime Diagram for the pool balls is in Fig. 6.12b. In fact, we've seen our first interesting Feynman Diagram and one that, with a little quantum mechanical tweaking, is highlighted on Richard Feynman's famous van show in Fig. 6.13. There's a story there.

Momentum Diagrams. Momentum Diagrams make the conservation of momentum visually apparent. In fact, they can be used to make predictions, as we'll see below. Time figures into Momentum Diagrams in the same way as our Space Diagrams by drawing one diagram for "before" and another diagram for "after." The point of them is that whatever the total vector sum of momentum is in the *before* picture has to be the vector sum in the *after* picture. The Momentum Diagram for the pool balls is in Fig. 6.12c.

Figure 6.12 shows the three diagrams side by side for emphasis. I need for you to be comfortable with all three.

Figure 6.13: I'll talk a lot more about Richard Feynman later, but he was famous for many things including his "out there" personality. This is a photograph of his van that I took a couple of years ago, which he'd covered in. . . Feynman Diagrams. The one right over the California license plate corresponds exactly to our "simplest" collision, when quantum effects are included. Stay tuned!

6.7.3 All of our collisions

We will consider a small number of different collisions and we can categorize them all by their diagrams. Our Second Diagrammatica Chapter 8, looks at each one and presents the diagrams for each and we'll refer back to it often.

6.8 Two and More Dimensions

We will not explicitly calculate in more than one dimension in QS&BB, but the sorts of collisions that you're most familiar with happen in more than one space dimension. Looking down on the pool table?

The balls move in x and y. In Demolition Derby, the cars scatter all over the infield. A pitch that's precisely horizontal as it passes home plate, is lofted into the air in collision with the bat. We could calculate the consequences of all of these kinds of two and three dimensional collisions, but we won't. We'll draw pictures.

In our previous considerations of apples and billiard balls colliding, we assumed that they had no size—point-like. But striking a billiard ball precisely on its center line with a cue ball is tricky. More likely is that they would strike just slightly off-center like in Fig. 6.14. In that case, if we define our x axis along the direction of the beam-ball, the target and the beam will both scatter into the y direction. Momentum conservation helps to determine the outcome.

The most general statement of the momentum conservation rule[10] is:

$$\mathbf{p}_{A,0} + \mathbf{p}_{B,0} = \mathbf{p}_A + \mathbf{p}_B \tag{6.10}$$

Equation 6.10 is really a symbolic statement—a *mathematical paragraph*, if you will—and not an equation that you can actually solve. Embedded in it are two (or three, if three dimensional space) "real" equations that you can actually manipulate...one for the momentum along the x axis and another for momentum along the y axis. Momentum conservation has to hold separately in each of them but we'll not actually solve the equations themselves. Rather, we can construct the Momentum Diagram and draw conclusions without doing any algebra. Let's put together a momentum diagram for this situation using our arbitrary units with the following parameters to start with. Pool balls? That's for children. We'll throw bowling balls at one another. I'm going to use the language of "beam" and "target" here, as that's more to our EPP liking. (In our previous collisions, A would have been the beam and B would have been the target.) So, now B will mean beam and T will mean target. Sorry.

- mass of the beam ball, B: 7 kg
- mass of the target ball, T: 7 kg
- speed of the beam ball: v(B) = 10 m/s

The coordinate system that we set up in Fig. 6.15a indicates the directions of motion in space, and so also, the directions of the momenta. In our previous Momentum Diagrams we were casual about the scale in our fake momentumunits, but let's do this one more precisely. We'll set the scale of the diagram as "1 inch equals 50 kg-m/s" and you can see that as the bottom line in Fig. 6.15. Now our ruler is calibrated in momentum units.

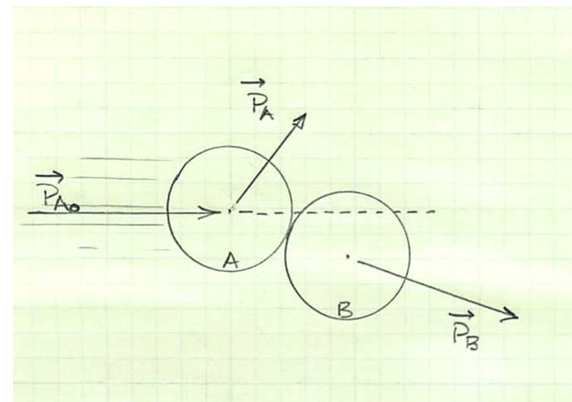


Figure 6.14: offcenter

[10] Remember? The total momentum of a system at the beginning of a collision is going to equal the total momentum of that system at the end

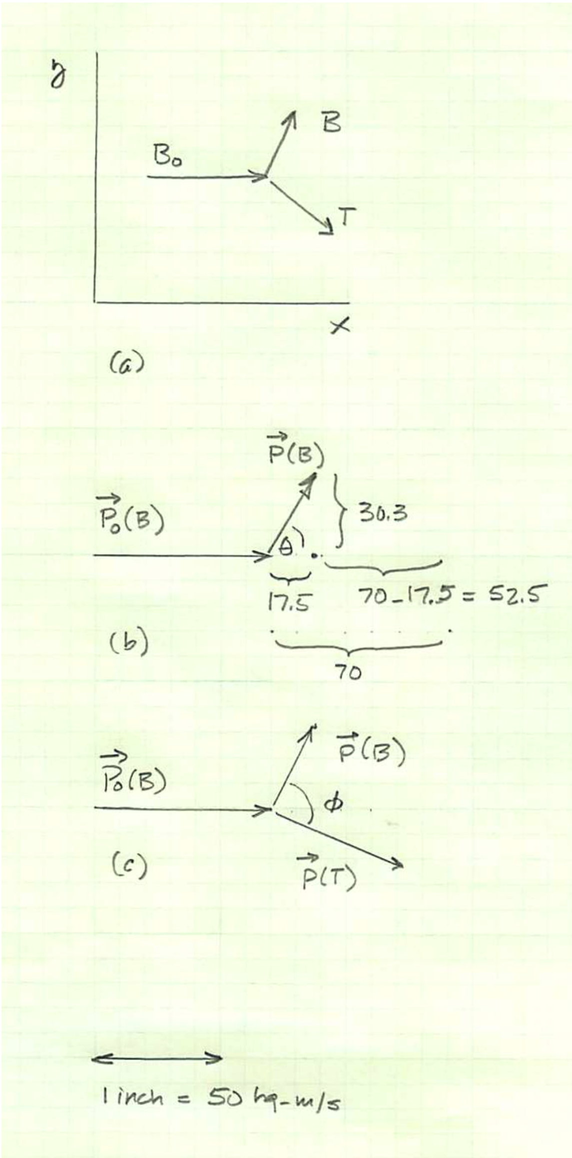


Figure 6.15: twobodyp

So, again, what do we know? We know the initial momentum of the beam. (Notice I've decorated the momentum symbol twice in the subscript. The 0 as the first of the pair indicates that this is an initial value like usual. The second, x means that this is in the x direction.)

$$p_{0,x}(\mathrm{B}) = m(B)v(\mathrm{B}) = (7)(10) = 70 \text{ kg-m/s}.$$

We also know the initial momentum of the target ball:

$$p_0(\mathrm{T}) = 0 \text{ kg-m/s}.$$

Let's suppose that the degree to which the centers are off is such that the beam ball bounces quite a bit off the target and shoots out in the $+y$ direction at an angle of 60° relative to the horizontal. Without doing the trigonometry, I'll just tell you that this means that the outgoing horizontal momentum of the beam is

$$p_x(\mathrm{B}) = 17.5 \text{ kg-m/s}$$

and that the vertical component is

$$p_y(\mathrm{B}) = 30.3 \text{ kg-m/s}.$$

This is all the information that we need to predict the motion of the target ball! Had Huygens and Newton been bowling buddies, they could have worked this out.

In Fig. 6.15b I've shown some distances and drawn in a vector momentum. The distance 17.5 is the x component of the final B momentum and the distance 30.3 is the y momentum. The arrow is the combination of the two components.

The target ball was just sitting still, minding its own business when it was hit by the beam ball. We can precisely determine what happens to it from the information we've gathered. We simply conserve momentum vertically and horizontally...by "eye." That is, by drawing in the **T** momentum vector according to momentum conservation.

- All of the x-directed momentum of the initial state system (all do to the beam) was 70 kg-m/s and all directed along x. So the sum of all momenta in the x direction of the final state system has to also add up to 70 kg-m/s.
- All of the y-directed momentum of the initial state system? Zero. There was no motion in the y direction. So the sum of the vertical components of final state momenta have to balance exactly, to sum to zero.

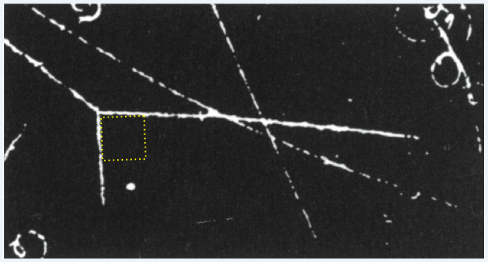
I've indicated the initial momentum of 70 in the final state as a bracketed length in Fig. 6.15b. That's so that I can use momentum conservation to find the x momentum of the target, it's just the amount left after the 17.5 from the beam which I've shown as 52.5. The vertical component of the target's momentum after the collisions? That's easy. It just balances the 30.3 value from the scattered beam.

So piece these together graphically, using the scale, I can sketch in the recoiling target's momentum vector and it's in Fig. 6.15c. By eye, the easiest thing to see is that the up and down momenta of the two final objects balance. You'd need your ruler to see that the sums of the horizontal components of each of the final state objects add to about the total of the original beam's momentum vector. We'll just do this by eye when we need to, but you get the idea!

Example 6.1

Resolving Elastic Scattering of the Target

Question:



Referring to our bowling ball example in Fig. 6.15:

a) What are the x and y components of the target ball's momentum?

b) Draw the total momentum vector of the target ball along with that of the beam.

Solution:

a) The total horizontal momentum has to be 70 kg-m/s, and we know that 17.5 of it is taken up by the beam ball's contribution to the final state motion. So the mome ntum of the target ball—in the x direction—must be $70-17.5=52.5$ kg-m/s. Now the initial and final state motions are balanced, horizontally.

The vertical component of momentum has to balance to zero and since the target's contribution is 30.3 kg-m/s, "up" $(+y)$, then the target's contribution must be -30.3 kg-m/s, "down" $(-y)$.

BTW...a fun fact. The angle that the target is ejected towards is a special one for this situation...in fact, it's 30° so that the sum of the two outgoing angles is $60^\circ+30^\circ=90^\circ$. This is always true when the target and the beam have the same masses and the collision is elastic. The outgoing particles themselves are separated by 90°. Look at the picture in Figure 6.1 which shows a bubble chamber photograph of a proton that enters the picture from the upper left and hits a proton in the liquid and scatters it in one direction while itself recoiling in another. The angle between the outgoing protons can be seen to be about 90°.

b) The whole thing is drawn on Fig. 6.15c.

What about the Feynman diagram of the whole "interaction"? Well, that's complicated. Let's use the x direction as our vertical space coordinate and of course time is always our horizontal coordinate. The above example shows us the relative values of the x momenta, and since the masses are the same, also the relative speeds. The relative speeds of the balls are then as follows: $v_{0,x}(\mathrm{B})>v(\mathrm{T})>v(\mathrm{B})$ so if we draw this diagram for just the x space coordinate versus time, it would look something like the sketch in Fig. 6.16.

The examples in the next Diagrammatica chapter, Section 8.4 will illustrate some two dimensional scattering, along with some real-life collisions.

6.8.1 Two Body Scattering in Two Dimensions

Finally, for EPP a relevant situation is one in which two balls, A and B, are head-on but also have a finite size and their centers are slightly shifted so that the collision starts off in one dimension, but the scattering is into two dimensions. Figure 6.17 shows this. Here the X marks where the collision happens.

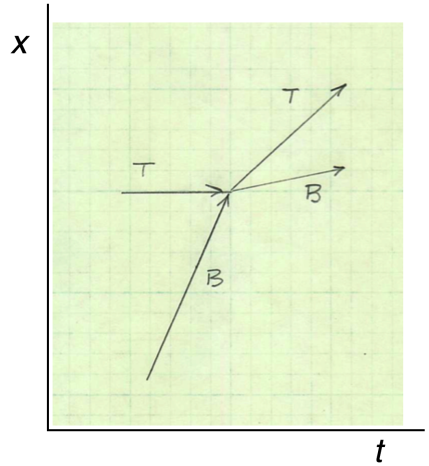


Figure 6.16: twoballsspacetime

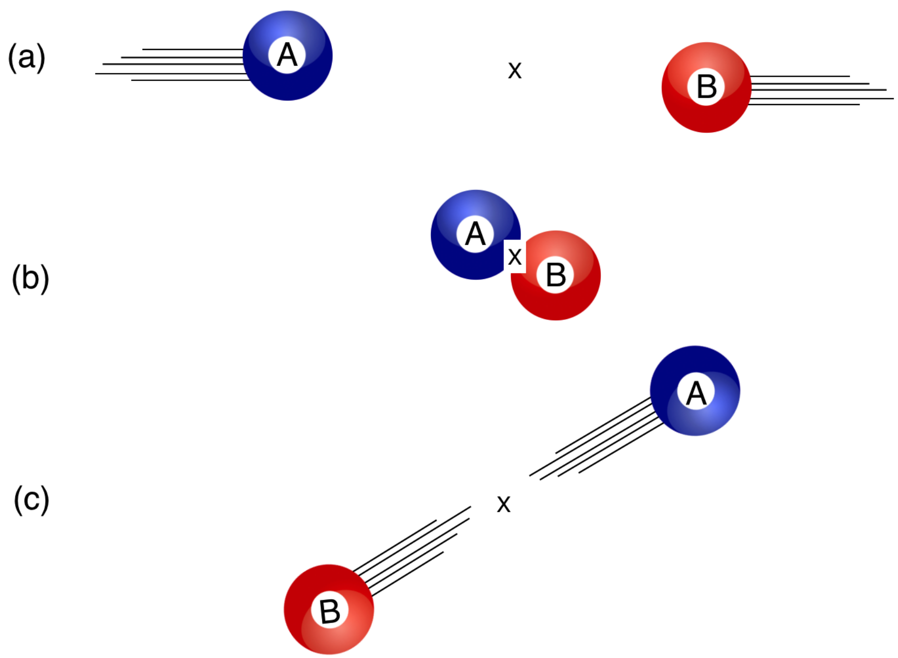


Figure 6.17: A more realistic collision. Again, looking down from the top two balls strike one another, but just off center so that they scatter in directions different from the original, oppositely converging directions. (a) and (c) are the initial and final states and (b) is the point of collision, where they bang together and momentarily stop. The X just indicates the point in space where they collide.

You Do It 6.2. Beams

or copy the solution

Refer to Fig. 6.17 and draw (a) the Space Diagram and, if we assume that the balls are both the same mass and each have the same initial speeds, draw (b) the momentum diagram.

6.8.2 Collisions At the Large Hadron Collider

Now let's look at some actual collisions in our experiment at CERN. Figure 6.18 shows a side view-slice of our "ATLAS" detector at the Large Hadron Collider at the CERN laboratory in Geneva, Switzerland. We'll learn what the various colors symbolize but we know enough now to understand what's happening.

A collision has happened and is one in which two protons collide head-on right in the center—one beam from the right and the other from the left (like the billiard-balls above and exactly like Fig. 6.17)—and produce, in this case, two electrons that emerge and leave their traces in our detector as the two

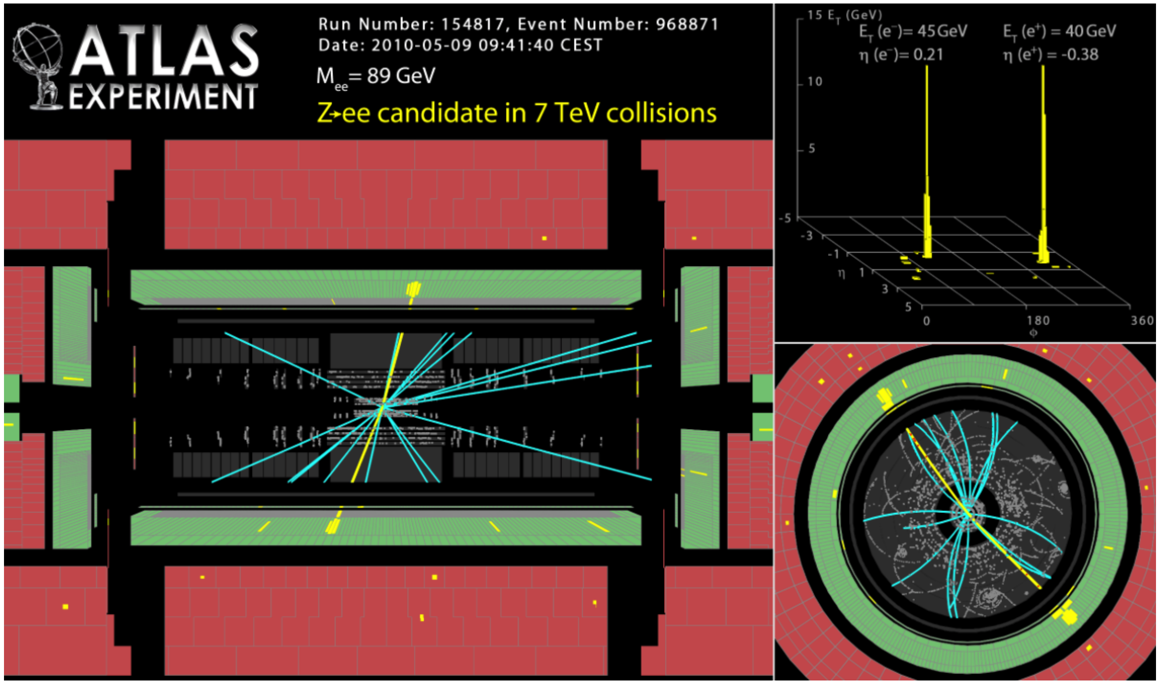


Figure 6.18: This is a computer reconstruction of the consequences of colliding two protons with one another in the ATLAS detector at the Large Hadron Collider at CERN. The left-hand view is a vertical slice through the detector with up being toward the surface (the device is 300 feet underground). The two beams of (identical) protons enter from the left and the right and collide in the center.

diagonal blotches of yellow color. You can think of the yellow lines and little blobs as the momentum vectors for the outgoing particles from the collision.

Momentum is being conserved in this simple reaction as you can almost see by eye. The initial state momentum in all directions is zero, since both beam particles are identical protons, so the momenta of the two electrons each have to be conserved in the horizontal and vertical directions. The amount of color in the "blobs" is a good measure of the magnitude of the momentum in each electron, and again, by eye you can just about see that they are equal and opposite.

Figure 6.19 is a completely different situation! Again, the collision is two protons, head-on. Again, in the final state there is an electron...but it looks like momentum is not conserved since there appears to be nothing emerging from the collision on the other side! That's our clue that a particularly elusive particle called a neutrino was produced with the electron, and we would even say, "Look, there's a neutrino in that event."

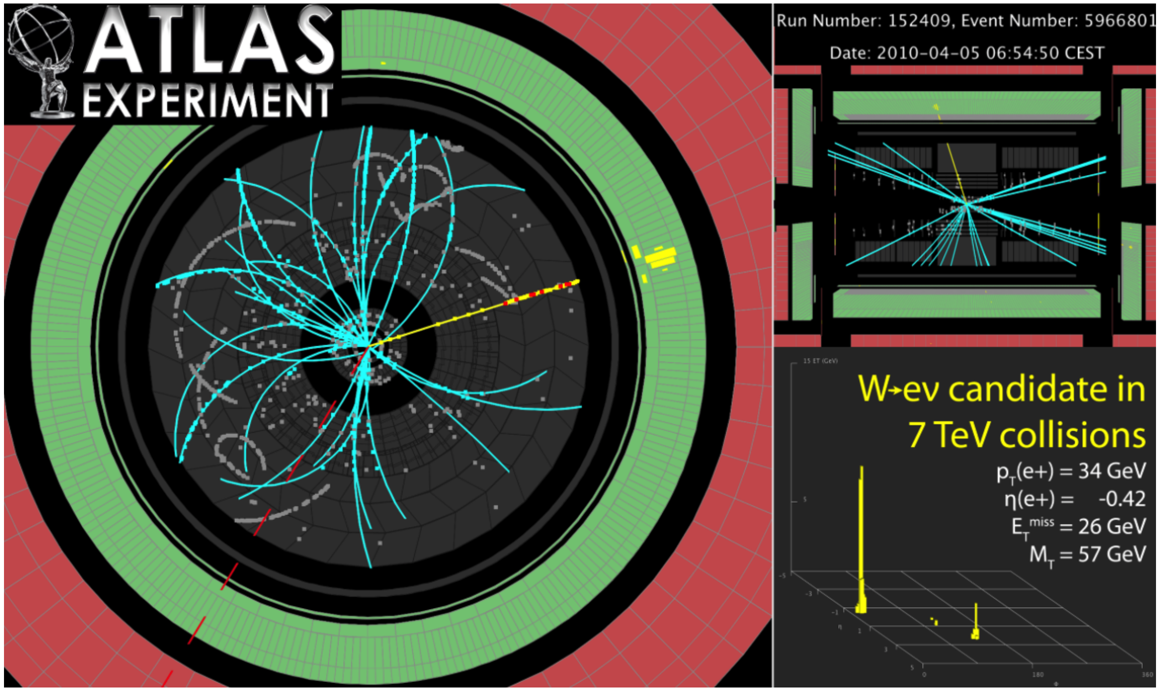


Figure 6.19: This is another event from the ATLAS detector, but now notice that the final state has only one yellow blob—one electron. Notice that there is nothing on the other side! Is momentum conservation violated?

Wait. *You mean that you observe this neutrino-particle by... seeing nothing?*

Glad you asked. *Yup. We believe in momentum conservation so strongly, that we are certain that some particle emerged that left no trace in the detector. Neutrinos are such a particle that interact with matter so rarely, that they essentially never show their presence in our detectors except by being the cause of an apparent momentum imbalance. That's our clue. Notice that this is really not the detection of a neutrino on an event-by-event basis. No, there are other hypothetical particles that might also leave no trace after they're produced and we are searching for exactly those kinds of particles.*

Wait. *Sorry to bother you again. But how do you know the difference then?*

Glad you asked. *No problem. We have to make such discoveries on a statistical basis. We know the rules for producing the "normal physics" with neutrinos and if there are other hypothetical particles that are produced, then we'd have to see them behave differently in many, many collisions. A statistically significant anomalous behavior.*

Have you got enough energy to learn about energy?

Chapter 7

Energy

A Long Time Coming

James Prescott Joule, Photogravure after G. Patten

James Prescott Joule, 1818-1889

"…wherever mechanical force is expended, an exact equivalent of heat is always obtained." *Joule, August (1843)*

The University of Manchester in that industrial city has been the home of to-be illustrious physicists as well as already-in-the-textbooks physicists for more than 150 years. Ironically, the Manchester scientist credited with one of the most fundamental statements about the word had nothing to do with the university. He made beer. James Prescott Joule was the son of a brewer who joined the management of the family business in his early 20's where he launched intensive research into how to increase the efficiency of or replace its large-scale steam engines. This led to a lifetime of largely private research into the nature of energy.

Figure 7.1: brewery

7.1 Goals of this chapter:

- Understand:
  - How to calculate kinetic energies of moving objects.
  - How to calculate potential energies of objects.
  - How to use the conservation of energy to calculate speeds. parameters
- Appreciate:
  - The importance of the conservation of momentum and energy.
- Be familiar with:
  - The importance of James Joule's work.
  - The importance of Emmy Noether's work.

7.2 A Little Bit of Joule

James Prescott Joule grew up in a wealthy family and was educated by private tutors and by the age of 16, the illustrious John Dalton,[1] also a resident of Manchester. Joule had an adolescent fascination with electricity, probably influenced by the famous work of Michael Faraday in London. When he and his brother were not (literally!) shocking their family and the household staff, James was beginning to conduct research on what causes heat. Motors were beginning to be conceived of and he built several and compared the amount of coal in a steam engine required to perform a fixed mechanical task with the amount of zinc to power a battery-driven motor towards that same task. All the better to figure out what was the best technology for the brewery. Coal won and they didn't adopt the new-fangled electric motor.

As a young man in the family business, Joule would go to the brewery by day and perform his management tasks, and then when he could find time, he would perform his private experiments in his homemade laboratory. From his 20's he carefully charted a course to unraveling three different phenomena, all of which caused objects to heat up, but none of which corresponded to the accepted picture of just what heat was supposed to be. He was suspicious of the commonly held theory that heat is a fluid, "caloric," that was neither created nor destroyed and moved (flowed) from a hot object to a cold one.

His first demonstration in 1841 was to show that when an electrical current flows through a wire, that it heats it. He could explain this by the heat being generated in the wire, and not having been transported from the source of the current. Caloric would have flowed to the point of heating. Today we call this Joule Heating and the formula for the amount of power associated with this heating is due to him: $P = I^2R$,

[1] John Dalton (1766–1844) is considered the father—or at least the favorite uncle—of chemistry. He worked out much of the picture of substances as made of atoms and that chemical compounds are made of atomic constituents. He lived in Manchester his entire adult life where he taught privately and at the university. As a Quaker, he was ineligible for education or employment at many British universities.

where R is the value of the resistance (which was a new idea when he was experimenting) and I is the current. This is the principle behind an electric stove or heater...and the villain to be defeated in the long-distance transmission of electrical power. His second demonstration was to show that when a gas is compressed, that the amount of force required translates directly into the temperature rise in the gas. It's the principle behind an internal combustion engine and the beginning of his notion of a "mechanical equivalent of heat" which led him to his next experiment, for which he's best remembered.

Heat and motion are both forms of energy which can be converted back and forth. Key Observation 4

If you mechanically stir a fluid, it gets warmer. Not a lot. But Joule had inherited from his tutor, Dalton, the idea that a gas was made of atoms (and developed his own theory of gases and the energy of molecules) and that making them move faster was to increase their temperature. He also applied this idea to water. He created a little system with paddles in a beaker of water that could be made to stir the water a specified amount because they were attached to a falling weight. The weight falls a given amount and the paddles reliably turn a specific number of rotations. Joule became skilled at making thermometers[2] and he found that a finite amount of stirring could raise the temperature of water by a single degree Centigrade. He reported this result to the British Association in 1845 and published a paper describing his results in the *Philosophical Magazine*.

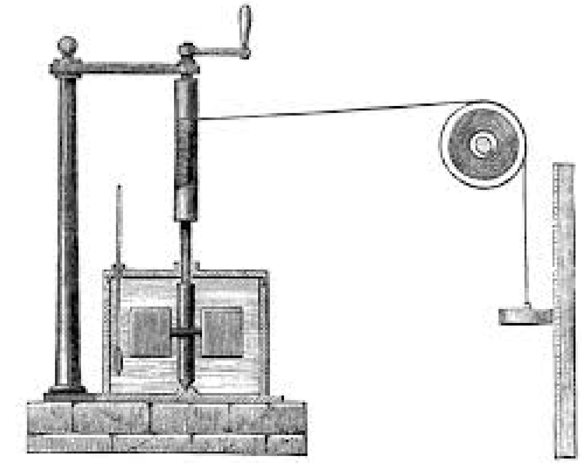


Figure 7.2: joulemechanical

[2] He once made a thermometer so precise that he could measure the temperature of moon-light. That is the temperature rise in air lit only by the moon.

He married Amelia Grimes in 1847 (who tragically died seven years later after they had two sons and a daughter). Their honeymoon was in Chamonix, France (near CERN, actually) where together they tried to measure the difference in temperature between water at the top of a waterfall and the bottom. You gotta love that as a scientist's honeymoon.

Joule was a little isolated while he did much of his work, but increasingly as a result of fortuitous speeches with just the right people in the audience, he became more and more well known and well regarded in Europe. Without any formal education, this recognition came slowly but eventually he was elected to Fellowship in the Royal Society in 1850 and received honorary degrees from Dublin, Oxford, and Glasgow. Finally, in 1872, he served as the President of the British Association. Not bad for a brewery lad.

William Thomson (later Lord Kelvin) wrote about his friendship with Joule and his surprise to discover that James was conducting experiments in waterfalls on his honeymoon."After that I had a long talk over the whole matter at one of the 'conversaziones' of the Association, and we became fast friends from thenceforward. However, he did not tell me he was to be married in a week or so; but about a fortnight later I was walking down from Chamounix to commence the tour of Mont Blanc, and whom should I meet walking up but Joule, with a long thermometer in his hand, and a carriage with a lady in it not far off. He told me he had been married since we had parted at Oxford! and he was going to try for elevation of temperature in waterfalls. We trysted to meet a few days later at Martigny, and look at the Cascade de Sallanches, to see if it might answer. We found it too much broken into spray. His young wife, as long as she lived, took complete interest in his scientific work, and both she and he showed me the greatest kindness during my visits to them in Manchester for our experiments on the thermal effects of fluid in motion, which we commenced a few years later."

Joule convinced everyone that heat and work (we'll see what the formal definition of work is below) are two sides of the same coin: *energy*. That "energy" can be transferred back and forth between heat and work is basically the First Law of Thermodynamics and the basis of the world's industrial economy and many of our household conveniences. It led to the notion of the conservation of energy and guides our thinking to this day.

Joule died in 1889 and is honored forever with his name used as the universal unit of energy: 1 Joule (J) is the equivalent of 1 kg-m^2/s^2. We pay our electricity bills according to Watts used, and its the only everyday metric unit in the U.S.: 1 Watt is 1 Joule per second.

7.3 Ability to Do Damage

Okay. "Ability to Do Damage" isn't a scientific phrase...but I'll bet you'll remember it better than our very specific use of a regular word: "work." If you want to do damage to something, you initiate some sort of contact with it and speed often figures into that process. Want to demolish something with a hammer? Gently pat it? or swing the hammer at high speed? Want to smash a teapot by dropping a rock onto it? Drop it from high up so it's moving really fast when it hits. So you need some speed to do damage. But mass figures in too: a hammer made out of balloons is not a damage-maker and neither is a pebble. So a question is: what's more important, mass or speed in inflicting damage?[3] Let's go back to High School and think about this.

[3] This comes up all the time with major league sluggers. Some will swear that in order to drive a baseball a heavy bat is better than a lighter one. Not true.

A regulation softball has a mass of about 0.22 kg while a regulation baseball has a mass of just about half of that, 0.145 kg. Now here's the question: A decent high school softball pitch is about 50 mph—faster than that, and you've got a college pitcher on your hands. But a 50 mph baseball is not so impressive, less than batting practice quality. Consider these two trade-offs, and think about being hit by each:

- Replace a baseball thrown of 50 mph with a softball of the same speed—a factor of 2 increase in mass, but same speed?
- Replace a baseball thrown at 50 mph with a baseball thrown at 100 mph—a factor of 2 increase in speed but the same mass?

Which replacement would do proportionally more damage? I'd take the first item any day.

Speed matters in this image more than mass, in fact it matters by a lot more. Since mass and velocity contribute to momentum in equal proportions, so this discussion of "damage" is referring to some other quality of motion. That additional quality we call Kinetic Energy. We'll use the symbol K to stand for it and in modern terms, it's written as

Definition: kinetic energy.
Kinetic Energy is the energy possessed by any object in motion.

$$K = \frac{1}{2}mv^2 \tag{7.1}$$

Equation: Kinetic Energy.
$K = \frac{1}{2}mv^2$

the ability to do damage is related to the square of speed and only linearly with mass.

The fancy way to speak about this is in terms of "work" which means something very specific in physics. Work is the product of force × the distance through which the force acts. This is similar to the way that

Impulse is the product of force × the time through which the force acts. Work is then equal to the change in kinetic energy, in the same way that Impulse is equal to the change in momentum.

So while the pitcher increases the momentum of the ball by applying a force to it through a full windup and follow-through (longer time), he also increases the kinetic energy of the ball by applying that force through a long distance. So a long-armed pitcher with a big arc has an advantage. The formal statement of this is:

$$\text{work} = W = F\Delta x = \Delta\left(\frac{1}{2}mv^2\right) \tag{7.2}$$

which looks a lot like

$$\text{impulse} = J = F\Delta t = \Delta mv. \tag{7.3}$$

The partnership between time and space is related to the partnership between energy and momentum, as we'll see a bit later.

7.3.1 Vis Viva

One of the remarkable achievements of Huygens, totally unanticipated by Newton, was the discovery of a second conserved quantity. In this, Huygens had a partner: Gottfried Leibnitz—Newton's bitter rival for the priority of the Calculus—had the same idea. They both found by experiment that if you add up all of the quantities: mv^2 for all of the objects in a collision that the total amount of that quantity before is equal to the total amount afterwards. . . *without regard to direction.* That is, since the velocity is squared these are not vector quantities, but scalar ones. Just numbers.

This was incorrectly given the name of "force" by Leibnitz, in particular the "Life Force" or *"vis viva."* Today (actually, about mid-18th century), a factor of 1/2 is added to create the quantity we call Kinetic Energy, KE= $\frac{1}{2}mv^2$. So, to summarize what's conserved in collisions, we separately conserve:

$$\mathbf{p}_{1,0} + \mathbf{p}_{2,0} = \mathbf{p}_1 + \mathbf{p}_2 \tag{7.4}$$

$$\frac{1}{2}m_1(v_{1,0})^2 + \frac{1}{2}m_2(v_{2,0})^2 = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2. \tag{7.5}$$

The first equation is the Conservation of Momentum, a vector equation and the second is the Conservation of Kinetic Energy.

Now, we can go back to the incomplete Example 6.9 where we were left hanging. Had we also added the Conservation of Kinetic Energy.

Definition: Kinetic Energy.
KE= $\frac{1}{2}mv^2$
Kinetic Energy is proportional to speed squared and mass.

Example 7.1 One Dimensional Collision. . . continued

Pencil 7.1.

Where we left off was Equation 6.9:

$$v_{1,0} = v_1 + v_2. \tag{7.6}$$

Now, let's include the Kinetic Energy relationship for this particular situation:

$$\begin{aligned} \frac{1}{2}m_1(v_{1,0})^2 &= \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2. \\ (v_{1,0})^2 &= v_1^2 + v_2^2, \end{aligned} \tag{7.7}$$

where in order to get the second line I canceled out the equal masses and the common factor of $\frac{1}{2}$.

Now we have two equations and two unknowns to solve, which can be done in a variety of ways (remember?). You always have to keep track of what you're looking for. Here, it's the final velocities. So, let's square the Eq. 7.6 and subtract it from the second one and you get the result:

$$0 = 2v_1 v_2. \tag{7.8}$$

Here are the few lines that lead to that simple conclusion:

$$v_{1,0} = v_1 + v_2$$
$$(v_{1,0})^2 = (v_1 + v_2)^2 = v_1^2 + v_2^2 + 2v_1 v_2$$
set equal to the RHS of Eq. 7.7
$$v_1^2 + v_2^2 = v_1^2 + v_2^2 + 2v_1 v_2$$
$$0 = 2v_1 v_2$$

So, either one or the other of the final velocities must be zero. One of these solutions doesn't make any physical sense. For example, if the target ball (2) is solid, then the target ball can't just fly right through it as if it were not there, so v_2 cannot be zero, it must be something else. That means, that $v_1 = 0$ and going back to Equation 7.6, we see that:

$$v_2 = v_{1,0},$$

which is what we expected.

We need **both** momentum and kinetic energy conservation to describe even the simplest of collisions!

7.4 Energy

The idea of Kinetic Energy was eventually appreciated as a part of a much broader concept. We use the term freely, but it's a subtle thing and the 17th, 18th, 19th and 20th centuries saw repeated recalibration of the idea. It was not until nearly the middle of the 1800s that heat was carefully studied by many, culminating when James Prescott Joule (1818-1889) carefully measured the amount of kinetic energy he put into a beaker of water by stirring it. He found that the water's temperature went up the same amount for the same input of energy—he suggested that heat was a equivalent to energy. Soon after Joule's death, it was decided internationally to honor Joule's memory by naming the basic unit of energy: 1 J = 1 N m after him.

Heat, then is a form of energy, adding to kinetic energy and potential energy as the classical trio of energy forms (nuclear, chemical, and elastic energies are additional kinds). Potential energy is just what the name implies…"the potential" for causing damage! Hold a barbell above your foot and let it go, it will change the shape of your foot when it lands, and maybe the floor as well. That suspended weight possess the *potential* for doing "Work," which is a technical term different from the everyday usage. If a force acts on an object through a distance x, then the work is defined as:

$$W = Fx.$$

7.4.1 Potential Energy

The subtle point about Work is that the force must have a part of its direction along the path through which it's acting. So, if I carry a heavy weight still, but walk across the room, I may be tired and think that I've worked hard, I've done no (technical) Work, since the direction I walked is perpendicular to the force that I exerted (up) in holding the weight. Work figures into the statement of an important theorem in mechanics, the Work-Energy Theorem: The change in kinetic energy in a collision is equal to the Work that's performed. In fact, the exchange of almost all sorts of energy involves doing Work.

For dropping things in a gravitational field, the Potential Energy is:

$$P = mgh \tag{7.9}$$

where h is the vertical distance above the point defined to be the zero value of potential energy. That's sensible since mg is the weight of the object, the force pulled on it by the Earth. So this too is a force times a distance, Wh. The typical example of potential energy at work (no pun intended...or is there?) is driving a nail into a block of wood by dropping a weight from some height as shown in Fig. 7.3. Potential energy is a funny concept and I'll have more to say about it when we talk about Einstein. But, it does have a slippery feature that's sometimes complicated to appreciate:

There is no *absolute* measure of potential energy. Only differences, before and after some change of configuration matter.

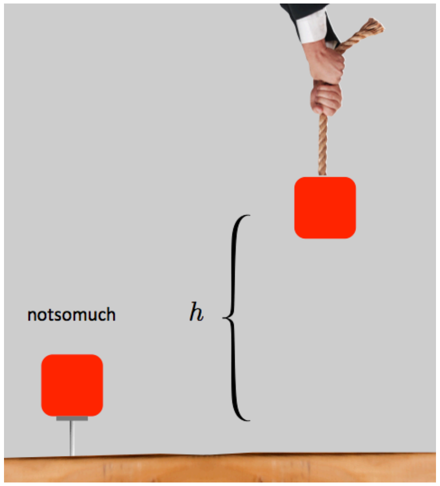


Figure 7.3: (left) Setting a block on a nail does not do much work against the fibers of the wood. (right) Dropping the block from a height onto a nail, drives it into the wood.

Definition: Potential Energy.
Potential Energy is possessed by an object by virtue of its "configuration"...height, distance away from a force center, located in a compressed or expanded spring, etc.

Equation: Gravitational Potential Energy, near the Earth.
$P = mgh$ at a distance h above an arbitrarily defined $P = 0$ location.

If I suspend the ball above the surface of a table, and if I assign the "zero" of potential energy to be that at the surface of the table, then when it falls to the tabletop, it has no potential energy left. But, if I take the zero of potential energy as that at the floor, then when it is done with its motion, still on the table, it still has potential energy left over relative to the floor—that associated with the height of the table. But, that doesn't change the tabletop result. The difference between before and after is still the same. Again, looking at Fig. 7.3 h is the same whether it's measured from the surface of the table or from the floor.

This also leads to the notion of a negative potential energy which is the standard idea in chemistry. When an electron is bound to a nucleus, we say that it has negative potential energy. When it's liberated (ionized), we say that it's free and has a positive energy and a positive energy must be supplied to the electron in order to free it from its bound state in the atom. Again, that's just the fact that the zero of the energy scale is defined for ease of use to be zero at the point of ionization.

Joule also pondered the Model that I mentioned earlier about a gas. You'll recall that picturing a gas as a collection of small, solid spheres colliding with the walls results in Boyle's Law: $PV =$ constant. Well, the constant can be shown to be the average kinetic energy of all of the little points in the gas. If each one has mass m, then $PV = C\frac{1}{2}mv^2_{\text{ave}}$. (Here C is a constant.) But, the Ideal Gas Law says that $PV = C'T$ (where C' is another constant). The really satisfying thing about this is that T is the temperature of the gas and is therefore simply a measure of the average kinetic energy of the molecules in the gas. So, heat *is* indeed a measure of energy and specifically an account of the motions of the individual molecules of any object with temperature. That's neat and had been hinted at since Newton's time, but it took 150 years for the idea to be fleshed out and understood during Joule's time (although not quantitatively by him).

Joule also studied friction and it became apparent that there was a conservation law at work that was broader than just that of motion alone. If one slides a real object down a plane, for example, it gains speed as it goes (increasing its kinetic energy) and it heats up the plane and the body through friction (heating as a loss of energy) and that adding up all of the energy at the end—kinetic energy gained, heat energy dissipated through the trip—it will all be equal to the potential energy that it had before it was let go.

7.4.2 What Comes In Must Come Out

Definition: Energy Conservation.
The total energy at the beginning of any process is equal to the total energy at the end of that process when losses due to friction and other dissipative processes are included.

That these energies add up is the statement of the Conservation of Energy—not just kinetic, not just mechanical, but *all forms of energy*. The idea was hinted at by the German physician, Julius Robert von Mayer (who always felt that he had been ignored by the physics community) and explicitly proposed by the formidable Hermann Helmholtz in 1847, who credited both Joule and Mayer. The statement of the conservation of mechanical energy is:

$$\begin{gathered}(\text{kinetic energy})_0 + (\text{potential energy})_0 = (\text{kinetic energy}) + (\text{potential energy}) + (\text{heat lost}) \\ \text{KE}_0 + \text{PE}_0 = \text{KE} + \text{PE} + \Delta Q \qquad (7.10)\end{gathered}$$

Total energy is always conserved. *Key Concept 14*

In order to make the point, let's consider the air-hockey example again in a wordy, rather than formal way in order to account for most of the energy transfers. We'll start with one puck already moving towards another:

- If you hear it sliding along the polished surface, then two things are going on: first, the rubbing of the surfaces together are heating up the surface of the table and the puck—a tiny bit. That energy loss reduces the kinetic energy.
- The rubbing sound is then propagating a compression of the air between the puck and your ears—heating it along the way. That heat dissipates throughout the room heating everything that is in contact with it. That energy loss reduces the kinetic energy.
- Once the compressed air oscillations reach your ear drum, they set it into vibration—and yes, heating it—which in turn triggers the electrochemical processes in your nervous system which your brain interprets as sound.
- Meanwhile, the puck has struck its neighbor and for a brief time more sound is emitted (more heat) and the two pucks distort slightly as the renaming kinetic energy is converted into potential energy of the lattice structure of the pucks which acts as a spring.
- The potential energy in the (slightly) springy lattice is released pushing the target puck away with the kinetic energy that's left over.

We could follow the energy all the way back to the source of the work that was done on the puck to get it started which would probably have an origin in chemical energy, either within the body of the person who shoved it or some electrical device getting its energy from an electrical grid (which could also have been a nuclear energy source). If it's a person, then the chemical energy in the food that was eaten was partially used to create the muscle action. And of course, if it's from food, then ultimately the energy of the Sun's radiation would have been responsible for the photosynthesis in plants as a direct source, or as food for an animal that was eaten.

But ultimately in any macroscopic mechanical event, what happens when everything has settled down? Everything has become...Heat. This realization, along with sophisticated thermodynamic notions like entropy (which we will not cover in this account) led physicists at the end of the 19th century to begin to speculate about the ultimate "heat death" of the universe as all energy eventually becomes aimless heat. The "death" part would happen when there are no differences among any sources of energy which are large enough to support life. We have a much different view of energy now and this will unfold as we follow Albert Einstein and his colleagues as they redefine the arguments in unexpected ways.

Wait. *You're telling me what energy does, not what energy is. . . What is energy?*

Glad you asked. *Well, a little uncomfortable that you asked. Let me try to explain with analogy. It's slippery.*

7.5 Okay, But What Is It...Really?

Energy is a sophisticated and abstract thing in physics. In fact, it's not a "thing" at all. It's not a substance. It's a concept that behaves mathematically in particular ways...and manifests itself physically in different guises. It's not surprising that it took more than three centuries to sort all of this out. We now know how to measure energy-guises. But, boy, what a mess for a long time.

Example 7.2 Diamonds are Forever

Energy as an abstraction is "just there." About the best analogy (but not a perfect one) is with the idea of economic value. Is the value of an object, or currency, a "thing"? No, it's a numerical concept which takes different guises and amounts which can at any point in a transaction be assigned a "value." Value is an economic energy.

Take a rough diamond. By itself, it has a *value* (unfortunately a value which often leads to violence and brutality) which is inherent: it can be traded with other objects which also have an equivalent value. . . like cash. In such a trade—a transaction—the total value of the two has not changed, just exchanged hands and in the process, changed kind. If you had diamonds, now you have cash. But you possess the same value.

But, suppose the diamond is cut and polished. Labor—which has a value—has been added and in turn the value of the diamond has increased and an exchange for cash would require more. But the total value of the labor, the raw diamond, and the cash has not changed. . . just shifted. The total value-amount at the beginning (the raw diamond plus the potential value of the labor before it's actually expended) is the same as at the end (the cash) but the potential value of the labor has been expended on transforming the diamond and adding to its value. All the while, this abstract quantity "value" has moved back and forth among the objects—exchanged hands, manifesting itself in various guises, but never actually standing alone as a substance.

Keep that in mind as we think about energy. We physicists tend to stop worrying about these sorts of things as we do calculations and measurements using the concept and so the delicate nature of the "what" gets pushed into the background in favor of the "how." The next example can show you how different energies are "exchanged" in a particularly useful "transaction," that of driving a construction pile into the ground.

Example 7.3 Pile Driver

Are you aware of how supports for bridges and large buildings are anchored into the ground? By brute-force! "Pile drivers" have been in use for centuries, to the present day and are impressive beasts. Even in the 1800s weights of nearly 5000 pounds would be pulled tens of feet into the air above the "pile" (an enormous nail—a beam or steel plate)—and then dropped. And then hoisted again. . . and dropped. Some pile drivers are still functioning after a century. Let's think about the effort and consequence of this machine.

From the point of the maximum height, the weight is just sitting there. It doesn't take any effort to release it but then, it's a different animal. The weight will head to earth, gaining speed as it goes, and eventually crashing into the pile with enormous force—so much that it will drive a very large steel object into the hard ground.[1] Remember that the only thing that can stop something with momentum is a force, corresponding to the total change in that momentum as it stops. Well, the pile driver eventually stops with the pile (and Earth) pushing back and providing that force. A lot has gone on during this transition from suspension to "stop."

[1] A modern pile driver can exert such a tremendous force that it actually heats up the air so much that it is capable of igniting. Diesel fuel is sometimes squirted into the region between the pile and the weight and briefly a one-cylinder diesel piston engine is produced with the fuel exploding and pushing the pile down even more

As we've seen, the trip down increased its speed, an increasing kinetic energy which is enormous since the weights are typically so large. But, free fall eventually ends and the weight begins to drive the pile into the Earth, slowing down considerably in the process as the Earth resists and eventually wins by stopping the pile and the weight. But, through some distance x, the blunt pile has shoved aside, compressed, and made room for itself in soil and rock. During each increment of time that the weight is driving the pile, the momentum of the weight is decreased and the momentum of the pile increased, conserving momentum like any collision. So, since the momentum changes, a force has been exerted on the pile and it's that force that rearranges the soil and rock. The force, created by the changing momentum, acts through a distance and *does work* on the soil.

Figure 7.4: Be the first on your block to own a Sennebogen 683 tp telescopic pile driver which can drop 32 tons over a distance of 34.8 meters.

Now, what are the different piece of energy in the pile driver example? Let's be precise. As the weight falls it is shoving aside and compressing the air which in turn locally heats it. So, potential energy of the suspended pile is going into the kinetic energy of the weight, and the kinetic energy (heat) of the air as it warms. You can probably hear the weight as it falls, and that's again more disturbance in the air that moves until it hits your eardrum. That air heats where the sound waves compresses it, and where it vibrates your eardrum, the air heats it up as well. The amounts of ear-air heating are again provided by the original potential energy. When the weight hits the pile, there's an enormous sound, which is again more air-heating, and it also locally heats the pile. Immediately, the pile (and Earth) push back on the weight which still has lots of momentum. But, that force of resistance slows the pile as it in turn does work on the pile and the earth, this time through friction and compression, heating the soil by—you guessed it—causing the molecules of soil to begin to vibrate, which is heating. Eventually, the rock is moved aside, compressing the surrounding rock and, yes heating it, until everything stops. All of the original potential energy of the weight suspended above the Earth has been converted into: heat.

Let's get a sense of the scale of Joule units of energy.

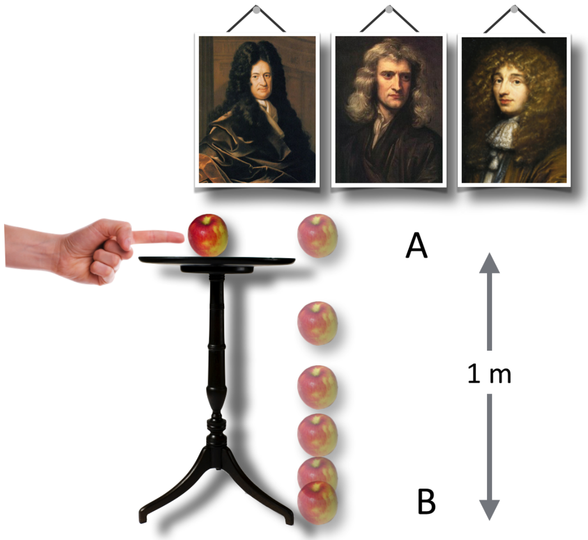


Figure 7.5: apple1m

Example 7.1 Another apple.

Question: What is the kinetic energy of an apple that falls a distance of 1 meter near the Earth?

Solution:

Suppose an apple falls from a table to the floor through a distance of 1 meter. An apple has a mass of 0.1 kg and for simplicity's sake, let's pretend that the acceleration due to gravity is 10 m/s^2 rather than its more precise value of 9.8 m/s^2.

What are the contributions to its energy at point A, point B, and halfway between them?

The contributions ot the energy of the apple would be combinations of potential and kinetic energy. Once we define where the "zero" of potential energy is located, it can be calculated at any height. Obviously, the most sensible thing to do is to define

$$\text{PE}(A) = 0.$$

When the apple is just tipped over the edge of the table, its energy is all potential and would have the value:

$$E = \text{PE}(A) = mgh_A = (0.1)(10)(1) = 1 \text{ J}.$$

That sets the scale of what 1 Joule of energy is like...Dropping an apple a meter above the ground provides it with a potential to do work on whatever it it lands on. When the apple has reached point B, its potential energy is spent, traded for kinetic energy as the apple has sped up from rest at A to the fastest that it will be just before hitting the floor (and deforming into a bruised fruit). So that energy is:

$$E = \text{KE}(B) = 1/2mv^2 = \text{PE}(A) = mgh_A$$

So we could ask how fast the apple is going, and this energy balance gives us the answer:

$$\begin{aligned} mgh_A &= 1/2mv^2 \\ gh_A &= 1/2v^2 \\ v &= \sqrt{2gh} = \sqrt{(2)(10)(1)} = \sqrt{20} = 4.5 \text{ m/s} \end{aligned}$$

But we could have gotten this same answer from Galileo's constant acceleration formula, Eq. 3.11 from Chapter 3. Finally, halfway between A and B, the energy is made up of less potential energy than A and less kinetic energy than at B.

$$\begin{aligned} E &= \text{PE(halfway)} + \text{KE(halfway)} \\ \text{PE(halfway)} &= mgH_{\text{halfway}} = (0.1)(10)(0.5) = 0.5 \text{ J} \\ E &= 0.5 \text{ J} + 0.5 \text{ J} = \text{always} 1.0 \text{ J} \end{aligned}$$

7.5.1 Classification of Collisions

In "regular life" we classify collisions into three kinds depending on how kinetic energy is handled: elastic, completely inelastic, and something in-between.

An **elastic collision** is one in which *kinetic* energy is completely conserved, which means that no energy is lost in any way. So the "normal" kinds of collisions in which the colliding objects make a sound, deform, or experience friction don't qualify. As we saw any of these circumstances take energy away from the motion and eventually all of it eventually becomes heat. We can't gather this energy up and use it efficiently and we say that these phenomena are "irreversible" which is why in part that so-called perpetual motion machines are impossible. Nature always takes energy away and doesn't return it.

Definition: Elastic Collisions.
Perfect collisions which conserve momentum and kinetic energy. Only elementary particles participate in pure elastic collisions.

A completely **inelastic collision** doesn't conserve kinetic energy and it doesn't do so...to the maximum degree possible. This happens when two objects collide and stick together, so a very particular kind of process.

Definition: Inelastic Collisions.
Collisions in which momentum is conserved, but energy is lost to heat so kinetic energy is not conserved.

Finally, in-between collisions are those which are not maximally inelastic, but not quite elastic. They're probably best represented in pool or air-hockey—the stand-in examples that I used to motivate (almost) elastic collisions. They bounce around almost conserving energy, but the fact that we hear them when they collide tells us that they're not quite perfect.

In fact, collisions of elementary particles like electrons and the whole zoo that we'll encounter are the only examples in nature of purely elastic collisions.

Definition: Totally Inelastic Collisions.
Inelastic collisions in which kinetic energy is maximally not conserved. These occur when the target and beam stick together in the final state.

To summarize:

- For Elastic Collisions: momentum is conserved and kinetic energy is conserved.
- For Inelastic Collisions: momentum is conserved, but kinetic energy is not conserved.
- For Totally Inelastic Collisions: momentum is conserved and kinetic energy is maximally not conserved.

Notice that momentum is always conserved. Note too that total energy of all kinds is always conserved. It is the *loss of kinetic energy into heat* through, say friction, that leads to kinetic energy itself being not conserved in the Inelastic Collision cases.

There is only one situation in the Universe in which collisions are perfectly, and precisely elastic: when elementary particles collide.

7.6 Energy and Momentum, From 50,000 Feet

The rules of momentum and energy conservation started as empirical observations. From the 1700s through the 1800s the science of mechanics became more and more mathematically formal. Rather than being a set of rough-and-ready tools at the disposal of engineers, mechanics and its mathematics revealed some neat things about how our universe seems to be put together. In particular, conservation laws went from a nice accounting scheme, to a clever way to solve difficult problems, to arguably the grandest of only a few universal concepts. I'll try to explain some of this later when we delve into symmetry as we understand it today but let's take a stab and meet Emmy.

Figure 7.6: A photograph of young Emmy Noether , probably around 1907, originally privately owned by family friend Herbert Heisig.

Amalie Emmy Noether (1882 - 1935) was the daughter of Max Noether, a well-regarded German mathematician from Erlangen University near Munich in the late 19th century. Max Noether was a contributor to algebraic geometry in the highly productive period where algebra was being abstracted as a very broad logical system, in which the puny subject that we learn in high school is only a small part. This particular apple fell very close to the tree and Emmy, as she was always known, turned out to be the most famous member of the Noether mathematical family (she had two brothers who had advanced mathematical training).

As a woman in Germany, only with an instructor's permission, was she was allowed to sit in on courses at a university—she could not formally enroll as a student. She did this for two years when the rules were changed and she could actually enroll an she steadily advanced to her Ph.D. degree at Erlangen in 1907. She was not able—again, due to German law—to pursue the second Ph.D. that's required in many European universities and so could not be a member of a faculty. So she stayed at Erlangen working with her father and colleagues. She even sponsored two Ph.D. students, formally enrolled under Max's name, but actually working under her. She developed a spectacular reputation and gave talks at international conferences on her work in algebra. Nathan Jacobson, the editor of her papers wrote, "The development of abstract algebra, which is one of the most distinctive innovations of twentieth century mathematics, is largely due to her—in published papers, in lectures, and in personal influence on her contemporaries."

Figure 7.7: Read about David Hilbert (1862-1943) and his 23 Problems. `http://www.famousscientists.org/david-hilbert/`

She was recruited in 1915 to work with the most famous mathematician in Europe, David Hilbert. He was racing Einstein to get to the conclusion of what became the General Relativity Theory of gravity and needed help with the complicated algebra and problems of symmetry, her specialty. Upon arrival at the Mathematics Capital of Europe, Göttingen, she quickly solved two outstanding problems, one of which has come to be known as Noether's Theorem, and which is of fundamental importance in physics today.

Hilbert fought for years for Emmy Noether's inclusion into the Göttingen faculty. He offered courses in his name, for her to teach. He led a raucous (in a early 20th century, gentile German sort of way) discussion in the faculty senate reminding his colleagues that theirs was not a bath house and that the inclusion of a woman was the modern thing to do. She was unpaid and yet still taught and sponsored a dozen Ph.D. students while at Göttingen. Einstein was particularly impressed and wrote to Hilbert, "Yesterday I received from Miss Noether a very interesting paper on invariants. I'm impressed that such things can be understood in such a general way. The old guard at Göttingen should take some lessons from Miss Noether! She seems to know her stuff."

Emmy's great grandfather was Jewish and had changed his name according to a Bavarian law in the early 1800's. However, this heritage became a dangerous burden for her and she emigrated to Pennsylvania in 1932 to Bryn Mayr College, outside of Philadelphia. There she resumed lecturing, including weekly lectures at the Advanced Institute at Princeton until she was suddenly and tragically stricken with virulent cancer that took her live in 1935. After her death, which was acknowledged around the world, Einstein wrote in the New York Times, "In the judgment of the most competent living mathematicians, Fräulein Noether was the most significant creative mathematical genius thus far produced since the higher education of women began. In the realm of algebra, in which the most gifted mathematicians have been busy for centuries, she discovered methods which have proved of enormous importance in the development of the present-day younger generation of mathematicians." But the most moving and personal obituary came from another eminent mathematician, Herman Weyl:

Figure 7.8: Emmy Noether later in life.

> You did not believe in evil, indeed it never occurred to you that it could play a role in the affairs of man. This was never brought home to me more clearly than in the last summer we spent together in Göttingen, the stormy summer of 1933. In the midst of the terrible struggle, destruction and upheaval that was going on around us in all factions, in a sea of hate and violence, of fear and desperation and dejection—you went your own way, pondering the challenges of mathematics with the same industriousness as before. When you were not allowed to use the institute's lecture halls you gathered your students in your own home. Even those in their brown shirts were welcome; never for a second did you doubt their integrity. Without regard for your own fate, openhearted and without fear, always conciliatory, you went your own way. Many of us believed that an enmity had been unleashed in which there could be no pardon; but you remained untouched by it all.

An amazing person, all the more so for her gender at time when the path for women scientists was non-existent. We'll see a few more as we go along. In any case, a crater on the Moon is named for her, a street and her childhood school are named for her, as are numerous prizes and scholarships around the world.

7.6.1 Noether's Theorem, In A Nutshell

As I mentioned, mechanics evolved into a formal mathematical framework that exposed a number of fussy, but important details. Encoded in this formalism is the regular Newton's Second law and also momentum conservation, but the wrapper is elegant and accidentally identically important in quantum mechanics and relativity. What Noether found was that this formalism included a hidden surprise. That surprise was how it would react if some of the terms were modified in particular ways.

If we were to take Newton's Second law, good old $F = ma$ and remember that the a term includes space and time coordinates, x's and t's, we can modify their appearance in the equation in particular ways. Suppose I were to take the appearance of every coordinate variable, x and change every one of them to $x + a$ where a is a constant distance, like an inch or a mile. In effect, shifting every space coordinate by a specific amount. What would you expect to happen? Should the rules of Newton change? This is in essence saying that Newton's Second law works fine here, but what if I'm not here, but I'm 20 miles away? Then I should be able to take the x and shift it by $x + 20$ and the rule should still work. My lawnmower works on the east side of my lawn as well as the west side of my lawn. And, the structure of the equation $F = ma$ is such that the 20 would go away. (Calculus is required to see this specifically.)

Definition: Noether's Theorem.
Requiring that equations of physics be invariant under symmetries in variables will insure conservation laws. A remarkable connection between mathematics and physics.

What Noether's theorem says is that this shifting of space coordinates actually speaks to an "invariance" that Newton's Second law respects...its form is not altered—and so my lawnmower works all over the yard—no matter where I am in space. This is a symmetry of nature. Nature's rules hold every*where* the same. And this symmetry has consequences that tumble out of her mathematical description of this symmetry in the hands of the fussy formalism that mechanics had become: momentum conservation falls right out.

Symmetries in physics equations imply conservation laws. *Key Concept 15*

But wait, there's more. My lawnmower works the same today as it did yesterday. And the same at the beginning of the job as at the end of the job. That means that if I take Newton's Second law...and everywhere that time, t appears, I replace it with $t + b$, where b is some constant, like 20 minutes or 24 hours. What tumbles out is another symmetry of nature and another conservation law: Energy conservation.

The remarkable consequence of these observations, is that we now can interpret our conservation laws as not an algebraic accident, or even because of an experimental result. No. Our conservation laws come about because nature requires that our mathematical rules are unchanged whether we use them today or tomorrow, or over there or over here. They hold every**where** and every**when**.

Boy, is this important! Using Noether's Theorem as a recipe, we can pick a symmetry as a test and then ask what our formal mathematical description of nature implies about physical conservation laws. If the laws work out, then we've found a symmetry of nature. If the laws are not observed in experiment, then we can discard that symmetry as not one that works in our universe.

We'll exploit this, but I've used the word "universe" many times. Let's go there.

Chapter 8

Diagrammatica: Collisions

Space, Spactime, and Momentum Diagrams

We're all about collisions, and we'll need a vocabulary and a working relationship with our Space, Spacetime, and Momentum diagrams for a few different types of collisions in QS&BB. We started this discussion in Chapter 6.5 with the simplest process of all, the dead ball billiards shot and nursed it through the discussion of Kinetic Energy. There are more.

In this Diagrammatica, I'll add to that list of one and order them all so that we have a readily available inventory of the kinds of collisions that we'll encounter. The emphasis will be on the diagrams, not on the algebra.

Nonetheless, you'll need your pencil. I'll wait…

Apart from some examples that you can work near the end, the presentation will all be about four different objects: A, B, D, and F.

A and B have the same masses, in our fake momentumunits, $m_A = m_B \equiv m = 5$; D is a big guy with $m_D \equiv M = 10$; and F is really well-fed with $m_F \equiv \mathrm{M} = 15$

In the examples that follow, A, B, D , and F will be crashing into one another at varying speeds and we'll fill in tables and draw the diagrams. The organization is intentionally very structured—the arrangement is by type of collision and I want you to return when we come across reactions of these types.

Each of the following sections will have the following features:

- There will be a table with the masses, velocities, and momenta for a particular examples of the collision featured in that section.
- There will be a cartoon of that collision for hopefully easy recognition and remembering later.
- The three diagrams will be shown for each:
- In the **Space Diagram**, imagine that the trajectories in the top frame are taken during the same time window as the bottom. So a longer line means that a particular trajectory is fast. These are sketches, and not to scale. I'll indicate the relative masses with little shaded balls on the arrows to give you a feel.
- In the **Spacetime Diagram**, the speeds are of course represented by the slopes and so pay attention to the size of the slope (steeper means faster) and the sign (positive means to the right and negative means to the left).
- In the **Momentum Diagrams**, the momentum vector's lengths correspond to the scale indicated above the diagrams.

By the way, we now know that both momentum and energy conservation are required in order to solve for most of these situations, but I'll tell you answers in the tables rather than work out the algebra in detail. It's the least I can do. No, seriously. I don't think I could do less than that.

8.1 Elastic, Two Body Scattering Event, $A_0 + B_0 \to A + B$

Two body, elastic scattering comes in many forms: the masses of the two objects matters and the relative motion of each in the laboratory results in different consequences. We'll look at four different kinds of elastic scattering involving two objects. Let's reprise our original "Simplest Collision" so that we will

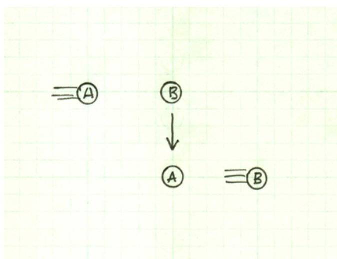
$A_0 + B_0 \to A + B$		BEFORE		AFTER	
		(mass)(speed)	$= p_0$	(mass)(speed)	$= p$
	A:	(5)(2)	= 10	(5)(0)	= 0
	B:	(5)(0)	= 0	(5)(2)	= 10
	total sum:		= 10		= 10

Table 8.1: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

have all of our diagrams in one place. The dead ball billiard shot means that the target is sitting still, so the designation "fixed target." This collision in in one dimension and the three diagrams are show in Fig. 8.1 where now I'm working according to our generic A, B, and D. The example in Table 8.1 is representative of this kind of collision which is very much a standard in many kinds of physics experiments. One prepares a target which could be solid, liquid, or gas, and shoots a beam of particles into it. Simple.

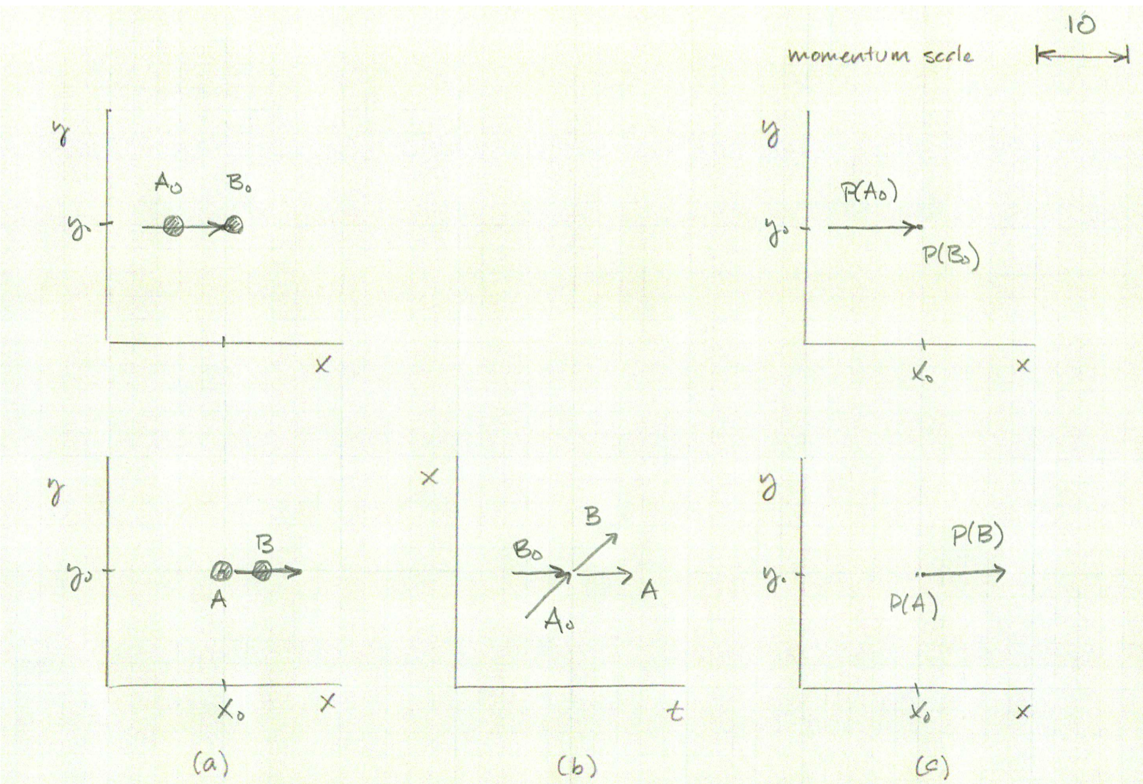


Figure 8.1: FTdiagrams2

8.1.1 Precisely Identical "Colliding Beam" Events: $A_0 + A_0 \to A + A$

Here's one more elastic collision of a particularly simple kind. Most of the particle physics laboratories of the last 30 years involve the head-on collisions of precisely identical particles (hence, the all A's above). Our relativity and quantum mechanics discussions will help to explain why this is advantageous. So I've separated these kinds of collisions for special treatment here. What follows is really a more general case of the elastic collisions that we started with in Section 8.1. It's always a characteristic of these collisions

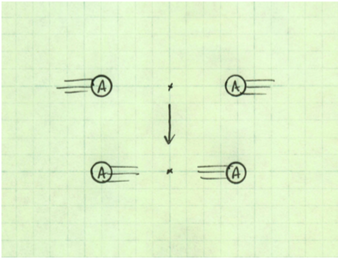
$A_0 + A_0 \to A + A$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
	A:	(5)(2)	$= 10$	(5)(-2)	$= -10$
	A:	(5)(-2)	$= -10$	(5)(2)	$= 10$
	total sum:		$= 0$		$= 0$

Table 8.2: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

that the total momentum in the initial and final states is zero. We'll see why that's useful as well. By now, you're not surprised to see that the recoil of the As from such a collision is such that they have the same velocities, but oppositely directed.

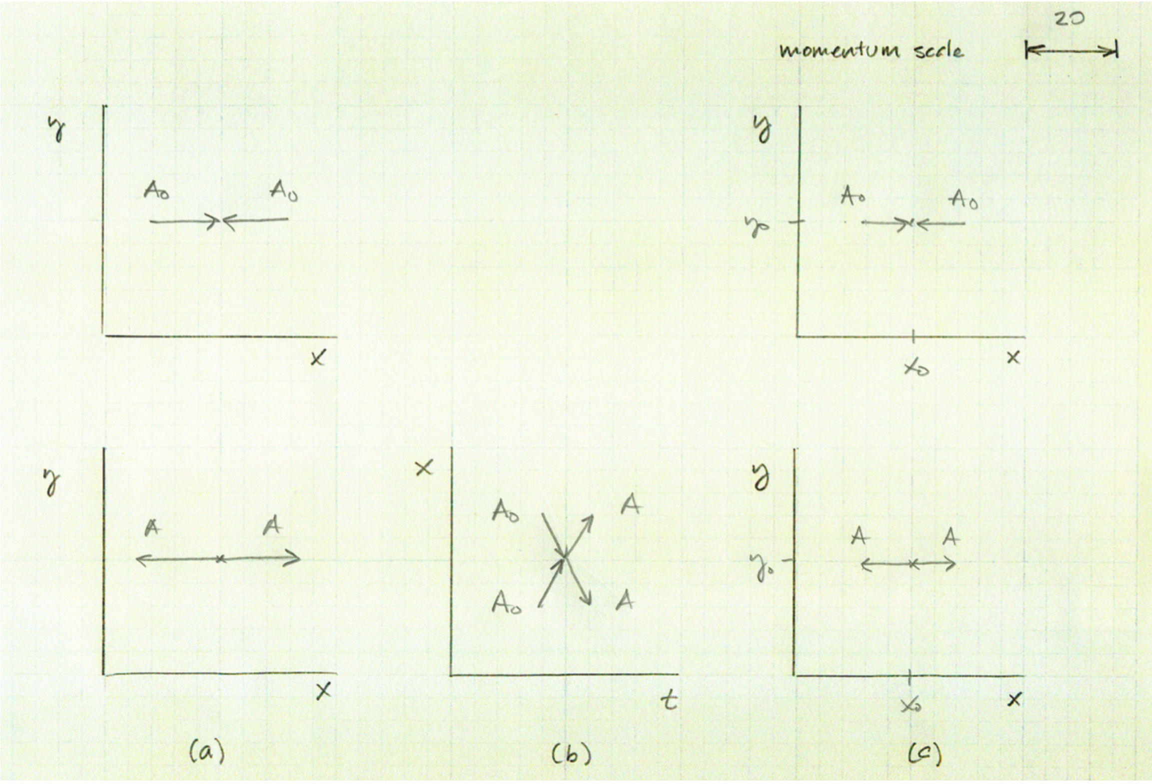


Figure 8.2: cm1Diagram

8.1.2 Moving, Bigger Target, Elastic Scattering Event, $A_0 + D_0 \to A + D$

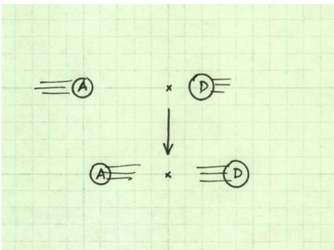
$A_0 + D_0 \to A + d$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
	A:	(5)(4)	$= 20$	(5)(-2 2/3)	$= -40/3$
	D:	(10)(-1)	$= -10$	(10)(2 1/3)	$= 70/3$
	total sum:		$= 10$		$= 10$

Table 8.3: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

Elastic scattering comes in many guises. More generally, if the target is moving, features change slightly. Here, let's involve the heavier D moving to the left slowly at -1. Then (with my back-door calculation) I've calculated the particular reaction parameters in Table 8.3. Notice how the little guy bounces off the big guy. The three diagrams for this reaction follow.

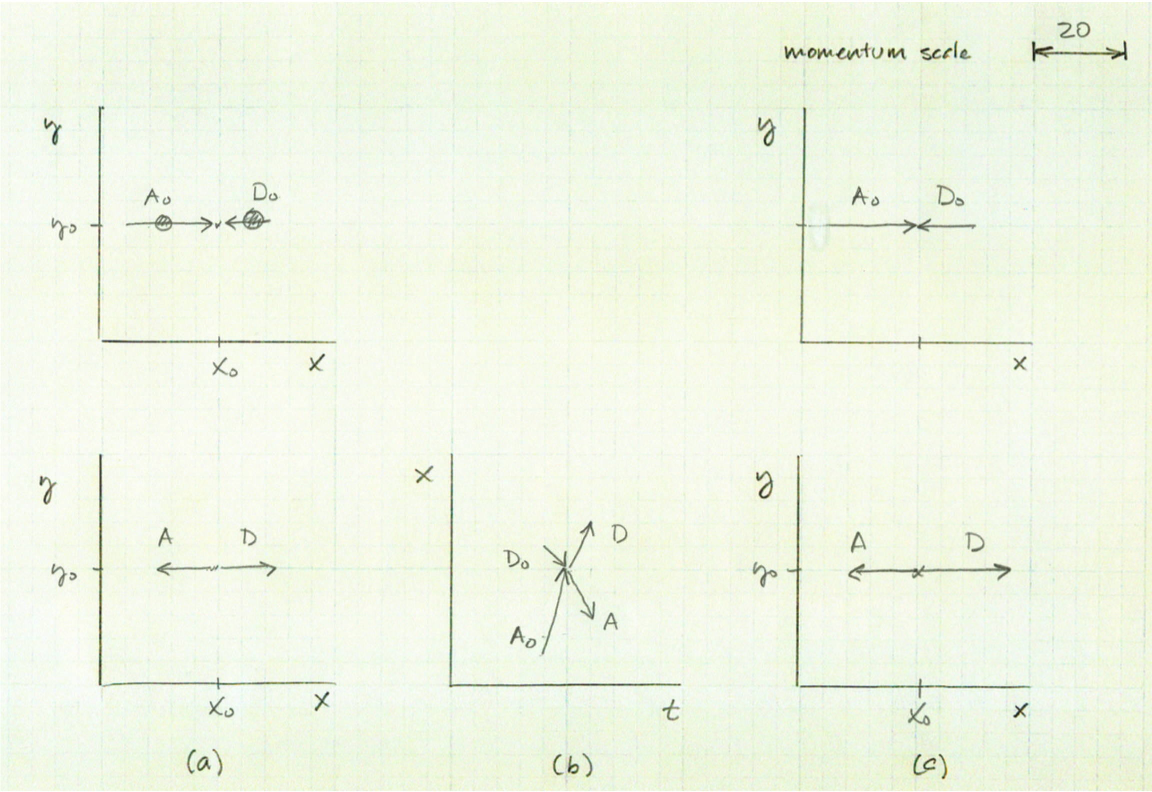


Figure 8.3: elastic2Diagram

8.1.3 Moving, Even Bigger Target, Elastic Scattering, $A_0 + F_0 \rightarrow A + F$

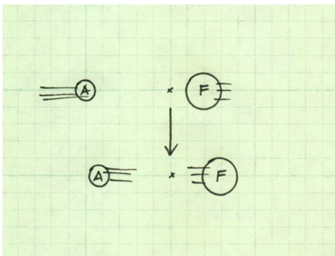
$A_0 + F_0 \rightarrow A + F$		BEFORE		AFTER	
		(mass)(speed)	$= p_0$	(mass)(speed)	$= p$
	A:	(5)(5)	$= 25$	(5)(-2.5)	$= -25/2$
	F:	(15)(-1)	$= -15$	(15)(1.5)	$= 45/2$
	total sum:		$= 10$		$= 10$

Table 8.4: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

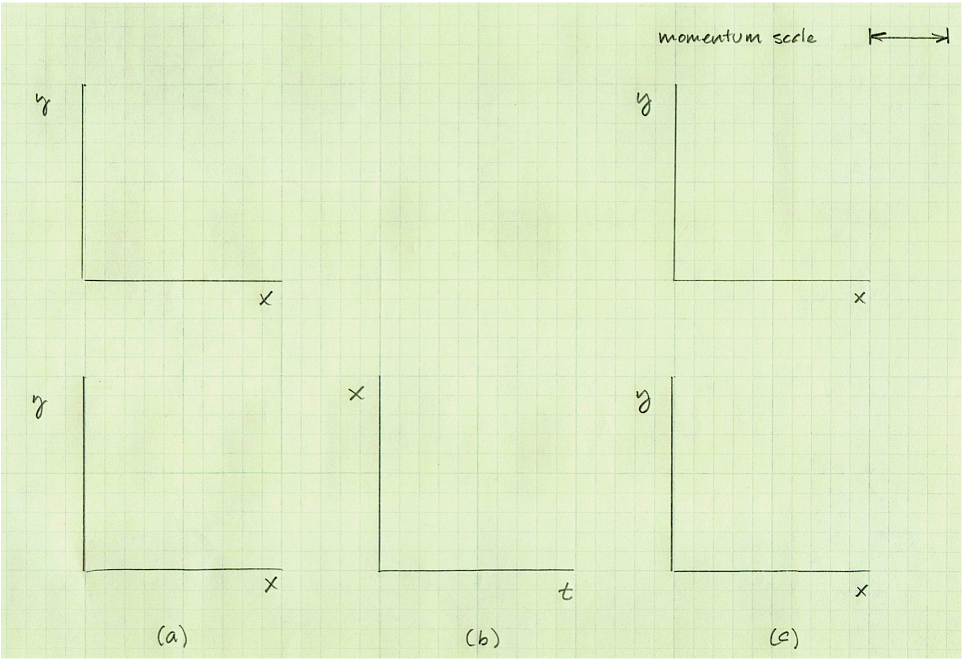
Now, let's involve the larger F moving to the left at only -1 but the smaller guy is moving much faster, and with more momentum. Then (with my back-door calculation) I calculate the reaction in Table 8.4. Notice how the big F bounces off the little guy, who has transferred a lot of momentum to it.

You should amuse yourself by drawing the diagram for this collision represented in Table 8.4.

You Do It 8.1. Spacetime

Draw the Spacetime Diagram for the $A_0 + C_0$ scattering in Table 8.4

or copy the solution



8.2 A "Decay," $D \rightarrow A + B$ Event

It's no secret that nuclei, atoms, and subatomic particles are sometimes unstable and decay into other particles. Momentum conservation is a critical part of how physicists understand these situations. Here we'll talk about the decay of one object into two objects, generically:

$$D \rightarrow A + B \tag{8.1}$$

where a big thing "decays" into two smaller, but here, identical things. Here, we'll assume that D is decaying at rest...it's sitting still. So this requires some thought!

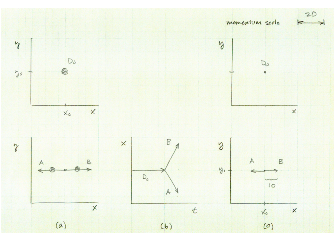
$D_0 \rightarrow A + B$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
	D:	(10)(0)	= 0	gone!	-
	A:	doesn't exist yet!	-	(5)(2)	= 10
	B:	doesn't exist yet!	-	(5)(2)	= −10
	total sum:		= 0		= 0

Table 8.5: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

Notice that momentum is happily conserved—always. But kinetic energy is not. There's kinetic energy in each of A and B in the final state, but there's no kinetic energy in the initial state. *Overall energy* is conserved...but Kinetic Energy is not. How does this happen?

Well, think about your experiences with things blowing up. If it's a firecracker or a cannon and cannonball or a person throwing something—these are all "decay" like events where one thing turns into more than one thing. In the first two, the kinetic energy of the final products comes entirely from the chemical energy that created the explosion. In the latter example, if you throw a ball the kinetic energy of the ball is entirely due to your ability to throw hard, which in turn is a feature of the elastic capabilities of your arm and the chemical processes in your muscles. So the energies are internal for all such decays...in real life.[1] We'll see that when relativity comes around, that this whole idea gets a whole new dimension added.

[1] So in Table 8.5, the speeds there are chosen at random. Their value would depend on whatever it was that caused D do explode.

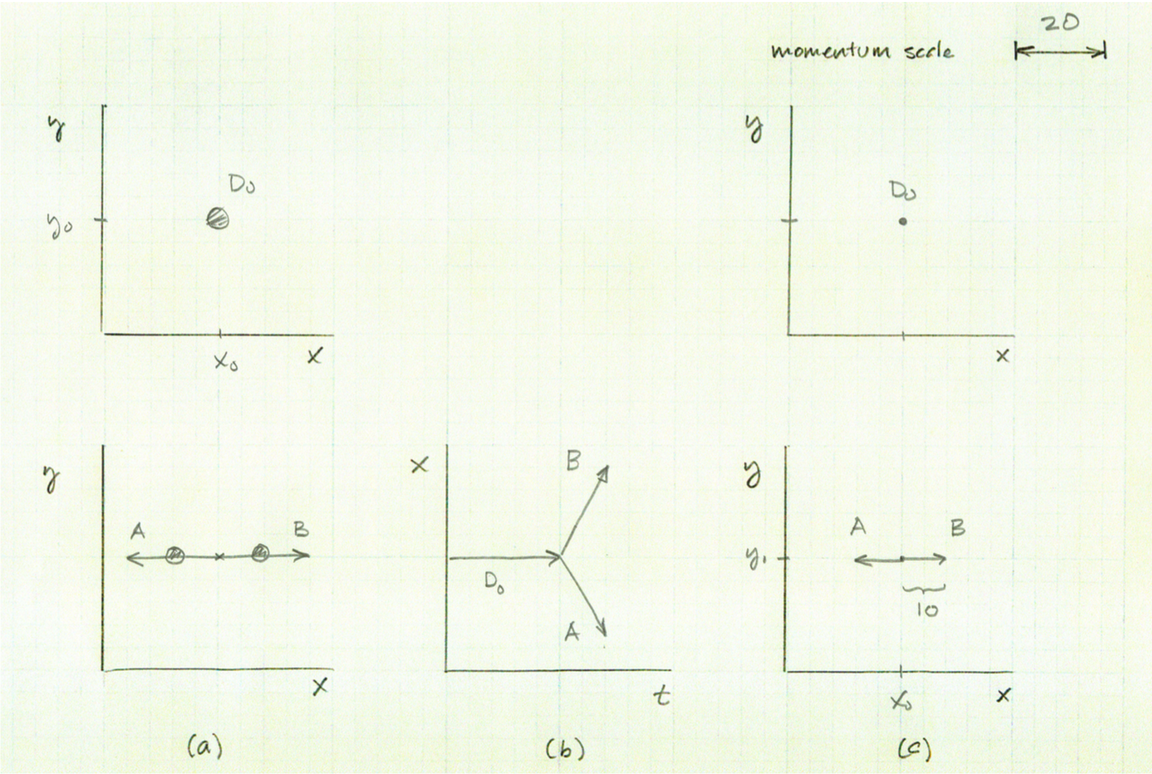


Figure 8.4: decay1DIagram

8.2.1 A "Decay," $F \to A + D$ Event

If the decay products are not the same mass, then that will affect their momenta, in order to make them balance.

$$F \to A + D \tag{8.2}$$

where a big thing "decays" into two different things.

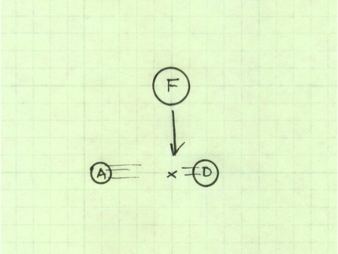
$F_0 \to A + B$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
	F:	(15)(0)	= 0	gone!	-
	A:	doesn't exist yet!	-	(5)(-2)	$= -10$
	D:	doesn't exist yet!	-	(10)(1)	= 10
	total sum:		= 0		= 0

Table 8.6: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

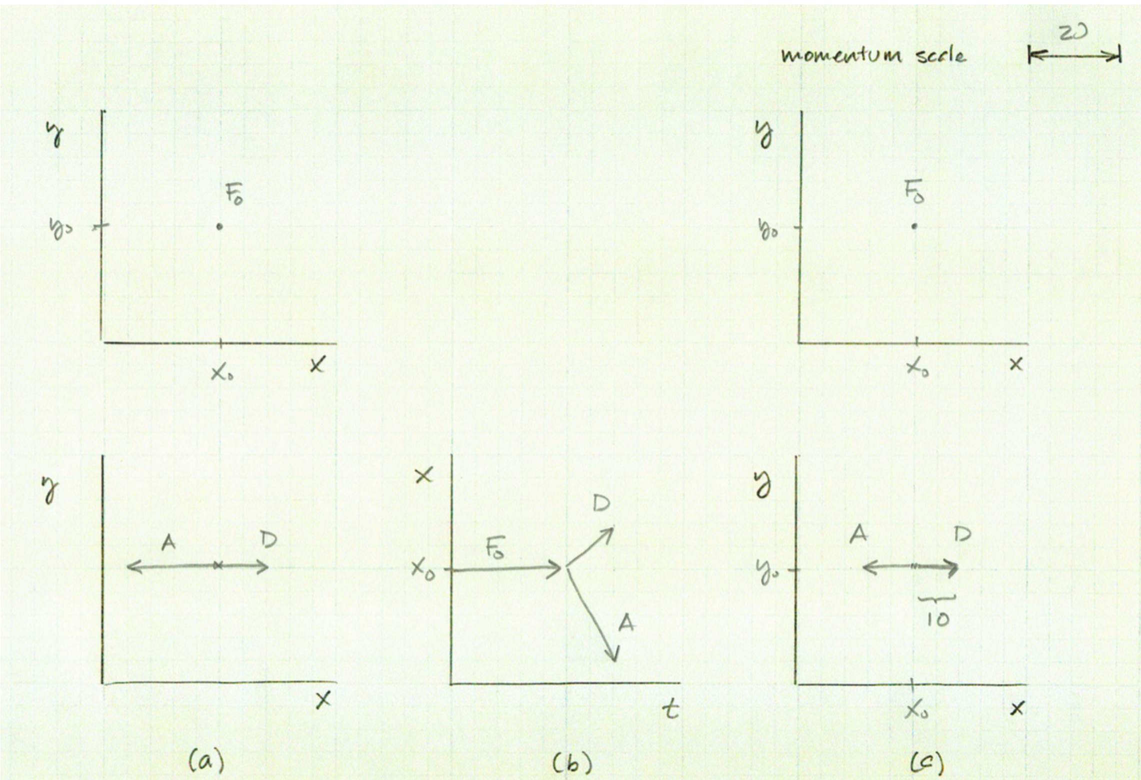


Figure 8.5: decay2Diagram

8.3 Fusion Collisions, $A + D \rightarrow F$

Opposite of a decay is a collision in which objects stick together after a collision. This could be a neutron being absorbed by a uranium nucleus in a reactor or an asteroid crashing into a planet, or two cars crashing together or an outfielder catching a fly ball. Generically, we can think of this like:

$$A_0 + D_0 \rightarrow F \tag{8.3}$$

Of course, this can be a collision in which the target is stationary, and the two move off together. Or the "target" could be moving toward or away from the beam. Here's a scenario where A and D crash together and make F. This is a completely inelastic collision and so Kinetic Energy is not conserved...but momentum always is. In this case, our objects have the same velocities, but one is twice as massive than the other. Notice that the momentum of the big guy wins, since F lumbers away to the left.

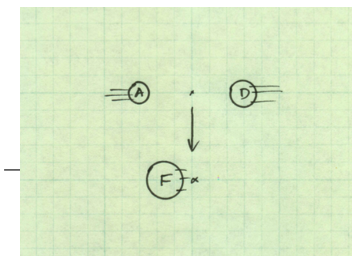
$D_0 \rightarrow A + B$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
	A:	(5)(2)	$= 10$	gone!	–
	D:	(10)(-2)	$= -20$	gone!	-
	F:	doesn't yet exist!	-	(15)(-2/3)	$= -10$
	total sum:		$= -10$		$= -10$

Table 8.7: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

There"s one for you to do at the end of the chapter. Involving middle linebackers.

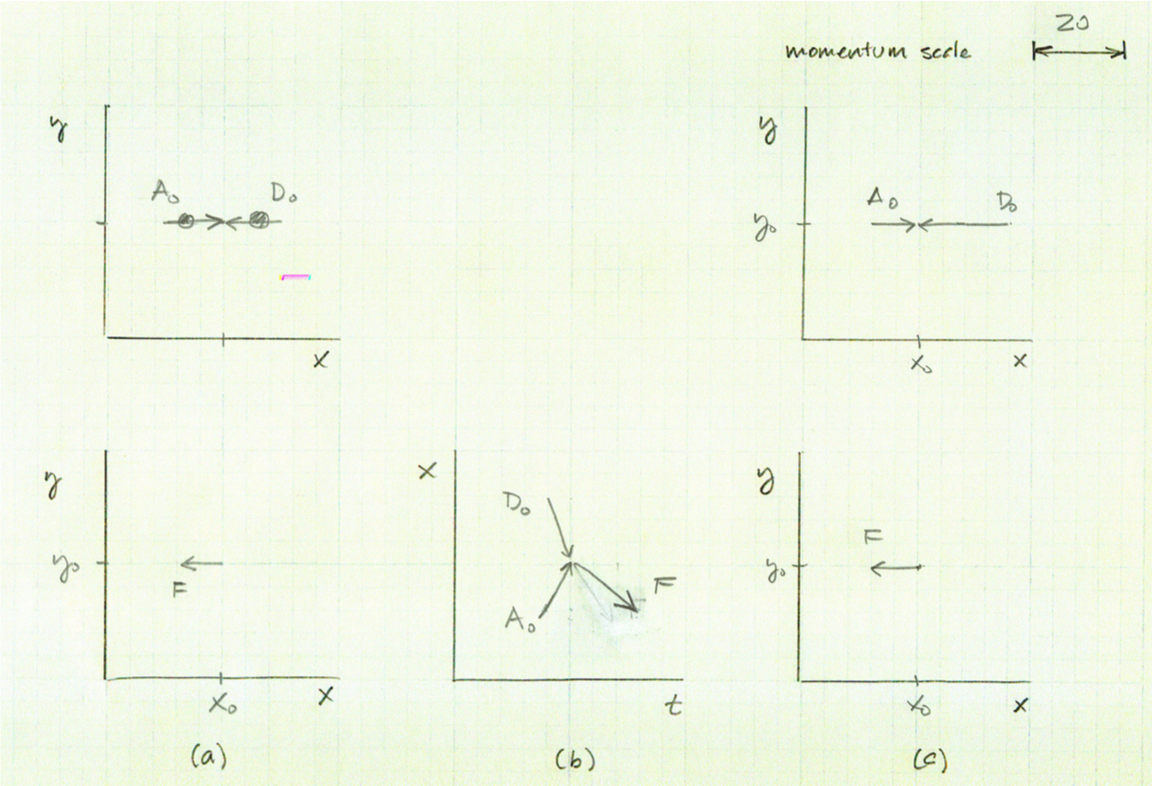


Figure 8.6: fuse1Diagram

8.4 Colliding Beam Events in Two Dimensions: $A_0 + A_0 \rightarrow A + A$

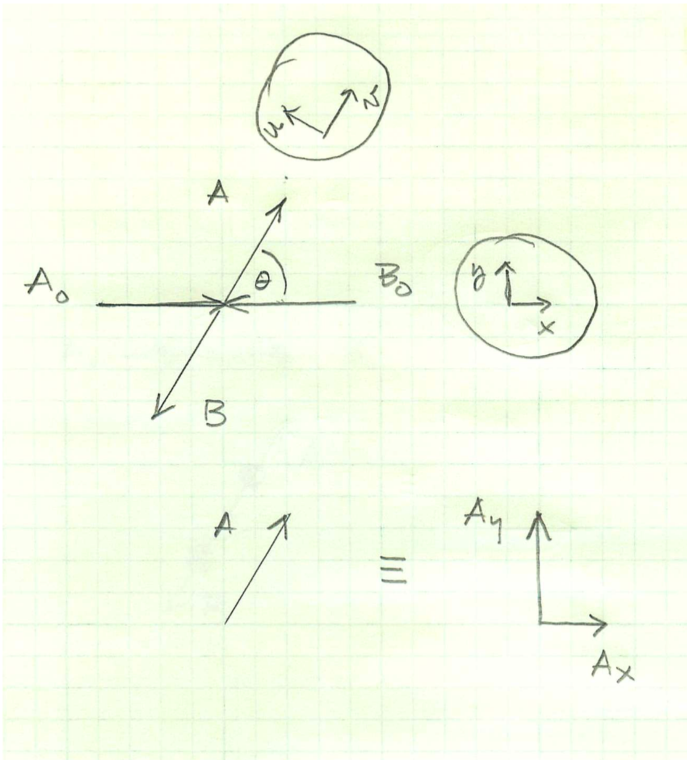


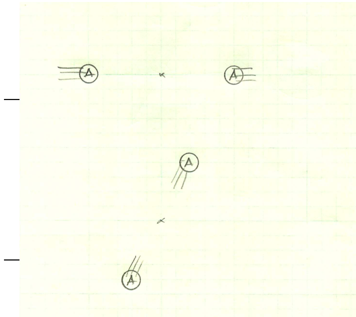
Figure 8.7: cm2dcomp

The only case in which we'll need to think in two dimensions is this: two identical beam objects scatter elastically into two identical outgoing objects. Let's do a little bit of algebra. Let's look at Kinetic Energy conservation, which works since we've defined this to be elastic scattering (no sound emitted, no heat lost, etc). Since the two initial velocities are the same, but oppositely directed and since the masses of all four identical particles are the same, this reduces to a simple relationship between the initial speeds of each and the final speeds of each.

$$
\begin{aligned}
\text{KE initial} &= \text{KE final} \\
1/2mv_0^2 + 1/2mv_0^2 &= 1/2mv^2 + 1/2mv^2 \\
2v_0^2 &= 2v^2 \\
v_0 &= v
\end{aligned}
$$

Now let's do an example with our now standard fake momentumunits. Refer to Fig. 8.7. But let's think about those to drawings. The one on the top is just the Space Diagram: A and B come in from the sides, collide, and then go off in another direction. I've indicated two different coordinate systems. The $x-y$ coordinate system is indicated with the circle around little reference axes. Particle A_0 is going in the $+x$ direction and B_0 in the $-x$ direction. Both of their momenta are anti-aligned and the total momentum of the initial state is 0.

Table 8.8: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

$D_0 \to A+B$		BEFORE (mass)(speed)	$= p_0$	AFTER (mass)(speed)	$= p$
				along x and y	
	A:	(5)(2)	$= 10$	(5)(1)	$= 5$
	A:	(5)(-2)	$= -10$	(5)(-1)	$= 5$
	total sum:		$= 0$		$= 0$
				along u and v	
	A:	(5)(2)	$= 10$	(5)(2)	$= 10$
	A:	(10)(-2)	$= -10$	(10)(-2)	$= -10$
	total sum:		$= 0$		$= 0$

8.5 Sort-of Real Life Examples

8.5.1 A Particular Decay Example

A quarterback who is initially holding the football and then throws it...is like a "decay" event. Now, quarterbacks are large and footballs are small (not anchors!), so if you've ever thrown a ball you've probably not experienced the same sort of recoil as with our boat. Let's see how much by supposing that we have the following: using arbitrary units...just numbers:

Here's what we know:

- Mass of football (F) is $m_F = 0.5$
- Mass of the quarterback (Q) is $m_Q = 100$
- Momentum of the football is $p(F) = 10$
- Call the system of Q + F (the quarterback holding the ball), LQ for "Loaded Quarterback."
- The momentum of the LQ state is 0, since the quarterback (with ball) is standing still.

Questions:

1. What is the momentum of the quarterback in the final state?
2. What is the mass of LQ?
3. What is the velocity of the football after being thrown, that is the final state?

Space Diagram

Let's orient our brains by drawing the Space Diagram, which will have before and after pieces, separating the initial and final states. Figure 8.8 shows just that. On the top is the quarterback and ball just standing there. After he throws the ball, the systems are now the Football (F) and the Quarterback (Q) and I've pictured the throw to be only in the x direction...and, I've assumed that the quarterback does indeed recoil (although our mathematics would show this). Roughly, the football goes further (from x_0 to x_f) during the "after" time interval than the quarterback, which is what you'd expect.

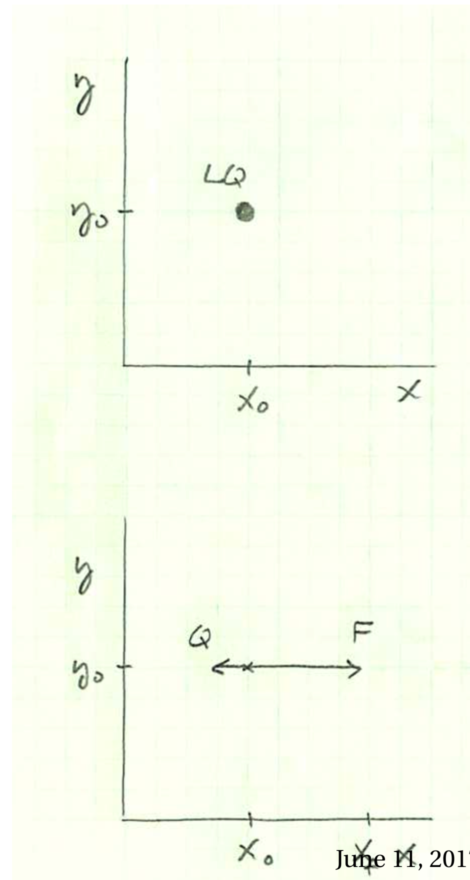


Figure 8.8: quarterSpace

Table 8.9: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

	Before	After
LQ	0	-
Q	-	b
F	-	10
total	0	a

Momentum Conservation

The particular situation is in symbols:

$$A \rightarrow b + c$$

$$\text{loaded quarterback} \rightarrow \text{ quarterback} + \text{ football}$$

$$LQ \rightarrow Q + F$$

Let's turn it into a momentum conservation equation:

$$p(\text{loaded quarterback}) = p(\text{ quarterback}) + p(\text{ football})$$

where I've removed the vector symbols since everything happens along the x axis. So any positive sign means "right" and any negative sign means "left." This is enough information to answer Question #1, what is the momentum of Q in the final state?

Let's analyze the quarterback as in Fig. 8.8. If he and the ball are both stationary, then the initial momentum of the LoadedQuarterback is zero (no velocity, so no momentum). So because of momentum conservation, after the event the final momentum of the resulting (different) system must also be zero. Since the football is thrown in the positive x direction, its momentum is in that direction as well. Think of the momentum analysis as a balancing of the books like in Table 8.9.

Let's unpack the missing entries. How about a? That's pretty easy: Because of momentum conservation, if the initial system's momentum is zero, then the entry for the total final state momentum is $a = 0$. And in order to balance, if the momentum of the football is $p(F) = 10$, then the momentum of the quarterback must be $p(Q) = -10$, so $b = -10$. Remember, while we got rid of the vector symbols for one dimension, directions still matter and the negative sign here means that the momentum direction (and hence the velocity direction) of the quarterback is in the negative x direction.

What's the meaning of this? Remember our boat and anchor? You and the boat move away from the direction you threw it. You recoil, and so does the quarterback from passing the football, precisely as a cannon jumps backward after the cannonball is explosively shot forward.

Wait. *The quarterback is much larger than the football, but they have the same value of momentum. How can that be?*

Glad you asked. *Ah. But the momentum of the quarterback is his mass times his velocity and the total momentum of that flying football can be shared between the quarterback's mass and velocity. His mass is high, so his recoil velocity is tiny. Let's see.*

Using arbitrary units for mass and velocity as well as momentum, the numbers in the list above are not too far from realistic football parameters.[2] So we have this little equation to solve:

[2] These are in the relative fractions associated with a 200 pound quarterback, a 50 mph thrown football, and a 0.5 kg football mass.

$$\begin{aligned}
p(\text{loaded quarterback}) &= p(\text{ quarterback}) + p(\text{ football}) \qquad (8.4)\\
p(LQ) &= p(Q) + p(F)\\
0 &= m(Q)v(Q) + m(F)v(F)\\
-m(Q)v(Q) &= m(F)v(F)\\
v(Q) &= -\frac{m(F)v(F)}{m(Q)} \qquad (8.5)\\
v(Q) &= -\frac{(1/2)(20)}{100} = -1/10
\end{aligned}$$

So the quarterback does recoil, but at a very, very low speed which is a fraction of the football's speed, governed by the ratio of the masses. So with such a large difference here, the recoil is probably not enough for him to even notice. So our more detailed momentum balance sheet would read:

	Before	After	(mass)(speed) = p before	(mass)(speed) = p after
LQ	0	-	(100.5)(0) =0	-
Q	-	-10	-	(100)(-1/10) = -10
F	-	10	-	(1/2)(20) = 10
total sum	0	0	0	0

Table 8.10: The momentum for the objects in the decay of a quarterback-with-the-ball into its constituent parts of quarterback and ball.

8.5.2 Decay: Summary

The best summary of a decay event are the diagrams which are all show in Fig. 8.9

a) We've already seen the Space Diagram with the stationary LoadedQuarterback (LQ) at position (x_0, y_0) so a single dot is the before picture. At time t_0 he throws the ball in the positive x direction—directly

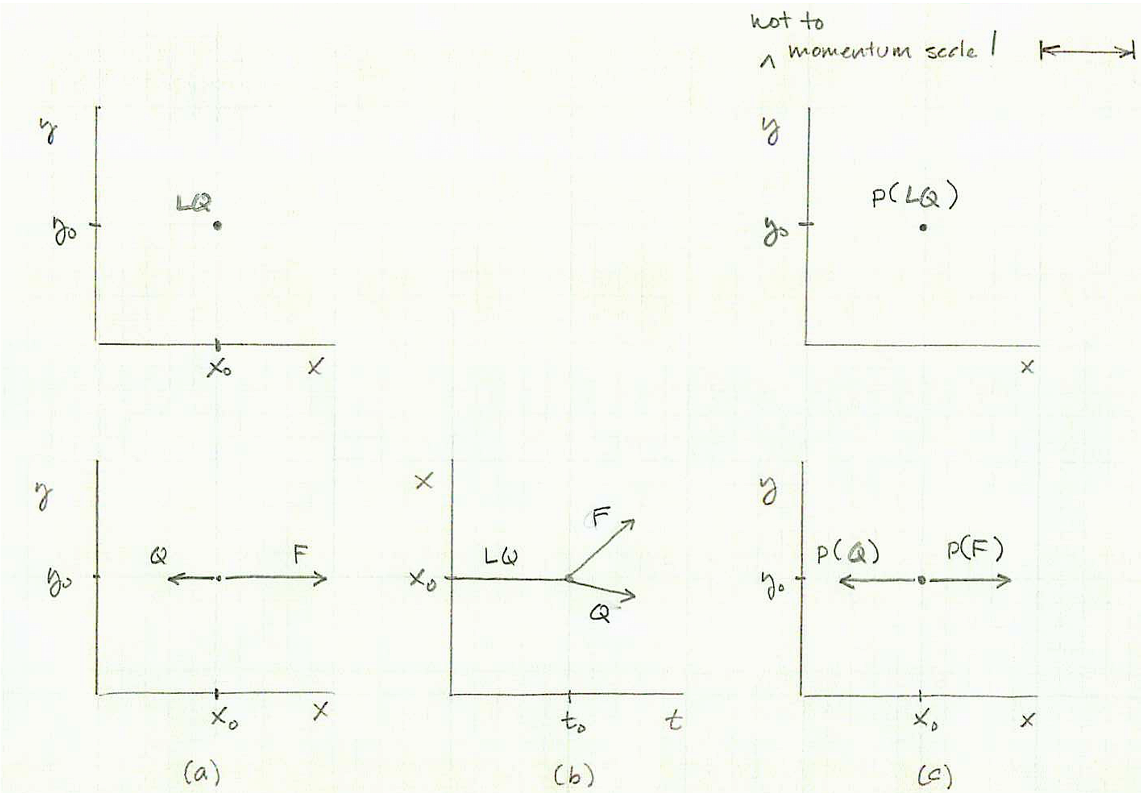


Figure 8.9: This shows all three diagrams for the decay (quarterback) scenario: (a) the Space Diagram, (b) the Spacetime Diagram, and (c) the Momentum Diagram. The top row (for the Space and Momentum Diagrams) shows the before situation and for them, the bottom row shows the after situation. The Spacetime Diagram explicitly calls out time, so before and after are all naturally there.

down-field—without any motion in the y direction (toward either sideline). The bottom picture shows the journey of the football (F) down-field and the recoil motion of the quarterback (Q) to the opposite direction. So, two objects' paths on the same "map."

b) The Spacetime Diagram shows just the x coordinate (as the y coordinate with be very unexciting in this example) as a function of time. The LQ is stationary on the field (and so no variation in x) and moves forward in time (and so the horizontal time journey line) until the event occurs at t_0 when he throws the ball. From that event LQ ceases to exist and the football goes down-field in the positive x direction (and hence the positive slope) and the quarterback (Q) recoils and goes in the negative x direction, away from the ball (and so the negative slope). Notice that the magnitude of the slope of F is much bigger than the magnitude of the slope of Q since the speed of F is much larger than the recoil speed of Q.

c) Finally, the Momentum Diagram is slightly different. First, time is implied and really only distinguished in the before and after intervals. Second, the axes are neither space nor time, but momentum units. Somewhere there could be a scale that told how many inches on the diagram would correspond to how many kg-m/s for any arrow on that diagram. Here, I'm being schematic and not precise. Notice that when LQ is stationary in the top diagram, like in space, the representation is a dot with the value of zero for both the x component of momentum (p_x) and the y component of momentum p_y. In the after representation, the momentum of the football is to the positive direction [which is governed by the positive velocity in (b) which is in turn determined by the direction of positive x] and that the momentum of Q is exactly the same length as that of F's...which is of course, because momentum is conserved. The total momentum of the system (LQ) in the top diagram is zero and so the total momentum of the whole system (now Q and F) must add to zero also. The two equal-length and opposing arrows make that happen.

You Do It 8.2. Lambert

Can we agree that Jack Lambert was the best middle linebacker in the history of the NFL? That's my story and I'm sticking to it. In fact, "sticking" is the very essence of a clean tackle. Lambert ("L") played at 220 pounds (100 kg), not so much bigger than a largish NFL quarterback ("Q" also at 100 kg). If his speed through an offensive line is 20 ft/s (6 m/s) and if a 220 pound quarterback is standing still ($v_0(Q) = 0$), when Lambert tackles him and they stick together as one unit in the final state, what is the speed of the pair of them after they collide? That is, fill in the table with $a, b, \&\ c$::

or copy the solution

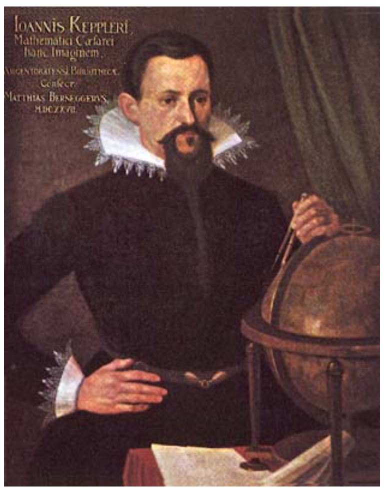
	Before	After	Your fill-in After
L	$m(L)v(L) = (100)(6) = 600$ kg-m/s		-
Q	$m(L)v(L) = (100)(0) = 0$ kg-m/s		-
LQ	-	m(LQ)v(LQ) = $(a)(b)$	( )( )
total	600	c	( )

Now we need some diagrams.

Chapter 9

Cosmology, the Old Way

Round 1: Wrestling With the Universe



Johannes Kepler

Johannes Kepler, 1571 - 1630

"I was almost driven to madness in considering and calculating this matter. I could not find out why the planet would rather go on an elliptical orbit. Oh, ridiculous me! As the liberation in the diameter could not also be the way to the ellipse. So this notion brought me up short, that the ellipse exists because of the liberation. With reasoning derived from physical principles, agreeing with experience, there is no figure left for the orbit of the planet but a perfect ellipse." *Johannes Kepler (New Astronomy, Based upon Causes, or Celestial Physics, Treated by Means of Commentaries on the Motions of the Star Mars, from the Observations of Tycho Brahe, Gent, aka Astronomia Nova)*

Everyone knows the Copernicus story. And everyone knows of the Galileo affair and Newton's apple. But the details are important and in some of these cases tell slightly different versions. Our heroes are indeed Copernicus, Tycho, Kepler, Galileo, and Newton. Characters, all. Brilliant, all. The entirety of intellectual life changed after they were done and together they, with Descartes, form the cast of characters who led the Scientific Revolution.

In Chapter 3 we considered Galileo's model of constantly accelerated motion. If that had been the only problem he'd solved, he would still have been a big deal in the textbooks. But he also discovered modern astronomy! Let's review how views of the cosmos evolved before him and then let the big three: Galileo, Kepler, and Newton take us home to a working picture.

9.1 Goals

- Understand
  - Be able to explain what the important predictions were from Copernicus' model and how they were confirmed.
  - Be able to explain how Kepler's, Tycho's, and Copernicus' models of the solar system are different.
- Appreciate
  - The ancient and Hellenistic Greek models of the solar system.
  - The Ptolemaic model.
- Be familiar with
  - The lives of Copernicus, Kepler, and Tycho.
  - The importance of Tycho's systematic approach to measurement.

9.2 A Little Bit of Kepler

One of the most interesting and courageous scientists of the 17th, or maybe any century, was the German Johannes Kepler (1571 - 1630). His personal life consisted of one disaster and tragedy after another. His professional relationships ranged from tempestuous to subservient. His first non-royal employer, Tycho Brahe, was a tyrant and upon his death, Kepler "liberated" Tycho's extensive observational data on the orbit of Mars and fought the Brahe family in and out of court for years. His relationship with Galileo was adolescent: Kepler always cheerful nearly begging for notice from the then famous Italian. Galileo basically ignored him unless he wanted something.

Kepler was continuously sick, perpetually destitute, a magnet for personal tragedy, formed by an awful childhood, lived in terrible environments, and was aggressively self-loathing.

> "That man has in every way a dog-like nature. his appearance is that of a little lap-dog. Even his appetites were like a dog; he liked gnawing on bones and dry crusts of bread, and was so greedy that whatever he saw he grabbed; yet like a dog he drinks little and is content with the simplest food...He is bored with conversations, but happily greets visitors like a dog; but when something is snatched from him, he sits up and growls. He barks at wrong-doers. He is malicious and bites people with sarcasms...He has a dog-like horror of baths..." Self-description at 25 years of age.

The single common feature of Kepler's adult life was the terrible Thirty Year's War which killed as much as a third of the German population. Whole towns would switch allegiance to Catholicism or Protestantism overnight depending on which army had passed through last. Kepler as a Protestant, was sometimes tolerated and sometimes evicted.

He was incredibly prolific, writing many books on astronomy, mathematics, and optics. He wrote one of the first science fiction novels, a third-person autobiography and left volumes of correspondence with intellectuals and political leaders from all over Europe.

He was educated to be a Lutheran minister, but stumbled into astronomy and mathematics and learned Copernicanism outside of classes with one of the early supporters.[1] He graduated but because of prodigious mathematics skills he became an atrocious math teacher at a Protestant school in Graz.

He was plagued by many questions in astronomy: Why are there six planets? Why do the speeds of the planets decrease the further away from Earth? Why are they ordered in the way they are? He was obsessively detailed in his scientific writings and so we know that on July 9, 1595, while he was boring himself to death in his own lecture, he had a sudden realization. He thought he'd come on a beautiful description of the planets' spacing. There are five so-called Platonic solids—all others can be broken down into combinations of the five: cube, dodecahedron, icosahedron, octahedron, and tetrahedron. What he found—he thought—was that if one imagined a sphere on the outside of each of the centered, increasingly bigger, and nested shapes in the order, octahedron, icosahedron, dodecahedron, tetrahedron, and cube...that the radii of those spheres correspond to the relative radii of the planets. Numerically, it was close. Figure 9.1 is the famous figure that he inserted by hand in his eccentric 1596 book, *Mysterium Cosmographicum* (*The Cosmographic Mystery*). It's madness of course. But in a nice way. Kepler's writing style was unusual, essentially a narrative, describing his highs and his lows. His passionate enthusiasm was naive-sounding, but it was who he was.

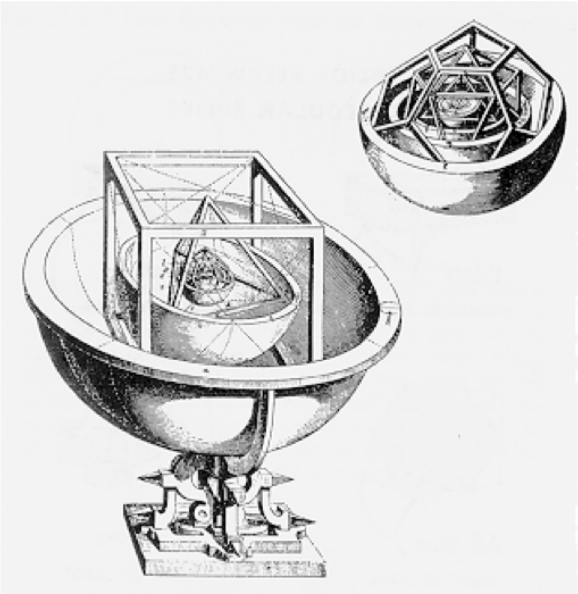


Figure 9.1: keplersolids

[1] Michael Maestlin taught the mandated Earth-centered astronomy during the day, and cultivated a private following of students who learned the sun-centered system in the evening.

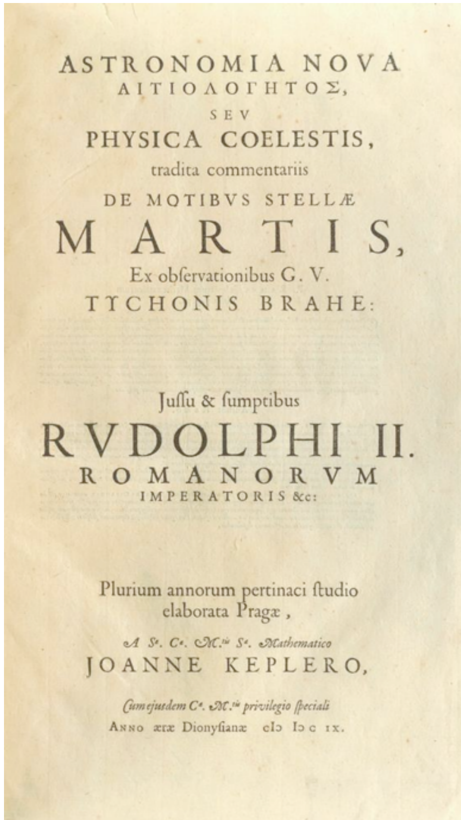


Figure 9.2: A NEW ASTRONOMY Based on Causation or a PHYSICS OF THE SKY derived from Investigations of the MOTIONS OF THE STAR MARS Founded on the Observations of THE NOBLE TYCHO BRAHE. Kepler's Big Score.

[2] whom he endearingly described as"...simple of mind and fat of body. . . "

> It is amazing! . . . although I had as yet no clear idea of the order in which the perfect solids had to be arranged, I nevertheless succeeded. . . in arranging them so happily... Now I no longer regretted the lost time. I no longer tired of my work; I shied from no computation, however difficult...day and night I spent with the calculations to see whether the proposition that I had formulated tallied with the Copernician orbits or whether my joy would be carried away by the winds. . . within a few days everything fell into its place. I saw one symmetrical solid after the other fit in so precisely between the appropriate orbits, that if a peasant were to ask you on what kind of hook the heavens are fastened so that the don't fall down, it will be easy for thee to answer him. Farewell! (Kepler, *Mysterium Cosmographicum*)

Not your standard scientific writing.

While this work was mostly fanciful at best, it set the stage for a number of similarly manic publications in understanding the motions of the planets that rhetorically meaner their way to world-changing conclusions. He started working out the questions that needed to be asked.

> "If we want to get closer to the truth and establish some correspondence ... [between the distances and velocities of the planets] then we must choose between these two assumptions: either the souls which move the planets are the less active the farther out the planet is removed from the sun, or there exists only one moving soul in the center of all of the orbits, that is the sun, which drives the planet the more vigorously the closer the planet is. . . "

It's not a spoiler to note that the idea of a force from the Sun in Newton's hands led to our classical idea of gravity. Kepler was the first to imagine such a thing, a generation before Newton.

The year after *Mysterium*, he married a rich, twice-widowed woman[2] and began a family. Within two years their first two children perished, Graz was overthrown, and he and his family had to leave. It was in 1600 when he and Tycho Brahe's paths crossed. By this time Tycho had been evicted from his island laboratory and moved his entire circus to Prague. As we'll see later, Tycho hired Kepler and assigned him the "problem of Mars." Kepler predicted that he'd work out the details of Mars' orbit in 8 days and he succeeded, but 10 years later. This he did, by basically stealing the data in the immediate aftermath of Tycho's death.

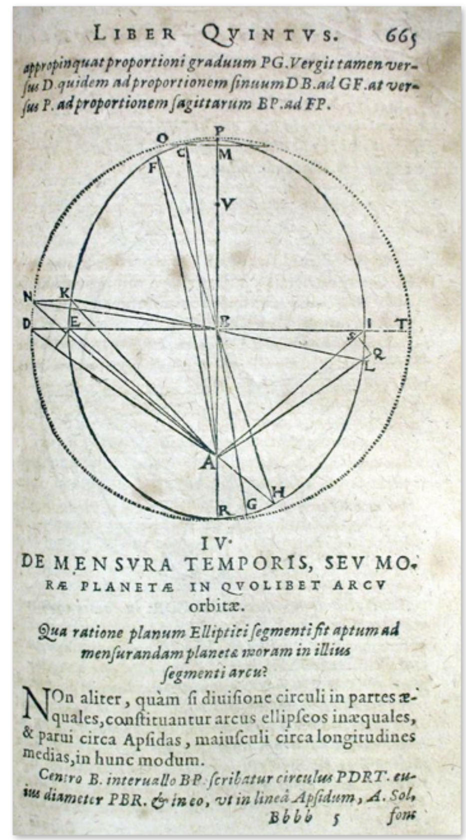
> "I confess that when Tycho died, I quickly took advantage of the... lack of circumspection... of the heirs, by taking the observations under my care..."

Kepler got Tycho's old job in 1601 as the Imperial Mathematician to the Holy Roman Emperor,[3] His Highness, Rudolf II...who was quite unstable. He spent the rest of his life trying to actually collect his salary, which became years in arrears.

[3] ...the Protestant Kepler working for the political leader of all of Catholicism.

Analytic geometry had not been invented when Kepler began working on the Mars orbit, so he relied on geometry, spherical trigonometry, and the newly invented logarithms for calculational help. Think about what he had to do. Tycho had thousands of individual position measurements of where Mars appeared in the Earth's sky through the years. Kepler's research program—as a good Copernican—was to translate those data into the apparent trajectory as viewed from the Sun. It took him six years of heroic calculations. He struggled with circles—which didn't work—then with an oval (!)—and then finally realized that the trajectory was an ellipse. He finally admits, in his Kepler-kind of way in Chapter 60 of his 1609 book, *Astronomia Nova* (*New Astronomy*):

> "Why should I mince my words? The truth of Nature, which I had rejected and chased away, returned by stealth through the backdoor... I thought and searched... as to why the planet preferred an elliptical orbit... Ah, what a foolish bird I have been."



LIBER QVINTVS. 665

appropinquat proportioni graduum PG. Vergit tamen versus D. quidem ad proportionem sinuum DB. ad GF. at versus P. ad proportionem sagittarum BP. ad FP.

IV.

DE MENSVRA TEMPORIS, SEV MORÆ PLANETÆ IN QVOLIBET ARCV orbitæ.

Qua ratione planum Elliptici segmenti fit aptum ad mensurandam planetæ moram in illius segmenti arcu?

NOn aliter, quàm si diuisione circuli in partes æquales, constituantur arcus ellipseos inæquales, & parui circa Apsidas, maiusculi circa longitudines medias, in hunc modum.

Centro B. interuallo BP. scribatur circulus PDRT. cuius diameter PBR. & in eo, vt in lineâ Apsidum, A. Sol,

Bbbb 5 fons

Figure 9.3: ellipse

This is one of the most important books in the history of astronomy. In it he enunciates two of "Kepler's Laws," which we'll talk about below. But he also—for the first time in 2,000 years of recorded history—asserts that planets do not move in (the perfect) circles that everyone believed was absolutely required of the cosmos. Think about how intellectually courageous this was, just a few years before Galileo was severely punished for an even revolutionary opinion than this.

In 1611, Kepler's son dies, as does his wife. Rudolf abdicates and his successor was not supportive and Kepler had to move with his remaining children, this time to Linz. There he advertises and interviews for a new wife, hires one, and has six more children. But the weirdness never ends for Kepler: In 1616, his mother is accused of witchcraft and by 1620 the charges were so serious that she was in danger of torture and execution. Kepler dropped everything and became in essence her defense attorney, winning the case after a year but at a cost to him.

In 1619, amid all of the disruption that was his life, he publishes *Harmonice Mundi* (Harmonies of the World). His mission was to understand the relative periods—how long it took for a planet to orbit the sun—and see if there was any pattern. Ever the mathematical-romantic he likens their motions to musical influences— the phrase "music of the spheres" comes from this work. While much of this is in the same category as his original idea of the Platonic solids, out of it came what we call Kepler's Third Law, which we'll talk about below as a crucial motivation for Newton's gravitational theory. He found a pattern that all of the planets exhibit that relates their orbital periods to the distances that they were from the Sun.

His fight with the Brahe family resulted in a settlement that required him to publish Tycho's data. He undertook this using his own funds, and so of course the printer's facility burned to the ground in the process.

Kepler's ability to rebound and find work was impressive, but sometimes he made judgment mistakes. His last employer was on the wrong side of a political divide and Kepler had to set out again looking for him in order get paid. The last we see of him, he is slowly and painfully[4] riding a horse into the rough country of continuing warfare. He never made it, dying from illness on November 15, 1630 far from home in Regensburg, Germany. His final resting place, a church graveyard, was ravaged by the Swedish army and there is no remnant of his remains. Of the 12 children that he had by his two wives, only two survived.

[4] Apparently he suffered from painful hemorrhoids. He was never shy in enumerating his physical ailments.

Kepler is considered among the greatest scientists. In 1949 Einstein considered writing a biography of him—there are many, since he was such an appealing figure. He described Kepler as one "...who had devoted himself passionately to the pursuit of deep insight into the nature of natural incidents, and who, despite all inner and outer difficulties also reached his high aim."

> "In 1930, he wrote, "In anxious and uncertain times like ours, when it is difficult to find pleasure in humanity and the course of human affairs, it is particularly consoling to think of such a supreme and quiet man as Kepler. Kepler lived in an age in which the reign of law in nature was as yet by no means certain. How great must his faith in the existence of natural law have been to give him the strength to devote decades of hard and patient work to the empirical investigation of planetary motion and the mathematical laws of that motion, entirely on his own, supported by no one and understood by very few."

Today we would call Johannes Kepler an astrophysicist, or a theoretical astronomer. He didn't make observations, but he used (new) mathematics—inventing mathematics along the way—to interpret the data and make predictions.

But we're ahead of our current story. Let's turn the clock all the way back to the first thinkers who thought deeply about the sky and recorded their thoughts. Of course, the Greeks.

9.3 Ancient Astronomy

Admit it: everyone is awestruck by the spectacle of the sky on a dark night. It's a basic human instinct and it links us to our ancestors of thousands of years ago. The rhythm of the variously repeating objects in the sky were easily mapped onto people's everyday lives which also had daily, monthly, seasonal, and annual rhythms. It's not hard to appreciate that they might have taken the sky's patterns as responsible for Earthly events and so people put a lot of thought into it! As a result, Astronomy is probably the oldest intellectual activity in all of humankind. By questioning the skies, we evolved into modern scientists with our current need to understand all of the physical universe!

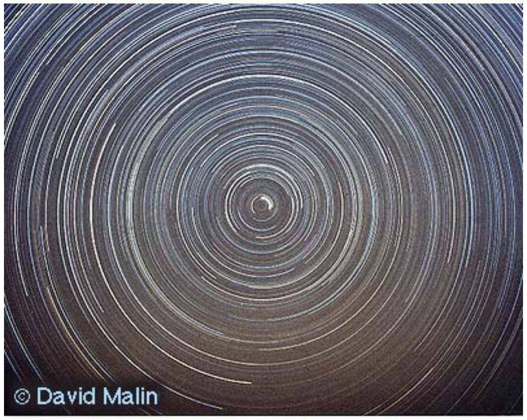


Figure 9.4: This is a long exposure of the sky around the North Celestial Pole, our North Star. The traces do suggest a circular path for each star, but we know it's actually indicative of the Earth's daily rotation on its axis.

Wait. *But what about astrology?*

Glad you asked. *Astrology is unfortunate. When it's used to make predictions or analyze a personality it means that someone's pretty gullible!*

While most of us have not looked at where the planets and stars are night by night, month by month, recording these positions and motions was a serious activity for the Babylonians, Egyptians, and Greeks. The regular motion around what we would call the North Star[5] as shown in a time-lapse image in Fig. 9.4 was highly predictive but the motions of the planets, notsomuch. In fact the word "planet" comes to us from the Greek meaning "wanderers" which is what indeed the planets seemed to do.

[5] Although going back thousands of years, Polaris would not have been at the location relative to the Earth's rotation, the center of the stars' rotation would have been a blank spot.

Basically there are three kinds of regular objects and three sorts of motions: the background stars which were familiarly constant year after year; the planets, which were relatively constant, but which executed odd motions every once in a while; and the Moon and Sun which seemed to execute their motions daily with subtle variations. Eclipses were frightening events, although information from the Babylonians enabled pre-Socratic Greeks to predict when they might occur. Comets were likewise startling, but supernovae, while few and far between, must have been deeply troubling.

While the Babylonian scholars were terrific recorders of events, the Greeks seemed to be the first to actually try to explain the cause of the stars and planets' motions. The difference between "description" and "explanation" is evident, even if their results are hard to swallow.

9.4 Greeks, Measuring Stuff

Explaining motions in the sky for the Greeks came with conceptual baggage–"theory-laden" would be the philosopher of science's term to describe their models. The mathematicians of the Pythagorean[6] school greatly influenced Plato, for whom everything is a poor copy of otherwordly: you can draw a circle, I can draw a circle, but for Plato and his followers there is only one True Circle, and that's the Ideal one. Circles were perfect in every respect–no matter how you orient them, they're identical, which you can't do with any other shape. So naturally (?) celestial motions had to be circular. Any explanation of the motions of the stars, Moon, and planets had to take this restriction as the starting place until our brave Johannes Kepler cast circles aside in the 17th century.

Look at Fig. 9.4 which shows a time-lapse photograph of the stars through a night. They appear to follow perfectly circular paths around the North Star.[7] So it's easy to believe that this observation on top of your predisposition to really liking circles, would lead Plato and others to extrapolate that motion to everything "up there." Likewise, the Sun's and Moon's motions are arcs through the sky also look to be close to circular. There was enough visual evidence that the Earth was a sphere (just a circle in 3 dimensions, right?)—understood a century before Plato and required by Aristotle.[8] The explanation that emerged placed the Earth at the center of the Universe with all of the celestial objects in circular orbit around it.

[6] Yes, that Pythagoras. Of The Theorem. That the square of the hypotenuse of a right triangle is equal to the sum of the squares of the two legs is called the Pythagorean Theorem, there are dozens of proofs and really none of them originated with him. For Pythagoras and his weird tribe of followers, the world *is* mathematics. I mean literally: the very existence of the universe *is numbers* and some of them are sacred. 10 was a sacred number. Whole numbers were sacred and when a member of the cult discovered that the length of a triangle with sides = 1 unit was not a whole number? He paid with his life, or so the story goes.

[7] This is completely arbitrary! The North Star–Polaris–just happens to be near the apparent axis around which the stars are traveling in a night. During the time of the Greeks, there was a different star and later, no star. This is because the Earth's spin is actually precessing–wobbling like a top– and over long times things like this happen. Enjoy Polaris' useful spot while it's available!

[8] Remember, he insisted that all Earthy matter would head for the center of the Universe, which was the center of Earth. So from all directions, the built-up mass that is our terra firma, would create a sphere. Also it was well-known that the Moon shines because of the Sun's light and the phases of the Moon were a reflection of the Earth's shadow across its face. So nobody of Greek influence believed that the Earth was flat.

9.4.1 Describing and Explaining

What did the Greeks observe? Well, of course they observed essentially the same things that we observe and the same things that the Babylonians and Egyptians observed! The Babylonians had a lot of data, but all they did was describe what they saw. The Greeks were the first to actually try to **explain** what they saw and for them, this was a job for Philosophers and Mathematicians. This is where they were first: using mathematical (meaning: geometrical) arguments to learn facts about the heavens. There were intellectual giants who set themselves on this task from the period between Plato (roughly 425 - 347 BCE) and Ptolemy (roughly 90 - 168 CE).

Among their accomplishments were a determination of the radius of the Earth, which was pretty close. An understanding of solar eclipses as a near-perfect blocking of the Sun by the Moon. An estimate of the distance from the Earth to the Moon (D_M) in terms of the radius of the Earth (R_E): about, $D_M = 70 \times R_E$. The creation of a very large and sophisticated star catalog, of course just positions of the stars as there were no telescopes.

The Greeks were very clever and invented the idea of not just describing Nature but trying to explain phenomena by interpreting measurements using mathematics. Explanation required some mechanism.

9.4.2 About Those Circles

How did all of those celestial objects execute those motions? The first to publish a mechanism was a contemporary of Plato's and another of the geometry giants of classical mathematics, Eudoxus of Cnidus (roughly 408 BCE - 355 BCE). He calculated that the stars, planets, the Sun, and the Moon were all attached to material spheres that rotated around axes that went through the center of the Earth. We can see through them, so he presumed that they were made of crystal: the "Crystaline Spheres." Because many of the motions were somewhat irregular and complicated, he needed many spheres with their axes of rotation inclined differently among them all. For example the Moon's motion alone required three such spheres to simulate the monthly and daily rotation and then a third to account for the fact that its orbit is slightly inclined to the horizon. Aristotle inherited this idea but took it a step further. Eventually his cosmos required 55 spheres, including one for the entire outer shell to which the stars were all thought to be collectively attached.

Make no mistake. This was not just mythology, although there was some of that in the naming of constellations. This was an attempt to describe what was actually happening so mechanisms were invented to explain why adjacent spheres would not rub against one another and impede the motions.[9] This model struggled to explain the data in three particular ways.

[9] The incorporation of little "ball bearing" like idler wheels were thought to be between spheres insuring smooth independent rotations and was an idea of Aristotles.

1. First, Venus and Mercury seemed to be related to the Sun, always very near it. That seemed hard to understand if the Sun, Mercury, and Venus all rotated around the Earth.
2. Second, Venus seemed to change its brightness in ways that no other planet did.
3. And, third, some planets appeared to suddenly go backwards!

This latter behavior is called "retrograde motion" and was very confusing. If you watch, say Mars (which is particularly obvious) night after night at the same time, it would appear to advance one way with respect to the background stars, and then reverse and advance the other. In the picture below, imagine that each line is a successively later (over days) observation of Mars ($\odot$) in relation to three background stars $(\star A, \star B, and \star C)$. Notice how on different nights, Mars moves faster than the background, but on the fifth observation... say a week or so later, that it starts appearing to move the other direction. Then, it turns around and goes back the way it originally was moving.

⊙ ★A	★B	★C
★A ⊙	★B	★C
★A	★B ⊙	★C
★A	★B	★C ⊙
★A	★B ⊙	★C
★A ⊙	★B	★C
★A	★B ⊙	★C
★A	★B	★C ⊙

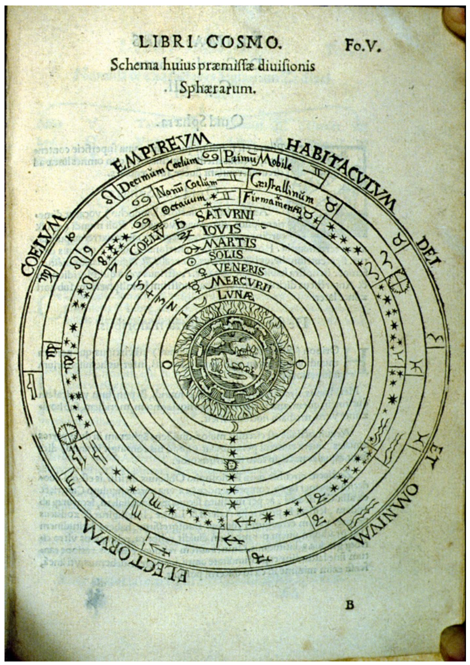


Figure 9.5: An engraving from the German mathematician and artist, Peter Apian, from a 1551 French edition of his 1524 Cosmographicus liber. Copernicus would surely have been very familiar with this popular book.

Retrograde motion happens with all of the planets and it's of course a function of our observing from the Earth. But the spheres couldn't account for this bizarre behavior since they would have intersected with one another!

Nonetheless, the spheres-picture that came from Aristotle's school was the dominant one. When Christianity took hold and Earth became the focus of everyone's attention, The Aristotelian picture was actually embedded into Church doctrine when Thomas Aquinas (1227-1274) merged Catholic dogma with the all-the-rage Aristotle's philosophy. Figure 9.5 shows a 1534 engraving by Peter Apian which conformed to the Christian cosmology: all planets, the outer stars, the Sun and the Moon revolving around the Earth. The crystaline shells were a part of the reality-picture in the Aristotelian-Church model.

9.4.3 Ptolemy

The original Greek-astronomy state of affairs lasted until a Greek-Roman-Egyptian by the name of Claudius Ptolemaios ("Ptolemy") of Alexandria, Egypt built a new model. His aim was different from his predecessor Greeks, and this is important: he wasn't trying to explain *how* the planets moved, but was trying to build a model that would accurately *calculate* where they would be at any time. This is an important distinction. Figure 9.6 shows a model...of his model. The big circle is the planet's main trajectory

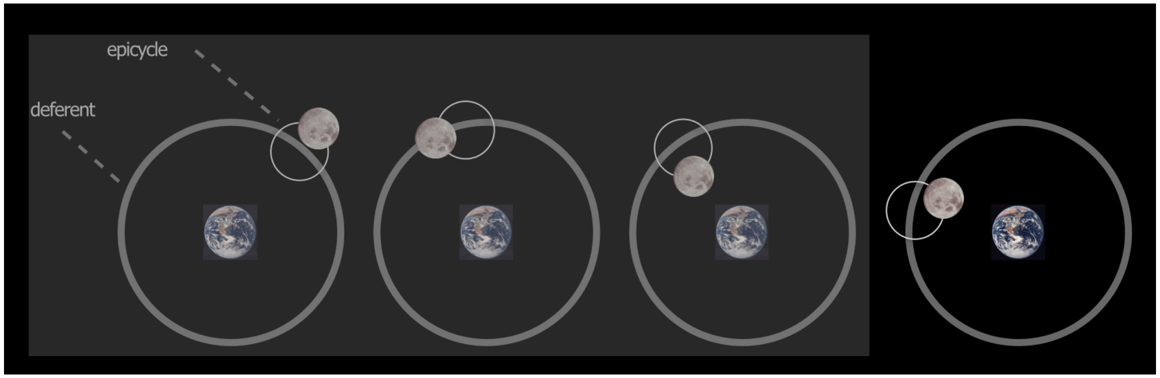


Figure 9.6: From left to right, the motion of a single planet around the Earth is shown where the planet "rides" on a little circular orbit (an epicycle) which in turn has a center that's attached to a bigger circle (the deferent) going around the Earth. In this way, one can imagine periods in which the planet would appear to be going the other way, as seen from the Earth.

around Earth called the "deferent." But the planet doesn't ride on that circle, rather it revolves around a little circle—an epicycle—the center of which revolves around Earth, riding on the bigger circle. So all of the planets, the Sun, and the Moon all have epicycle parameters where he changes the sizes of the deferent and the rotational speeds to match what he saw. This idea of "epicycles" stays in astronomy until the early 17th century.

Ptolemy was a very good observer who took careful measurements of the positions of every object and the times of those observations. In fact his data were the best until the 17th century! Increasing precision showed him that his model needed tweaking, and so tweak he did: Fig. 9.7shows that the center of the

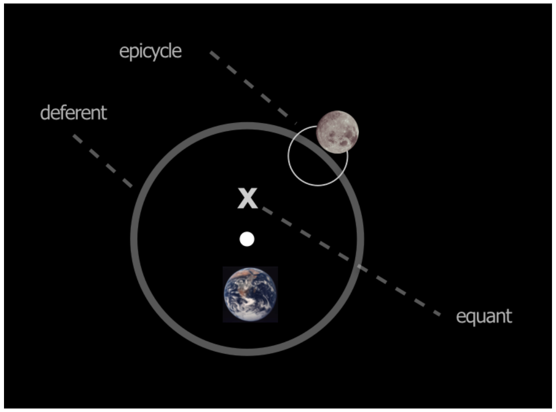


Figure 9.7: The addition of the equant and even non-uniform rate of motion was required in order to fit the data.

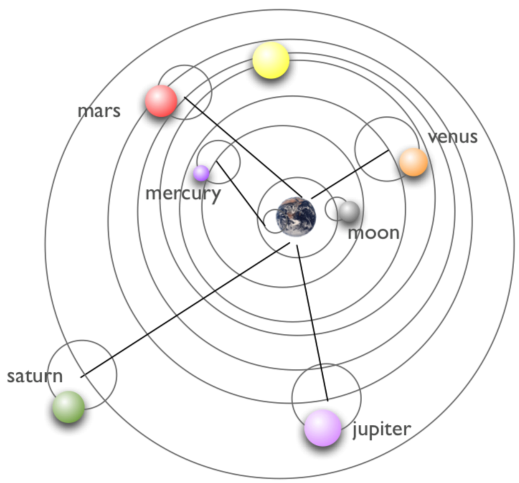


Figure 9.8: The Ptolemaic model of the solar system.

deferents for each planet were then not required to be at the center of the Earth and the rates at which the individual epicycles moved around the deferents were not required to be uniform.

Ptolemy's scientific work was far-reaching beyond astronomy, but it was his model that stood the test of time, indeed, 1500 year's worth of time. The Arabls called his text,, *Almagest*, meaning "the greatest" and it followed them into Spain, where it was eventually translated into Latin around the 14th century. Figure 9.8 shows a sketch of the whole solar system as he modeled it. If what you want is an accurate prediction of various astronomical or astrological events and you lived before 1600, then Ptolemy's model is what you needed. You just turn the crank with all of his little circles moving at their individual rates, and out would pop the positions of everything. And since many agricultural as religious events were timed by particular, near-annual astronomical events, it mattered to you.

Wait. *You mean that there was the model of planets in spheres and the model of planets in epicycles? They're very different!*

Glad you asked. *Yes, these were two different–completely different–ways to model the solar system motions. The way that people (uneasily?) managed is that the model of Aristotle was how things really were. The model of Ptolemy was just a calculator–to make predictions without expecting that the planets actually moved in the way his model pictures.*

What Ptolemy believed, is precisely what everyone believed: orbits of the extraterrestial objects were circular[10] and that the Earth was the center of the Universe and all of the planets and stars moved around it.

9.4.4 Aristarchus

Was everyone on board? There had been other models of the solar system, most memorably by Aristarchus of Samos (310 BCE - 230 BCE) who lived during the time of Archimedes, who commented on his work. Aristarchus proposed that the Sun was the center of the Universe and that the Earth and all of the planets revolved around it. This idea had also been put forward as a logical possibility before even Plato's time. But Aristotle insisted that the Earth could not possibly move since when an arrow is shot directly overhead it should land behind the archer since the Earth would have moved out from underneath. Or if the Earth moved, then it would leave its atmosphere behind.

The more serious problem with an Earth revolving around the Sun can be tested while you're reading this. Look across your room and close one eye, take note of where something in your foreground is relative to the far wall and then close the other eye and open the first. Go back and forth and you'll see that the foreground object seems to move from left to right depending on which eye you're looking through. This phenomenon of your binocular vision is called parallax.

If your left eye plays the role of some hypothetical postion of the Earth and your nose is the Sun, then your other eye is the position of the Earth six months later, on the other side of the Sun. But nobody saw the presumably fixed stars appearing to move relative to the Earth at those extreme positions…so they concluded that the Earth is not moving. Or so the argument went, since for there to be no parallax the Universe would have to be so large as to be beyond comprehension. That was Archimedes' argument: the Universe cannot possibly be so large. It was not until the 1838 that measurements could be made with telescopes precisely enough to confirm stellar parallax and hence the first demonstration that the Earth indeed moves around the sun. But Aristarchus had a fan.

Figure 9.9: The late 1500's saw the publication of a corrected version of Almagest by Johann Müller, known as Regiomontanus in 1497. This Epitome of Ptolemy's Almagest was the textbook would have been familiar to every astronomer of the 15th century.

9.5 Putting the Sun Where It Belongs

The Renaissance and the rise of humanism brought with them a freedom of thinking. Universities began to flourish, especially in Italy, Paris, and Oxford. A century before Galileo joined the faculty at Padua, an unassuming Pole also went to Padua to study medicine, for which it was particularly renown, but similarly to Galileo, he couldn't shake his fascination for mathematics and astronomy. Nicolaus Copernicus (1473-1543) was sent to Italy by his uncle who was the Bishop of Warmia to study canon law at Bologna but he actually studied in Padua, Rome, Bologna, and Ferrara. While in Bologna, he lived with the faculty astronomer and made many observations with him. (The primary job of a late medieval astronomy professor included teaching mathematics and astrology.) He eventually went on to Padua to study medicine and believe it or not, astrology was an important tool for doctors and so Copernicus was well-prepared. While he obviously had trouble "declaring a major" he did manage to receive his canon law doctorate degree and have sufficient training in medicine that he would be a practicing physician and personal assistant to his uncle for the rest of his benefactor's life.

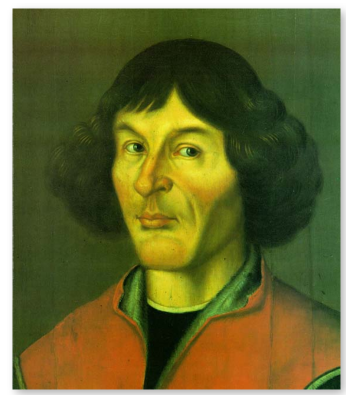


Figure 9.10: Nicolaus Copernicus

His uncle set him up with a local academic appointment, but with the stipulation that he never needed to appear on campus! He was solely detailed to tending to his uncle and as a canon[11] with minimal bishopoiric duties.

[11] Which is a secular position.

Copernicus never took vows and so was a lifelong lay-clergyman. He took a mistress and his hobby: was astronomy. He had learned Greek in Italy and slowly began to question the Aristotelian and Ptolemaic pictures, being especially irritated with Ptolemy's use of the equant, believing that it destroyed the symmetry. He knew of Aristarcus and began to think differently.

Figure 9.11: The medieval tower in cold, marshy northern Frauenburg on the Baltic Sea in Prussia (Poland) where Copernicus wrote his famous book.

NICOLAI COPERNICI

net, in quo terram cum orbe lunari tanquam epicyclo contineri diximus. Quinto loco Venus nono mense reducitur. Sextum denique locum Mercurius tenet, octuaginta dierum spacio circũ currens. In medio uero omnium residet Sol. Quis enim in hoc

pulcherrimo templo lampadem hanc in alio uel meliori loco poneret, quàm unde totum simul possit illuminare? Siquidem non inepte quidam lucernam mundi, alij mentem, alij rectorem uocant. Trimegistus uisibilem Deum, Sophoclis Electra intuentẽ omnia. Ita profecto tanquam in solio regali Sol residens circum agentem gubernat Astrorum familiam. Tellus quoque minime fraudatur lunari ministerio, sed ut Aristoteles de animalibus ait, maximã Luna cũ terra cognationẽ habet. Concipit interea à Sole terra, & impregnatur annuo partu. Inuenimus igitur sub hac

Figure 9.12: The page from Copernicus' book that inspired all of our images of the solar system. Too bad it's wrong!

What intruiged him was that the order of the planets was arbitrary in the Ptolemaic system—he thought there should be some correlation of motion with the positions of the planets. In Fig. 9.5 the planet's ordering was Moon, Mercury, Venus, Sun, Mars, Saturn, Jupiter, and the stars. Sometimes people put Venus closer to Earth. What he knew however was that the years of each planet were ordered and perhaps the Humanist fascination with the Sun rubbed off on him a little. In any case, he made a stab at suggesting a Sun-centered picture with the planets in the order that we know them now, following the lengths of the years of each as one gets further away from the Sun. His little attempt was written some time before 1514 and distributed to friends and called *Nicolai Copernici de hypothesibus motuum caelestium a se constitutis commentariolus*, a "little commentary," or *Commentariolus*. In it he lays out his plans in about 40 pages, but not his reasoning. It made it to Rome and and it's known that Pope Clement VII heard a lecture on it 20 years after its production and was intrigued. It listed Copernicus' objections and a set of assumptions: basically, the Sun is stationary, the Earth moves around it annually, the Earth rotates on its own axis daily, and that retrograde motion is a natural consequence of the relative orbit of Earth and the other planets.

This went okay and high ranking clergy even offered to support him in the production of a more complete book. But there was enough criticism and it seems that Copernicus had thin skin and he waited almost 40 years to write the complete story: *De revolutionibus orbium celestium* (*On the Revolutions of the Celestial Orbs* [orbits])... the densest treatise on spherical geometry, maybe ever.

He came to produce *Revolutionibus* somewhat reluctantly. It took decades. It seems he required an odd companion.

9.5.1 *Revolutionibus and Scandal*

The Copernican System

In school you probably learned of the Copernican system of the planets. The Sun in the center of the solar system and the planets all orbiting in perfectly circular orbits. Fig 9.12 is familiar and from *Revolutionibus*. In it, he criticizes the Ptolemaic system as a Frankenstein monster of sorts:

> “...the true symmetry of its parts...they have been like someone attempting a portrait by assembling hands, feet, a head and other parts from different sources. These several bits may be well depicted, but they do not fit together to make up a single body. Bearing no genuine relationship to each other, these fragments, joined together, produce a monster rather than a man.” To him, there was no alternative than to order the planets according to the length of their years. "Thus we discover in this orderly arrangement the marvelous symmetry of the universe and a firm harmonious connection between the motion and the size of the spheres. . . ."

Finally, the Sun, rather than just another orbiting bit takes on a central role:

> "Behold, in the middle of the universe resides the Sun. For who, in this most beautiful Temple, would set this lamp in another or a better place, whence to illumine all things at once? For aptly indeed do some call him the lantern--and others the visible god, and Sophocles’ Electra, the Watcher of all things. Truly indeed does the Sun, as if seated upon a royal throne, govern his family of planets as they circle about him."

Wait. *We all learned that the orbits of the planets are not perfect circles, but the orbits are elliptical in shape. How did Copernicus get away with circles?*

Glad you asked. *He couldn’t! In fact, he required the use of epicyles as well as Ptolemy. His were not around deferents that went around the Earth, but rather the Sun. But clearly, he could not make circles work by themselves.*

But Copernicus needed help with his circles and that came in the form of as many epicycles as Ptolemy!

Was Copernicus afraid of the Church? Not really. Remember, he had supporters and he was respected in the Vatican. He dedicated *Revolutionibus* to Pope Paul III! Things got bad for Copernicus long after he had left the scene...and after the home office-Church decided to reboot and reassert its dominance in opposition to Protestantism and the general corruption of its far-flung clerical satellite offices.

When he finished the work, his assistant had to leave to go back to his home university. He took the manuscript with him, intending to drop it off at the publisher in Nürnberg, but left oversight with another Lutheran minister, Andreas Osiander, a dabbler in mathematics and familiar with this kind of publishing.

Osiander had been in communication with Copernicus and urged him to not state that the world was the way he presented it, but that his work was just a hypothesis. Catholic Copernicus ignored that ad-

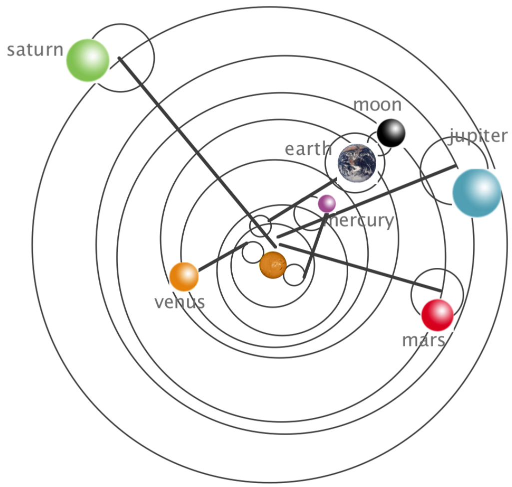


Figure 9.13: A rendering of Copernicus' actual calculational model. Many epicycles, now to mimic elliptical motion, not to solve the retrograde motion problem.

vice. But Osiander did Copernicus a dirty trick. He added a preface of his own construction, which was a scandal:

> "Since [the astronomer] cannot in any way attain true causes, he will adopt whatever suppositions enable the motions to be calculated.... **For hypotheses need not be true nor even probable**. On the contrary, if they provide calculations consistent with the observations, that alone is enough.... Different hypotheses are sometimes offered for one and the same motion (for example, either an eccentric or an epicycle model will explain the Sun's motion). The astronomer will adopt whichever hypothesis is easier to grasp.... So as far as hypotheses are concerned, let no one expect anything certain from astronomy... lest he accept as truth ideas conceived for another purpose, and depart from this study a greater fool than when he entered it."

Copernicus surely didn't know that this had been added to his book as he'd suffered a debilitating stroke and died at the age of 70 on May 24, 1543. The touching legend is that he was presented with the published version on his deathbed, but that's unsubstantiated.

Where he was buried was a mystery until 2008 when archaeologists found a skeleton under the Frombork Cathedral floor. DNA from grave matched DNA from hair found in a book that Copernicus owned. He was given another funeral in 2010 in the Cathedral, where his grave is now adorned with a handsome replica of the Solar System as we know it today.

Copernicus came a long a the right time and in the right place to re-imagine the planets in orbit around the Sun. His arguments were not driven by data—his model wasn't more accurate than Ptolemy's. Rather his argument was basically one of symmetry and philosophy. He thought that Ptolemy had described an ugly circumstance and could not explain the order of the planets.

The planets of our solar system orbit the sun. *Key Concept 16*

9.6 Astronomy: If It Ain't Baroque, Don't Fix It

Not much notice was taken of Copernicus' work. The most dismissive was Martin Luther,[12] as the new Protestantism was literal in Biblican interpretation. Copernicus' work was taught in a few places, but mostly "after school" by individuals who might lecture on the standard Aristotelian model during the day, but privately instruct students off-hours. One such instructor was Michael Maestlin at the University of Tübingen. He taught Ptolemy and Copernicus and one of his avid pupils was the young Johannes Kepler who became a Copernican at this time. He was studying for the ministry, but his mathematical skills were unusually advanced and upon graduation he was recommended for a position as mathematics instructor Protestant school in Graz, which he accepted.

[12] "There is talk of a new astrologer who wants to prove that the earth moves and goes around instead of the sky, the sun, the moon, just as if somebody were moving in a carriage or ship might hold that he was sitting still and at rest while the earth and the trees walked and moved. But that is how things are nowadays: when a man wishes to be clever he must needs invent something special, and the way he does it must needs be the best! The fool wants to turn the whole art of astronomy upside-down...However, as Holy Scripture tells us, so did Joshua bid the sun to stand still and not the earth.

As we've seen, Kepler was an amazing specimen, but physically and emotionally...a wreck. . . all of his life. He was an extreme Platonist, or even Pythagorean. This means that he believed deeply that the universe was governed by perfect mathematics—that it **is** mathematics—and this jived with his deeply religious, almost mystical inclinations. His devotion to the curious model of the Platonic solids and their seeming relationship to the planets' orbits made perfect sense to him. A believer in the reality of math.

He worked for various bigger-than-life people in Germany, Austria, and Czechoslovakia. His first big employer was the very unusual, Tycho Brahe.

Figure 9.14: tycho1596

9.6.1 The First Laboratory Director

The Dane Tycho Brahe (1546 - 1601) was, like Copernicus, another nephew of a powerful man who directed his education. However, unlike Copernicus who's father had died, Tycho's uncle actually kidnapped him to raise as his own. Tycho was not exactly a shy guy. Yes, the famous nose story is true: he was sent to the University of Copenhagen and as a 20 year old got into an argument over mathematics which led to swords, and he lost the tip of his nose. Because everyone defends mathematics to the death. For his entire life he famously wore a nose made of metal—gold, silver, brass—that he would attach with glue.[13] He'd be talking, his nose would fall off... and he's reattach it like it was a normal thing. Well there was nothing normal about Tycho. You can just make out this odd attachment in Fig. 9.14.

[13] Recent exhumation and analysis suggests that the nose was brass. Legend suggested precious metals. Oh well.

He had become a particularly astute observer of the positions of the stars and planets and even though his uncle wanted a lawyer, he was fascinated by astronomy. His uncle died and then his father, and another uncle helped him to build an astronomical observatory. This family was extraordinarily wealthy.

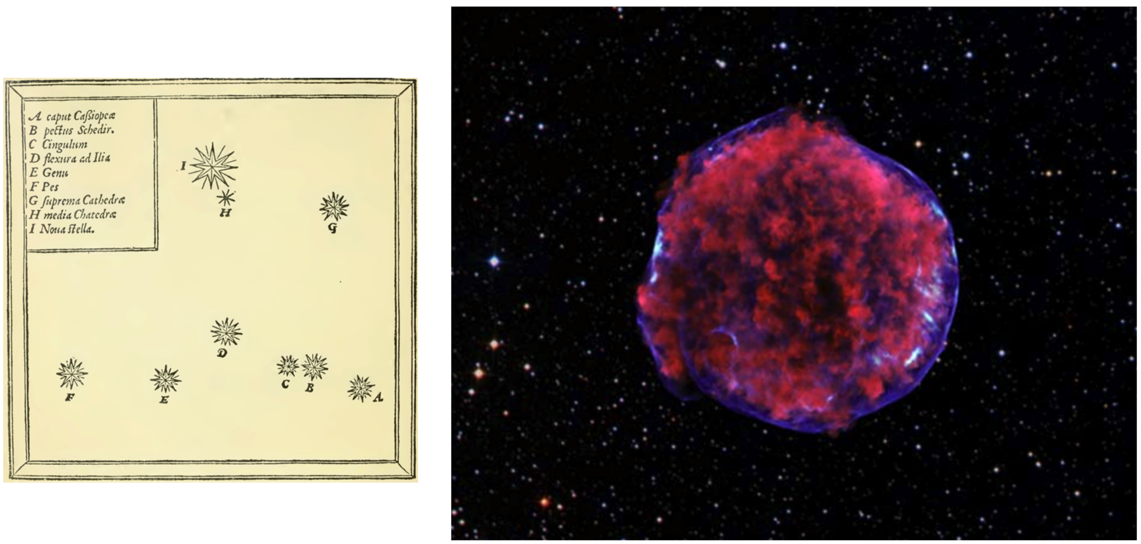


Figure 9.15: On the left is Tycho's Supernova as he drew it in his publication. It's the star labeled "I" in the constellation Cassiopeia. On the right is the image of the remnant today as observed by the orbiting Chandra X-ray telescope. (Only very powerful optical telescopes can see it now.) The blue edge is X-rays emitted from a very hot shock wave that's still expanding at a ferocious rate. The red are lower energy X-rays. The stars are a composite in the visible region It's about 55 light-years across and 13,000 light years from Earth. This is a Type Ia supernova which will become an important class of exploding star when we study 21st century cosmology.

Shattering the Crystaline Spheres

On November 11, 1572 he observed a new star in the constellation Cassiopeia. This object was so bright that it would shine during the day, which must have been unnerving. By this point in his development he was an expert observer and by measuring positions carefully he determined that there was no parallax, and that hence this *nova stella* (new star) was far outside of the Moon's orbit, and likely with the distant stars. His designation of "new star" is how we get our name "supernovae" and the event that he witnessed, and wrote about in a popular book, is called now Tycho's Supernova, or in astrophysics-geek-speak: SN 1572. Figure 9.15 is from his book, *De nova stella*.

With this observation the cracks began to form in Aristotle's model that the planets and stars were firmly attached to crystalline, Earth-centered spheres which rotated, carrying the cosmic objects with them. Aristotle also insisted that the only motions in the heavens were circular and that they were permanent. Here's the first problem for which evidence suggested otherwise.

This analysis vaulted Tycho into celebrity status in Europe as a result of his dramatic explanation. The heavens were not permanent! He further observed a comet and showed that it too had to be outside of the Moon's orbit. One of the absolute certainties in Aristotle's cosmos was that every celestial object beyond the Moon was permanent. Tycho had demonstrated convincingly that this couldn't be true since a new star was born, literally before everyone's eyes. The King of Sweden, fearing losing him to another country, gave him island of Hven in Oresund. And built him a laboratory in 1576 that was unlike anything since Alexandria. Tycho inherited not only the land, but the people who lived on the island. Uraniborg was a complete national laboratory. It had more than a hundred lab assistants, carpenters, machinists, gardeners, a police force, a printing office, and the best instruments in the world. And there for the next 23 years, every night, he and his assistants recorded the positions of 1000's of stars and full orbit positions for all of the planets over two decades. And, he came up with a model of the solar system that was Copernican, but not Copernican.

The next shoe to drop was the Great Comet of 1577. Tycho again made meticulous measurements and found a number of startling things. First, the comet seemed to be related to the Sun—its tail always pointed away from it. Second, it was also clearly outside of the Moon's orbit among the planets themsevles with an apparently varying speed. Somehow this object pierced the crystalline spheres—without any effect in its motion—while following a new circular path. Maybe those fanciful planet-carriers didn't exist? "Now it is quite clear to me that there are no solid spheres in the heavens, and those that have been devised by authors to save the appearances, exist only in their imagination, for the purpose of permitting the mind to conceive the motion which the heavenly bodies trace in their courses." He had a model.

Figure 9.16: comet1577

The cosmos changes and is not perfect. *Key Observation 5*

The Tychonic "System of the World" was a clever way to solve some of the problems that Copernicus' system also solved. In his picture, shown in Fig. 9.17, the Earth is too ponderous to move and so it's indeed stationary with the Sun and the Moon orbiting around it in a circle. All of the other planets then *revolve around the Sun*. This way Mercury's and Venus' relationship to the Sun was fixed and likewise

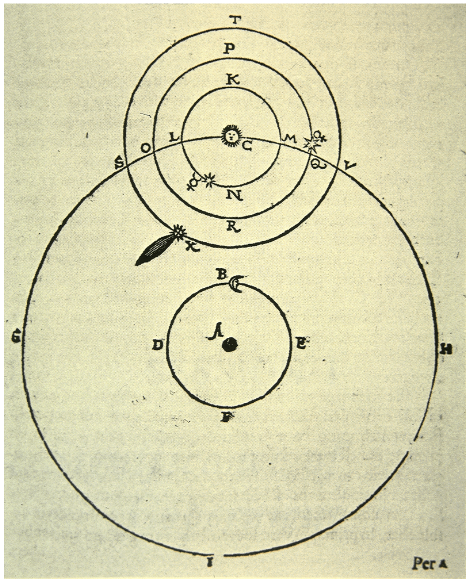


Figure 9.17: Tycho's model of the solar system is geometrically the same as Copernicus' but has the Earth still while the Sun orbits it with all of the planets then orbiting the Sun. (Notice the comet in its own little Sun-centered orbit). No crystalline spheres here.

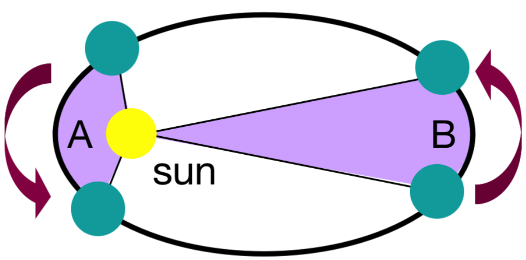


Figure 9.18: From Kepler's Astronomica Nova showing a part of his construction of elliptical orbits

retrograde motion was also accounted for and the changing brightness of Venus is solved. His objection to Copernicus was both scientific and religious. He was able to determine from measurement that if the Earth were orbiting the Sun that the lack of retrograde motion required the stars to be more than 700 times the distance to Saturn. While a long way, he was not motivated just by the sheer magnitude of that distance, but the fact that the stars appear to have a size and that size could not be so bright if they were that far away. Now of course the stars do have a size, but not that we can see from Earth. Optical effects give the impression that they are extended. Notice how the crystaline sphere idea can't work in his model. They'd overlap and crush one another. So,

Tycho's model was a legitimate competitor with the Copernican model until stellar parallax was definitively observed in the 19th century. The Catholic Church loved it as we'll see. But that wasn't enough to save Uraniborg. The new King of Denmark would not sustain funding for his lab–which was costing about 1% of the entire national budget—and so Tycho had to move. His entire menagerie, including his pet moose, moved to Prague where he became the Imperial Mathematician to the Holy Roman Emperor, Rudolf II. Yes, the Lutheran Tycho was employed by the Catholic Emperor. Tycho's moose drank too much beer and died falling down some stairs and Tycho drank too much at a state dinner and because of protocol would not leave the table. Later his bladder burst and he died painfully in 1601. And then started a brawl.

The Battle of Mars

When Kepler went to work for Tycho in Prague it turned out to *not* be a marriage made in heaven. Kepler inherited Tycho's job as Imperial Mathematician and a long headache of law suits. Tycho was possessive of his data and Kepler wanted desperately to get his hands on it and was frustrated at it being fed to him piecemeal. He was determined to solve "the Mars problem," namely the retrograde motion problem and Tycho's decades of precise Mars data was just out of his grasp. However, Tycho's mishap at the banquet left Kepler with an opportunity, which he took. Or rather, he took the data. "I confess that when Tycho died, I quickly took advantage of the... lack of circumspection... of the heirs, by taking the observations under my care..." Only to be hounded by Tycho's heirs for nearly the rest of his life. In a legal decision, he agreed to publish Tycho's data in a book dedicated to the Emperor and the *Rudolphine Tables* were eventually published, but not until Kepler had thoroughly analyzed the data. He won the war of Mars.

Aristotle's model of crystalline spheres could not be correct. *Key Concept 17*

Remember that circles were an unquestioned feature of all of astronomy for two millennia. What Kepler found was that the orbit of Mars would just not fit a circular path. After anguishing over this, in heroic calculations, he determined that the orbit was elliptical with the Sun, not at the center of the ellipse, but at one of the foci. This was an intellectually brave conclusion given the dominance of the circular prejudice. In 1609 he published these ellipse results (Kepler's First law) in *Astronomia nova*, *New Astronomy,* along with his discovery that the planets sweep out equal areas in equal times as their speeds change as the approach and move away from the Sun (Kepler's Second Law, but it actually came first). Notice in Fig. 9.18 that the area A is equal to area B, so that the planet is going faster close to the sun.

The planetary orbits are in the shapes of ellipses. *Key Observation 6*

Later in 1619 he published an extension of his cosmology in *Harmonices mundi libri*, *Harmonies of the World*. Plato's solids were still there as was a considerable amount of new mathematics, and his Third Law: for any two planets the ratio of the squares of their periods[14] is proportional to the ratio of the cubes of their average distance from center of their orbits. Said in symbols, where T is the period and R is the mean orbit radius:

$$T^2 \propto R^3. \tag{9.1}$$

Kepler was quite amazing. While considerably under pressure of family and health he came close to inventing calculus (in trying to calculate the volume of wine casks), conceiving of gravitational attraction as the source of planetary binding (he was familiar with the work of Walter Gilbert in England who determined that the Earth was a large magnet and imagined the Sun as emanating a magnetic attraction to the planets), optical telescopes (he worked out the correct geometrical optics of concave and convex lenses after Galileo's publications), and fiction (he wrote the first science fiction novel).

Kepler is one of my personal scientific heroes. He, before anyone, was intellectually brave enough to abandon circles as the path of the planets. He followed the data, rather than authority.

In some ways, he straddles the Renaissance and the birth of physics and astronomy. So in the next chapter we'll take the leap into the Scientific Revolution with Galileo's telescope and Newton's enormous intellect.

[14] The Period is the time that it takes for any repeating motion to come back to its starting point. So the period of the Earth's orbit is one year.

Supernovae happen all the time in the universe and astronomers regularly catalog them in other galaxies (not the astronomers, the supernovae!) from the Hubble Space Telescope. But there have not been any supernovae in the Milky Way galaxy since the 17th century. We saw how Tycho's Supernova led to the downfall of the crystaline sphere model of the solar system. But what's astonishing is that within a few years, on October 9, 1604 Kepler studied and then wrote about *another* Milky Way galaxy supernova: called today the Kepler Supernova! None since. Zero.

Chapter 10

Newtonian Gravitation

the lion roars*

Galileo Galilei, by Justus Sustermans, 1637.

∗ In 1697, a calculus-based mathematical competition was held throughout Europe. Newton was long out of physics, but entered anonymously. Noted the sponsor of the challenge, "we recognize the lion by his claw."

Galileo Galilei, 1564 -1642

""They know that as to the arrangement of the parts of the universe, I hold the sun to be situated motionless in the center of the revolution of the celestial orbs while the earth rotates on its axis and revolves about the sun. They know also that I support this position not only by refuting the arguments of Ptolemy and Aristotle, but by producing many counter-arguments; in particular, some which relate to physical effects whose causes can perhaps be assigned in no other way."" *Letter to Grand Dutchess Christina, 1615.*

Physics got real with Galileo's telescopic discoveries. Everyone knows the highlights of the Galileo story and his embarrassment at the hands of Pope Urban VII. The real story is perhaps different from the urban legends. As in his experiments on terrestrial motion, his conclusions on the moon's and planets' motions were more descriptive than causal. They "why"—the dynamics—of the cosmos was left to Isaac Newton to figure out. His Gravitational Model was so successful, that in the space of his lifetime, Europe went from ignorant of how Nature worked, to believing that everything can be known. The Enlightenment itself owes much of its origins to Newton's work.

10.0.1 Goals of this chapter:

- Understand
  - How to calculate the gravitational force between two masses.
  - How to calculate the weight of any object on any planet.
  - How to calculate gravitational potential energy.
- Appreciate
  - How Galileo's astronomical discoveries charted new ground in astronomy.
  - How Galileo's approach to science laid groundwork for the modern version.
  - Newton's argument regarding the Moon and the Apple.
  - What being in orbit implies about "falling."
  - Escape velocity.
- Be familiar with
  - The later lives of Galileo and Newton.
  - The importance of Galileo's Letter to the Dutchess Catherine.

10.1 A Little Bit More of Galileo

Figure 10.1: One of Galileo's original telescopes. *Museo Galileo*

When we last left Galileo, he was in Padua working out the correct understanding of falling bodies and projectiles. He didn't publish that work until he was under house arrest in his villa outside of Florence and had to smuggle it out of Italy to the Netherlands. How his arrest-story came about is legendary, and not necessarily how most people imagine it. First, some more science, then some of the back-story to his troubles with the Inquisition.

Galileo had been in Padua for 16 years when in May of 1609 he heard of a novelty that was being sold in France, Germany, England, and the Netherlands where it was invented. This was of course the telescope. Remember that he was good with with his hands and eventually employed an instrument maker. Together, within a month, from only a word of mouth description, he was able to grind and polish lenses and construct his own telescope. It was just 3x magnification, not as good as what was "out there." But he persisted in his technique and built 8x, 15x, and 30x versions Figure 10.1 shows one of his first prototypes from the science museum in Florence, Italy. He used it to look across the land and water and then...he looked up.

In August of 1609 he took his then 8x version to Venice and demonstrated it to the intellectual community, and also politicians. Remember, Galileo always had his family's debts on his mind and he gave an exhibition from the top of the St. Marco tower and showed that one could see ships much further away than with the naked eye.[1] Venice, was a maritime power and sometimes the target of naval attack from the East and so the Venetian Senate had a vested interest in this new early warning system. They were impressed, doubled his salary, and awarded him lifetime tenure at the university... but also froze his salary at the new level. Although he was now one of the most highly paid professors in the Venetian Republic, the prohibition of any raises for the rest of his life didn't sit well with the ambitious 45 year old.

[1] Famously, an Aristotelian philosopher, Guilio Libri refused to look through the telescope. We'll learn that Galileo's mouth often got him in trouble and he suffered fools badly. Libri died soon after the incident and Galileo remarked that now he could see Jupiter's moons as he passed by it on his way to heaven.

10.1.1 What Galileo Saw!

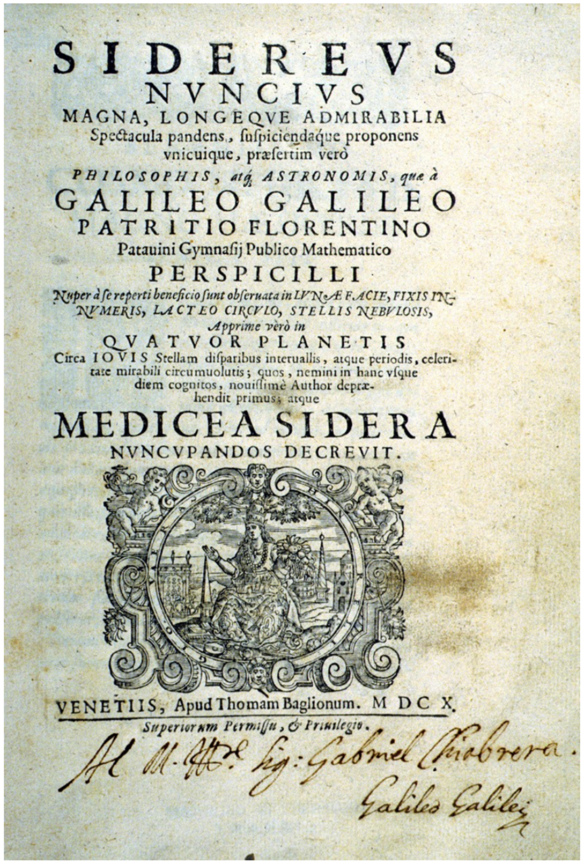
Through the next year Galileo observed things that nobody had previously imagined. In November and December of 1609 he carefully studied the Moon and with his excellent artistic abilities, drew detailed images showing the mountains and craters. This was revolutionary because the Aristotelian model required the celestial bodies to be perfectly unblemished spheres. By carefully mapping the shadows of the Moon, Galileo estimated the height of crater edges and found them to be Earth-like in size. Then a month later he studied Jupiter and found what looked like three bright stars, all in a line. He continued looking in successive nights and saw a fourth "star" peek out from behind the planet and found all four of them to be moving together! Subsequent observations convinced him that they were bound to Jupiter, and not stars at all: Jupiter has moons which today we call the Galilean Moons. And, when he looked into deep space the stars multiplied. He found hundreds of stars that nobody had ever seen before.

It's important to realize that Galileo did not invent the telescope (one of those persistent myths) and he was not the first to use it to discover things in the sky. A British natural philosopher, Thomas Harriot, was first to observe many of the things that are credited to Galileo. Harriot was not as self-promoting nor did he publish as quickly as Galileo, so he lost his historical moment.

In 1611 he published *Sidereus Nuncius*, or *Starry Messenger*, reporting these and other revolutionary observations and interpretations. Figure 10.2 shows the elaborately constructed title page and Fig. 10.3 shows a few pages from the text. When he was at Pisa, he had become friendly with the Medici family, especially the the Grand Dutches Christina. A few summers while he was in Pisa he was brought back by her to Florence in order to tutor her young son, the young Cosimo d'Medici. His bold naming of the moons after his former student—by then the reigning Duke of Tuscany—and his dedication of *Sidereus Nuncius* to him was an obvious ploy to again improve his circumstances and to get a new job without teaching responsibilities.[2]

[2] "...scarcely have the immortal graces of your soul begun to shine forth on earth than bright stars offer themselves in the heavens, which, like tongues [longer lived than poets] will speak of and celebrate your most excellent virtues for all time." A tad syrupy perhaps?

Figure 10.3 shows his sketches of multiple nights' viewing of the Medician Moons. What he had found was a miniature Copernican system within the bounds of our own solar system. Further, *Siderius* described his discovery that Venus had phases like the Moon which explained why it appeared to change brightness periodically, just as Copernicus had predicted. And finally, the number of stars visible with the telescope dwarfed what everyone believed was the full compliment of stars that had been carefully tallied by the Babylonians, Greeks, and Tycho. The universe appeared to be a much more interesting place than anyone had imagined.



THE HERALD
OF THE STARS
unfolding
GREAT, and HIGHLY ADMIRABLE
Sights, and presenting to the gaze
of everyone, but especially
PHILOSOPHERS, and ASTRONOMERS,
those things observed by
GALILEO GALILEI
PATRICIAN OF FLORENCE
Public Mathematician of the University of Padua
with the aid of a
TELESCOPE
which he has recently devised on THE FACE OF THE MOON, INNUMERABLE FIXED STARS, THE MILKY WAY, CLOUDLIKE STARS, and especially concerning
FOUR PLANETS
revolving around the star of JUPITER with unequal intervals and periods, with wonderful swiftness, which, known to no-one up to this day, the Author most recently discovered for the first time; and
DETERMINED TO NAME
THE MEDICEAN STARS

Figure 10.2: The title page of *Sidereus* from a copy in the University of Oklahoma science library. The translation is from a 19th century translation. Notice that Galileo signed this copy.

The effect of all of this news electrified Europe and overnight, Galileo became famous, and remained so for the rest of his life. The good news? He got the job back in Florence. And the bad news: Florence was within the sphere of influence of Rome and the Pope. In fact, there had been a number of Medici popes in the family. Venice was much more liberal and was often at odds with the Vatican over one or another issue.[3] Galileo's. . . unusual views. . . were safe in Venice, but dangerous in Florence.

He negotiated the position that not only paid well, but also importantly, raised his stature: he was The Chief Mathematician of the University of Pisa and Philosopher and Mathematician to the Grand Duke. The last title was important, for it was Philosophers who ruled the academic roost and mathematicians

[3] Pope Clement V excommunicated the entire population of Venice in 1309! Interdicts—forbidding any ecclesiastical functions were instituted against Venice in 1202, 1284, 1480, 1509, and again in 1609.

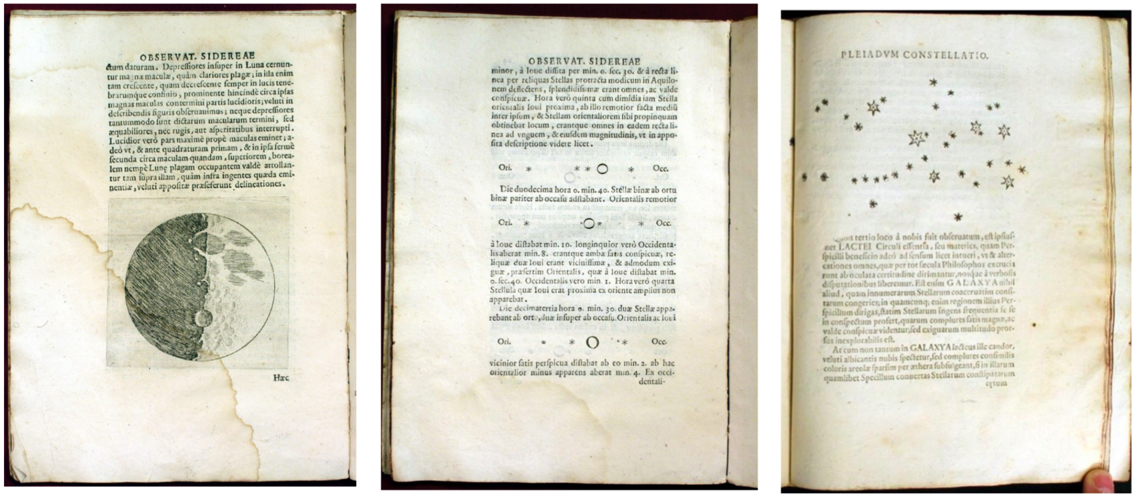


Figure 10.3: Three of the many sketches in Sidereus Nuncius. The Moon picture is famous and meticulous. The middle drawing is one of many documenting the motion of Jupiter's four moons orbiting the planet. The right figure is his sketch of the Pleiades constellation with its seven ("seven sisters") stars and then all of the new ones visible through his telescope.

were the least respected. Galileo insisted on this dual, contradictory title. He took multiple victory laps in Rome where he was celebrated by the College of Jesuits and where the then Cardinal Barberini took great pleasure in Galileo's friendship. The 47 year old was riding high.

The Moon has a rough surface with mountains and valleys. *Key Observation 7*

Other planets in our solar system have moons that according to Kepler's model. *Key Observation 8*

Figure 10.4: Dominican friar, Tommaso Caccini, raised the first public attack on Galile from this pulpit in Santa Maria, Novella in Florence in 1614.

In years to come, Galileo studied many things and wrote books on Sunspots (He learned to train his telescope on the Sun, but a student taught him to project the image onto a piece of paper so that he would not damage his eyes. The result was another kind of blemish in a heretofore perfect celestial sphere: sunspots.) and buoyancy. Here he began to get himself in trouble as a respected Jesuit competitor disagreed with him on the origin of sunspots (were they just another set of planets?) and buoyancy...what caused things to float. In both cases, Galileo was over the top and *ad hominem* in his nasty criticisms of his scientific adversaries. Ths cost him support among some of his Jesuit colleagues.

The Sun has blemishes on its surface that change in time. *Key Observation 9*

10.1.2 The Most Famous Letter in the History of Science

In 1614 Galileo was denounced by name from the pulpit (Fig. 10.4)of Santa Maria Novella—in Florence—by a conservative Dominican priest. He had begun to be suspected of heresy—was formally denounced to the Inquisition in 1615—and there was a growing unhappiness with him within the most doctrinaire of the Church's hierarchy. He reacted in what was to become the Galileo-way: a strong defense is always a strong offense.

In 1615 Galileo circulated a long, open letter[4] to the Grand Dutches Christina purporting to explain a debate at a meal that he was not at, but where his views were the topic of discussion.[5] It's worth quoting in length, for it forms the rallying cry of the new approach of Natural Philosophy as it morphs into a real scientific attitude:

> Some years ago as Your Serene Highness well knows, I discovered in the heavens many things that had not been seen before our own age. The novelty of these things, as well as some consequences which followed from them in contradiction to the physical notions commonly held among academic philosophers, stirred up against me no small number of professors–**as if I had placed these things in the sky with my own hands in order to upset nature** and overturn the sciences. . .

> Showing a greater fondness for their own opinions than for truth, they sought to deny and disprove the new things which, if they had cared to look for themselves, their own senses would have demonstrated to them. To this end they hurled various charges and published numerous writings filled with vain arguments, and they **made the grave mistake of sprinkling these with passages taken from places in the Bible which they had failed to understand properly**.

He's just getting warmed up:

[4] http://inters.org/galilei-madame-christina-Lorraine

[5] He admits that he was influenced by the bumper-sticker comment of Cardinal Baronies: "The intention of the Holy Ghost is to teach us how one goes to heaven, not how heaven goes."

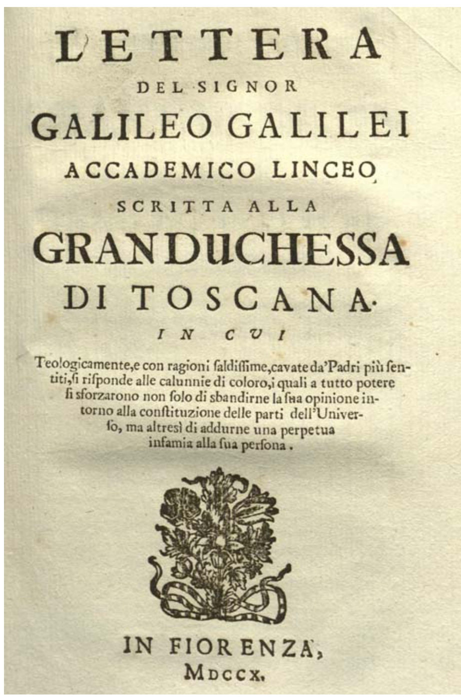

LETTERA
DEL SIGNOR
GALILEO GALILEI
ACCADEMICO LINCEO
SCRITTA ALLA
GRANDUCHESSA
DI TOSCANA.
IN CUI
Teologicamente,e con ragioni saldissime,cavate da' Padri più sentiti,si risponde alle calunnie di coloro,i quali a tutto potere si sforzarono non solo di sbandirne la sua opinione intorno alla constituzione delle parti dell'Universo, ma altresì di addurne una perpetua infamia alla sua persona.
IN FIORENZA,
MDCCX.

Figure 10.5: Title page of the Letter to the Grand Dutches. Galileo wrote it formally in 1616 and it was eventually printed in Latin in 1636, three years after he'd been incarcerated.

> Again, to command that the very professors of astronomy that they must not see what they see and must not understand what they know, and that in searching they must find the opposite of what they actually encounter is beyond any possibility of accomplishment.

And the punch-line:

> Now, if truly demonstrated **physical conclusions need not be subordinated to biblical passages, but the latter must rather be shown not to interfere with the former**, then **before a physical proposition is condemned it must be shown to be not rigorously demonstrated**... and this is to be done not by those who hold the proposition to be true, but by those who judge it to be false.

Finally:

> Inasmuch as the Bible calls for an interpretation differing from the immediate sense of the words, it seems to me that as an authority in mathematical controversy it has very little standing... **I believe that natural processes which we perceive by careful observation or deduce by cogent demonstration cannot be refuted by passages from the Bible**....The primary purpose of the Holy Writ is to worship God and save souls...

The nub of the argument was:

> They know that as to the arrangement of the parts of the universe, I hold the sun to be situated motionless in the center of the revolution of the celestial orbs while the earth rotates on its axis and revolves about the sun. They know also that I support this position not only by refuting the arguments of Ptolemy and Aristotle, but by producing many counter-arguments; in particular, some which relate to physical effects whose causes can perhaps be assigned in no other way.

Figure 10.6: christina

His war with theology has begun. He's saying: the Bible does not determine what is the case in the physical world. What's observed does. In order to overturn a fact in the world, only another observation can condemn it, not scripture. His critics accused him of re-interpreting the Bible, which smacked of protestantism and was against Church law according to the defensive Council of Trent.

Only measurements can challenge observations about the physical world. *Key Concept 18*

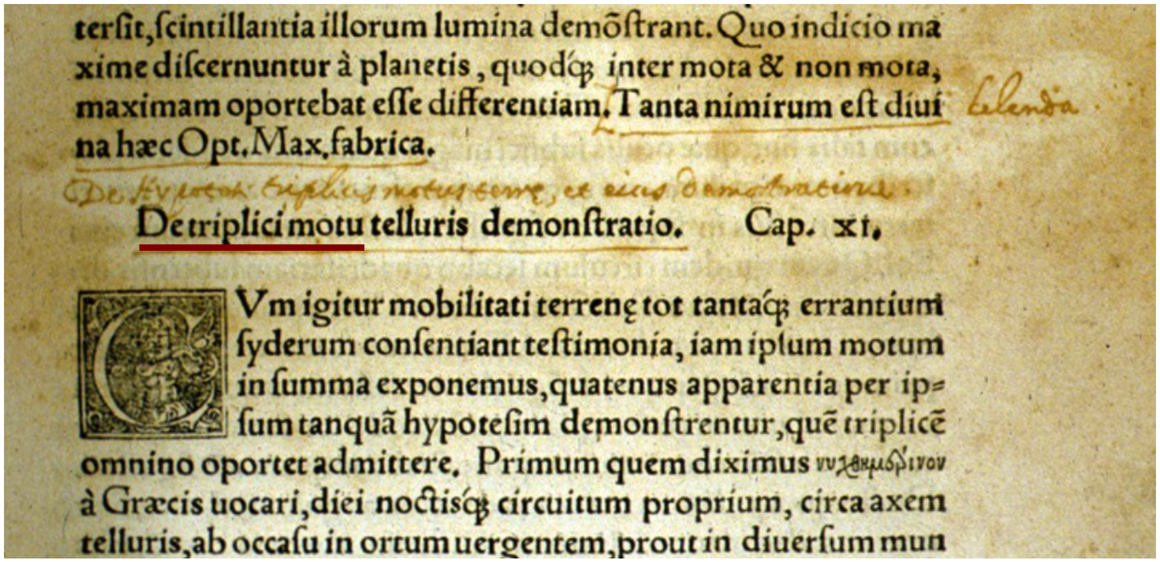
Galileo had become a Copernican and there's some evidence that this evolution in his belief happened early, but it was first enunciated in a letter to Kepler in 1597 ("....Like you, I accepted the Copernican position several years ago") but the letter to Catherine was his coming-out.

> "All our Fathers of the devout Convent of St. Mark feel that the letter contains many statements which seem presumptuous or suspect, as when it states that the words of Holy Scripture do not mean what they say; that in discussions about natural phenomena the authority of Scripture should rank last... [the followers of Galileo] were taking it upon themselves to expound the Holy Scripture according to their private lights and in a manner different from the common interpretation of the Fathers of the Church..." Letter to a member of the Inquisition.

> "All our Fathers of the devout Convent of St. Mark feel that the letter contains many statements which seem presumptuous or suspect, as when it states that the words of Holy Scripture do not mean what they say; that in discussions about natural phenomena the authority of Scripture should rank last... [the followers of Galileo] were taking it upon themselves to expound the Holy Scripture according to their private lights and in a manner different from the common interpretation of the Fathers of the Church..." Letter to a member of the Inquisition.

A council of advisors was established by Pope Paul V to review the theological aspects of Copernicanism. They reached conclusion on two issues: Does the Sun sit immobile? Does the Earth move?

- On the first, they concluded that to hold that the Sun was immobile and at the center of the solar system: "...foolish and absurd in philosophy, and formally heretical since it explicitly contradicts in many places the sense of Holy Scripture."
- On the second, that the Earth moves: "...receives the same judgment in philosophy and... in regard to theological truth it is at least erroneous in faith."



terſit,ſcintillantia illorum lumina demõſtrant. Quo indicio ma
xime diſcernuntur à planetis, quodq́ʒ inter mota & non mota,
maximam oportebat eſſe differentiam. Tanta nimirum eſt diui Delenda
na hæc Opt. Max. fabrica.
De Hypothesi triplicis motus terrae, et eius demonstratione.
De triplici motu telluris demonſtratio. Cap. xi.
CVm igitur mobilitati terrenę tot tantaq́ʒ errantium
ſyderum conſentiant teſtimonia, iam ipſum motum
in ſumma exponemus, quatenus apparentia per ip=
ſum tanquã hypoteſim demonſtrentur, quę triplicē
omnino oportet admittere. Primum quem diximus νυχθημέρινον
à Græcis uocari, diei noctisq́ʒ circuitum proprium, circa axem
telluris, ab occaſu in ortum uergentem, prout in diuerſum mun

Figure 10.7: At the end of the paragraph, Copernicus wrote: "Tanta nimirum est divina haec Opt. Max. Fabrica." (*"So vast, without any question, is the divine handiwork of the most excellent Almighty."*. That was to be eliminated. The beginning of the next chapter entitled, "*De hypothesi triplicis motus telluris eiusque demonstratione*" ("On the explication of the three-fold Motion of the Earth"was too much for the Inquisition and they suggested instead, "On the Hypothesis of the Three-fold Motion of the Earth and its Explication" which is written in above.)

This led to a banning of Copernicus' book until corrections were made. Figure 10.7 is a page of Revolutiononibus showing Inquisitor's corrections.

The consequences were not terribly significant. The Pope's advisors reported to him on February 24, 1616 and Paul asked the respected Robert Cardinal Bellarmine (who had previously defended Galileo's letter to the Grand Dutches to the Pope) to advise Galileo to not claim that Copernicanism as fact. This was in a written document signed by both that hypothetical discussions were okay. Galileo said, "okay." He had a nice meeting with the Pope and went home. Some years later a letter surfaced that suggested that Galileo had been admonished to not speak of Copernicanism even in hypothetical terms, but there's ample reason to suspect that this letter was fraudulent and created in order to create a legal case to silence or imprison Galileo.

It wasn't necessary. Galileo was perfectly capable of creating his own problems, all by himself.

10.1.3 Unforced Errors

Paul died in fall of 1616 and was replaced by Cardinal Scipione Caffarelli-Borghese as Pope Greggory XV, who then was succeeded in 1623 by Maffeo Barberini who became Pope Urban VIII. This was good, thought Galileo. Urban was a personal friend! Barbarini had supported him with poems and a stipend... even supporting Galileo's son.

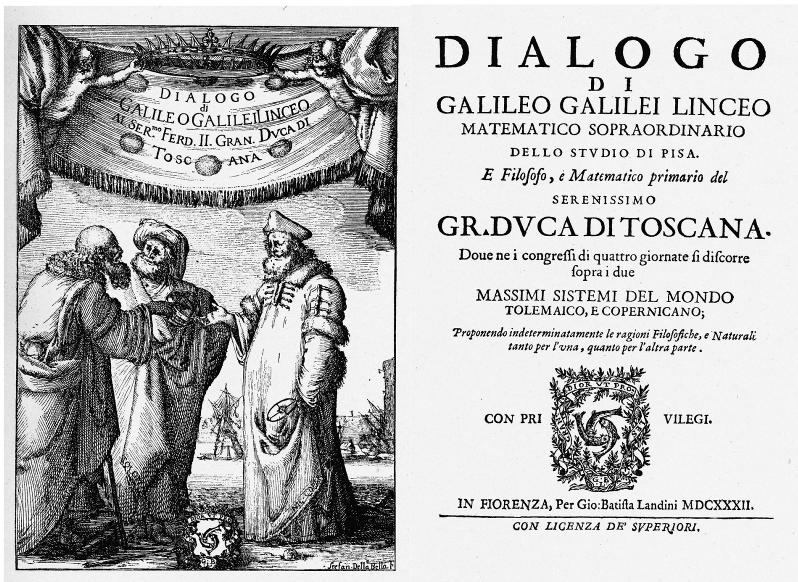


Figure 10.8: The cover shows from the left, Salviati (actually, Copernicus, who looks more like Galileo...in the next editions a more Copernicus-looking young person is depicted), Sagredo (actually, Ptolemy, hence the turban), and Simplicio (actually, Aristotle).

SIxteen years after his meeting with Bellarmine and Paul XV, Galileo finally went to print with his definitive publication on Cosmology, *Dialogue Concerning the Two Chief World Systems*. He chose to write it as a "dialog" among three people: Salviati (an actual friend of Galileo's) is the enlightened modern thinker who defends Copernicanism, Sagredo is an intelligent layperson who's slowly convinced by Salviati, and Simplicio who is an Aristotelian, whose name says it all about how he's portrayed. Figure 10.8 is the cover of Dialogo.

For some reason, Galileo puts Urban's own words to him in the mouth of Simplicio and that was his undoing. Within two months, the *Dialogue* was removed from all shops (it was in Italian, so laypeople could read it) and Galileo was summoned to Rome to stand trial for heresy.

Figure 10.9: ghome

Wait. *So Galileo's problems weren't necessarily because of his views of Copernicus?*

Glad you asked. *No. Galileo could surely have survived if he'd managed his "mouth" better than he did. Remember, after the original Inquisitional investigation, he was free to work for 16 more years with the friendly backing of multiple Popes. It was only when he ridiculed his former friend and patron that he was arrested. He brought it on himself.*

10.1.4 The End

Galileo was sentenced to house arrest for the balance of his life. Eventually, he was allowed to live in his villa outside of Florence where he was tended to by his son and other supporters. He slowly went blind and suffered many physical ailments, but was forbidden by the Pope Urban to be allowed to see a doctor in Florence. He still managed to put his Paduan work on motion into a new book which is also a dialog among the same three characters, *Discourses and Mathematical Demonstrations Relating to Two New Sciences*. But now the characters are representative of Galileo himself at different stages of his intellectual life. So no Pope is in the cast of that story.

Galileo died in 1642 within a few months of the birth of Isaac Newton. His burial was a mess, as he was not originally allowed to be buried in consecrated ground. Urban refused to allow it and was buried just outside of the famous Basilica of Santa Croce which includes many Renaissance heroes like Michelangelo and Machiavelli (and Galileo's famous namesake relative). Finally in 1737 he was reburied in the main room of the Basilica, but during the transfer three fingers and a tooth were taken from his body. One of those fingers is on display at the Museo Galileo in Florence.[6] Figure 10.10 represents Galileo still editorializing to the world from beyond it.

[6] Actually, after writing this I visited the museum again and found that they'd recovered the missing pieces and now all fingers are on grissly display. Terrific.

Figure 10.10: finger

The damage had been done to the Aristotelean picture of the solar system in a steady stream of observational blows from Tycho through Galileo. Our favorite Italian was famous and persuasive. His book on mechanics was the basis for further work in motion and with the adoption of algebra, Descartes' analytic geometry, and the decimal place...mathematics was brought to bear in Britain. But also damage to Italian science was serious and stifled for nearly 200 years after the Galileo embarrassment.

The Inquisition lifted the ban on Galileo's books in 1718! In 1741, the Pope authorized a publication of his works, somewhat edited. Not until 1758 was heliocentrism allowed in other publications, although Copernicus' books remained banned until 1835. Finally in 1992 Pope John Paul II publicly regretted the Galileo affair and the Church's handling of it.

10.2 The Apple Moment

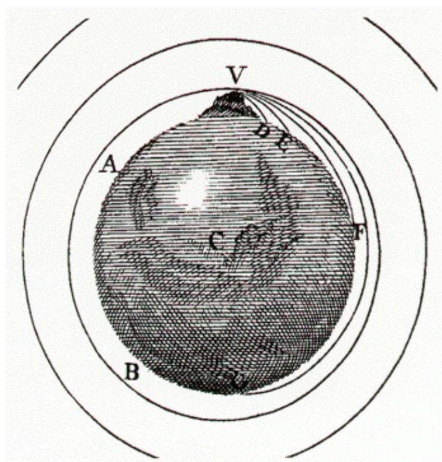


Figure 10.11: From the *Principia*. Perhaps the most wihimsical thing that Isaac Newton might ever have done!

Suppose you could go to a mountain and shoot a cannonball horizontally, like Galileo's table-top. If you were to increase the charge so that the cannonball is given more and more horizontal velocity. . . it would go further and further and eventually—it misses the ground. Figure 10.11 from the *Principia* is perhaps the most fanciful thing that Newton ever depicted. [7] He surmised that if given enough velocity the cannonball would continue to “fall" around and around: it would go into orbit. [8] This is a part of the famous idea that transformed physics forever. Yes, the Apple.

Box 10.1 The Apple Changed Everything

What I’m about to describe arguably changed not only natural science, but was the catalyist for the creation of the Enlightenment itself. What came from the Enlightenment, you ask? Everything we know today as how to think, how to govern, and the role of rationality in deciphering how the world works.

There’s no way to minimize the importance of Newton’s Gravitational law. It made precise predictions about a number of physical phenomena, which were tested and shown to be confirmed. The very idea that a model of the universe would be quantiative and that predictions would be worth testing was itself a new idea. Before Newton there was superstition. After Newton, there was science. Gravitation theory was the reason. The Aristotelian view disappeared. Ptolemey’s model disappeared. The solar system and the Sun’s rightful place was established, not to be unseated again.

During the 17th century the rules governing celestial objects were supposed to be different from those on Earth. Copernicus didn’t question this. Kepler hinted at a common set of rules. Galileo didn’t go there. And so Newton would have been taught in Cambridge eduction with still a strong hint of Aristotle. But somehow he–unlike anyone before him–imagined that there was only one set of rules that governed Earth-bound and celestial objects. And he, first among all, figured out how to make a model and test it.

When he was at the farm during the plague he had a number of remarkable ideas, among which one presumably came from the apple story. There was (and still is) an apple tree at his childhood home. He might indeed have watched one fall. The only account of an apple in this story comes from his first biographer, William Stukeley, about a dinner he and the great (now old) man enjoyed in 1726:

> "After dinner, the weather being warm, we went into the garden & drank tea under the shade of some apple tree; only he and myself...

[7] ...an adult. As a child he built intricate little kites on which he mounted burning candles. Then he launched them all one evening terrifying the townspeople. A Newton-prank.

[8] So this is important: “weightlessness” as a description of life in the International Space Station is a misnomer. Everything has weight as everything is still attracted to the Earth by its gravity, albeit at a slightly smaller value than the g that we experience on the ground. But if everything in the space station is falling together, it looks like nothing has weight. You would have to go much further than the Station’s orbit to be virtually free of a gravitational attraction.

> "Amid other discourse, he told me, he was just in the same situation, as when formerly the notion of gravitation came into his mind. Why should that apple always descend perpendicularly to the ground, thought he to himself; occasioned by the fall of an apple, as he sat in contemplative mood.

> "Why should it not go sideways, or upwards? But constantly to the Earth's centre? Assuredly the reason is, that the Earth draws it. There must be a drawing power in matter. And the sum of the drawing power in the matter of the Earth must be in the Earth's centre, not in any side of the Earth.

> "Therefore does this apple fall perpendicularly or towards the centre? If matter thus draws matter; it must be proportion of its quantity. Therefore the apple draws the Earth, as well as the Earth draws the apple."

There it is. That's the story. Further, if the Earth has "drawing power," maybe whatever it is about the Earth that draws an apple to the ground might also pull on the Moon. How much? Galileo thought that the force of the Earth on objects would be constant everywhere, but Newton guessed that it was not... that it should be diluted as one moves away from the Earth. He guessed (and later showed mathematically) that it could be presumed to be the center of the Earth.

He had to assume that the Moon moves in a (near) circle around the Earth. And he had to presume that the same rules governing objects moving in a circle on the Earth would hold for the Moon. He (privately in 1666) and Huygens (publicly in 1659) demonstrated mathematically that a centripetal acceleration must be pulling to the center and have the form v^2/R. Maybe that causal force—the centripetal force—is the one and same force that attracts the apple. Hmm?

Wait. *So did the apple hit his head as we're all taught?*

Glad you asked. *He never made mention of it, so it's another fable told by supporters later. Did Washington cut down the cherry tree? Does fruit always figure into Great Person Myths?*

With this set of ideas, he's way outside of the normal way of thinking. He's violating Aristotle's principles and he's violating his hero, Descartes' principle of reasoning from first principles. He's going to work out the consequences of such a guess, *without first identifying the original cause.*

> **Wait.** *Where did that original motion of the Moon come from? What's the Moon-cannon? Didn't he worry about that?*
>
> **Glad you asked.** *That's the brilliance of Newton. He chose not to start with a set of deductions from a principle, which is like a "why" question. He thought it appropriate to answer the "how" question and if satisfactory explanation resulted, then progress has been made. We understand this primordial velocity of the Moon as related to whatever early spinning was going on 4.5 By ago when the Earth and the Solar System were formed from rotating dust.*

Here's what he wrote of his summer many years later:

> "I began to think of gravity extending to the orb of the Moon & (having found out how to estimate the force with which [a] globe revolving within a sphere presses the surface of the sphere) from **Kepler's rule of the periodical times of the Planets being in sesquialterate proportion** of their distances from the center of their Orbs, I deduced that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve & thereby compared the Moon in her Orb with the force of gravity at the surface of the Earth & found them answer **pretty nearly.**"

What's he saying here?[9]

[9] I'll bet you didn't know that "sesquialterate" means "...in a ratio of one and a half to one." Neither did I.

10.2.1 How to Support a Moon In Its Orbit

What's coming in the next three pages is the longest mathematical story in the whole book. But it's the very definition of "game changer" and you know enough to be able to enjoy it with me! Let's develop the most important physics chapter in the book of western science.

Newton's model asserted the following:

1. The force of gravity that pulls on the Moon is the same force that pulls on an apple. That force is the centripetal force that causes the Moon to move in a circle. (He's using his first law here.)

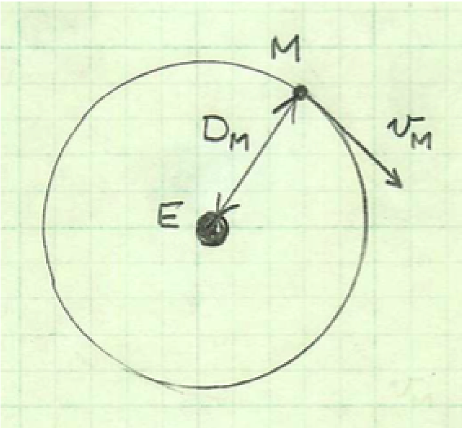


Figure 10.12: earthmoonsetup

2. The Moon's circular trip can be approximated by an infinite number of straight, tangential paths, which are pulled back by gravity to the circle. Over and over.
3. That "pull" back...is "falling" and that's what he models for the Moon and compares to falling near the Earth.
4. He uses Kepler's 3rd Law as a guide, and shows that this implies that the force of gravity on an object decreases by the inverse of the distance away from the source. Kepler introduced it for the planets around the Sun, but Newton extends the idea to also work for the Moon around the Earth.

Pencil 10.1.

The setup is shown in Fig. 10.12. I'll use R_E for the radius of the Earth and D_M to mean the distance from the center of the Earth to the center of the Moon, and v_M to be the speed of the Moon in its orbit.

What's the Moon's Centripetal Acceleration?

He needed to find the centripetal acceleration of the Moon using what he knows about the Moon's motion (it takes a month to go around the Earth) and Kepler's Third Law: $T^2 \propto D_M^3$.

$$a_C(\mathrm{M}) = \frac{v_M^2}{D_M}$$

How does he actually know v_M? Well, that's the easy part. The speed is the distance traveled—the circumference of the Moon's orbit—divided by the time that it takes to go around, its period (1 month), which we'll call T. So the speed is

$$v_M = \frac{\text{circumference}}{\text{period}} = \frac{2\pi D_M}{T}.$$

which we can substitute into the centripetal acceleration relation:

$$v_M = \frac{2\pi D_M}{T}$$

to get

$$a_C(\mathrm{M}) = \frac{4\pi^2 D_M^2}{T^2 D_M}$$
$$a_C(\mathrm{M}) = \frac{4\pi^2 D_M}{T^2} \tag{10.1}$$

Then he used Kepler's rule from Eq. 9.1

$$T^2 \propto D_M^3$$
$$T^2 = kD_M^3 \tag{10.2}$$

where I've inserted a constant of proportionality, k.[10]

Substituting the period of the Moon's orbit into Eq. 10.1, we get:

$$a_C(\mathrm{M}) = \frac{4\pi^2 D_M}{kD_M^3}$$
$$a_C(\mathrm{M}) = \frac{4\pi^2}{k}\frac{1}{D_M^2}. \tag{10.3}$$

This is huge. He's demonstrated that the centripetal acceleration (and hence the force) for the orbiting Moon would vary as the inverse-square of the distance that the Moon is from the Earth. This is something that everyone suspected, but nobody figured out before this moment. It was buried in Kepler's law all the time. Keep it in mind.

[10] This is a big leap from Kepler! The constant k for the planets, that Kepler assumed, would be completely different from that assumed for the Earth-Moon relationship. Newton picks out the idea and applies it in a direction that Kepler never intended.

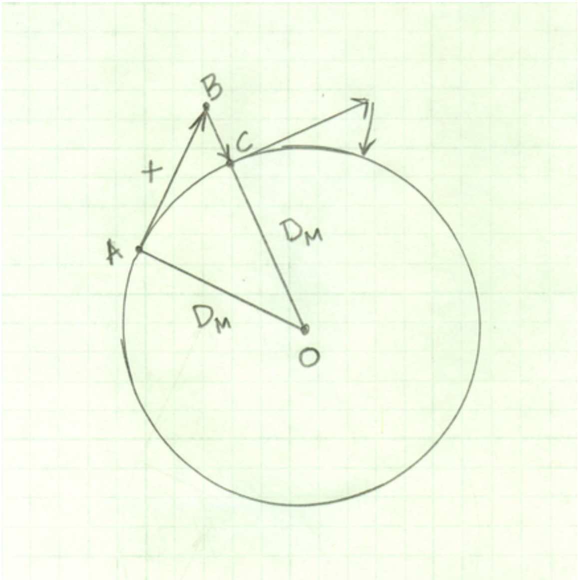


Figure 10.13: moonstraight

How Newton Confirmed His Model of Gravity

The second bold (in both sense of the word "bold"!) phrase refers to a calculation using his model which is embodied in the simple diagram in Fig. 10.13. The Moon is traveling in a circle, but thinking like a calculus-inventor, Newton imagined that this circular orbit is really an infinite number of little tugs across the intervening space and to the center of the Earth. The Moon goes in a straight line from point A to point B at speed v_M and then is pulled to the center—to point C—by the Earth's gravity. This pull happens in some time interval, which we'll take to be 1 second. Then it goes in *another* straight line and is pulled back.

In essence he asked how far is it tugged? It's as if the Moon is at point B and in one second "falls" —FALLS!—to point C...just like the apple falls. How far is that $\overline{BC}$ distance?

Actually, that triangle OAB is a right triangle and the hypotenuse is $\overline{BC} + D_M$.

- He knows D_M to be about $60 \times R_E$ and he can calculate the $\overline{AB}$ leg by knowing how fast the Moon is going (v_M) and using the regular formula for speed: $\overline{AB} = v_M t$.
- We have the form of the Moon's speed and he used an average month to be 27.3 days, which is $T = 2,360,000$ seconds.
- We'll use a modern value for the distance from the Earth to the Moon, where he used an ancient result that D_M is about $60 \times R_E$.[11]

[11] He also made some other mistakes and approximations. For example, he assumed that a mile is 5,000 ft. But hey, when you're inventing a whole discipline, sometimes you gotta cut corners.

Let's put all of this together and calculate that speed:

$$v_M = \frac{\text{circumference of Moon's orbit}}{\text{1 month}} = \frac{2\pi D_M}{T} = \frac{(2\pi) \times 1,031,400,000}{2,360,000} = 2,746 \text{ ft/sec.}$$

We found the speed above, so the distance $\overline{AB} = x(\text{Moon})(1 \text{ second}) = 2,746$ ft along that tangent, for 1 second.

$\overline{ABO}$ is a right triangle and we can find all three legs: we just found $\overline{AB}$ and we know that $\overline{AO}$ is D_M. The leg $\overline{BO}$ is really $D_M + \overline{BC}$. We can use Pythagoras' Theorem to find $\overline{BO}$ and therefore, to isolate $\overline{BC}$. This is the "fall" of the Moon through that 1 second! The result of the calculation is:

$$\overline{BC} = 1/20\text{th of an inch} = 0.004167 \text{ ft.}$$

Let me repeat this, because it's important: Newton calculated that in 1 second, our Moon "falls" to Earth by 0.004167 ft.

Let's put this together with the discovery that the Earth's gravity is diluted by the square of the distance, then the distance that an apple would fall on Earth can be related to the distance that the Moon falls away from the Earth!

$$\begin{aligned} x(\text{apple}) &= x(\text{Moon})\frac{D_M^2}{R_E^2} \\ (\text{apple}) &= 0.004167(60)^2 = 15 \text{ feet} \end{aligned} \tag{10.4}$$

Now, if an apple falls through a distance of 15 feet in 1 second, what's the acceleration due to gravity for it from what Newton called "Galileo's Theorem"?

$$\begin{aligned} x &= 1/2 g t^2 \\ g &= 2x/t^2 = 2(15)/1^2 = 30 \text{ ft/s}^2 \end{aligned} \tag{10.5}$$

This must have been satisfying: 30 ft/s^{2}is pretty close to what he knew little* g* to be, 32 ft/s^2. So indeed, "pretty nearly."

He's (you've!) done an amazing thing!

He's measured the acceleration of gravity on the Earth parameters from the Moon!

Think about that.

As he said in Book III of *Principia* where he summarized this earlier work, "And heavy bodies do actually descend to the earth with this very force." Understated as the most important few lines of work in the history of science!

The same physical theories govern motion on Earth and the cosmos. *Key Concept 19*

Now Think Big!

In looking back to the days on the farm in 1666 where he first tried out this idea, he used some incorrect numbers and was still working out the mathematics but even though he never published his results. . . he worked off-and-on for years. But eventually, he started to think about the actual force that the Earth would exert.

> If we stuck in the Moon's mass, then we'd arrive at a formula for the gravitational force, similarly diluted by that same factor.

That is, with his Second law $F = ma_C$ and the centripetal acceleration we would find that the force of attraction by the Earth on the Moon is:

$$F_{EM} = K\frac{m_M}{D_M^2}$$

where I've rolled all of the constants ($\frac{4\pi^2}{k}$) into one big K for tidiness.

Now we remember Newton's *Third* law that says that if you push on me, I'll push back on you. So what's the force of attraction *for* the Earth, *from* the Moon? The same. We can't just replace the Moon's mass with the Earth's mass, since that would change the force. The only way for to work is that whatever this force is it has to be symmetrical between the Moon and Earth. So it hast to look like this:

$$F_{EM} \propto \frac{m_M m_E}{D_M^2}.$$

The proportionality constant has a name...and a long history to this day.

So this is rather remarkable. For the first time the physics of the Moon is convincingly shown to be identical to physics on the Earth. But he's not done.

10.2.2 Universal Gravitation

With the idea of centripetal force and his third law of motion—and confirmation using the Moon's parameters—he's connected the Moon's orbital motion to regular Earth-bound circular motion and connected the Moon's motion to objects falling on the Earth. The nature of these forces seem to not care about the objects that cause and experience them—like Galileo insisted—indeed, they're not different forces, but manifestations of a single force. He's connected the Earth to the Moon: one theory.

Later he analyzed the motions of the moons of Jupiter and Saturn and eventually comets and showed that they obeyed Kepler's Law, like Kepler's planets and now, like the Moon. Suddenly, the whole solar system, including the moons of all planets hung together in a single mathematical system. One set of rules for the whole of our visible universe and our terrestrial home.[12]

At this point, Newton makes a breathtaking leap. He assumes that a gravitational attraction exists between *any two objects with mass.* Right now you are being attracted by the Earth, but also by the Sun, and the Moon, and Jupiter, and by the banana on your desk that you're saving for lunch . All objects in the Universe attract one another according to the following universal rule, which we played with in our mathematics review way back in Chapter 2. The force acting on 1, because of the pull of 2 is:[13]

Figure 10.14 shows the situation. Some object #1 with a mass M_1 attracts some other object #2, also with a mass M_2...and *visa versa.* This attraction is along a line connecting their centers which are $R_{1,2}$ apart. This is called the Universal Law of Gravitation and the constant of proportionality, G is Newton's Constant or the Gravitational Constant. The force of attractoin *on* 1 due to 2 is $F_{1,2}$ while the force of attraction *on* 2 due to 1 is $F_{2,1}$. From Newton's Third law? They're equal.

An interesting fact about this equation is that it can only be solved exactly for two objects. Add a third object—or a fourth, or fifth, etc—and the equation cannot be solved. Rather it is necessary to solve it *approximately* and after Newton people became very skilled at doing very complicated approximation calculations called "perturbations."

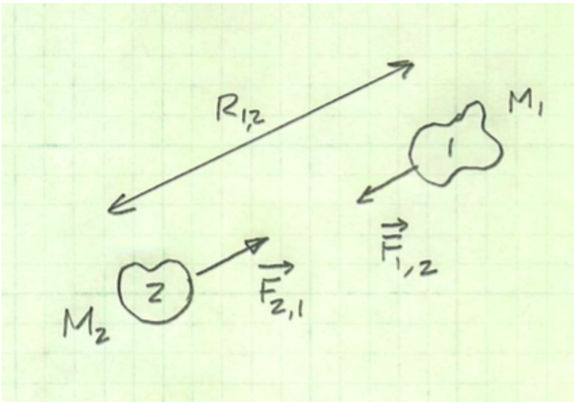


Figure 10.14: force12

[12] Aristotle has now left the building. Never to return.

[13] Of course there is the force acting on 2 because of its attraction by 1, $F_{2,1}$. And, they're equal.

Equation: **Newton's Gravitational Law.**

$$F = G\frac{Mm}{R^2}$$

Constant of nature: **Newton's Gravitational Constant.**

$G = 6.67 \times 10^{-11}\ \mathrm{m^3kg^{-1}m^{-2}}$

The Gravitation Constant

The constant of proportionality, G, is very tiny and not known well.[14] Newton had no estimate for its value, rather he worked in ratios of forces but it was measured in a laboratory by the very odd Henry Cavendish about a century later. It is a fundamental constant of nature. It just is. There's no deriving it. Were it different by a little, our world would be very different.

[14] the uncertainty on that number is 0.021 out of 6.67, or about 0.3%. For a fundamental constant of nature, that's not very precise.

The gravitational force is very weak and characterized by a single constant of nature. *Key Concept 20*

Little *g* Again

Now we can understand Galileo's results from a modern point of view. With the Universal Law of Gravitation and Newton's Second Law, the acceleration due to a gravitating body can be isolated from Newton's rule by finding the "a" and the "m." To see what I mean, look at Fig. 10.15.

Pencil 10.2.

Keep in mind Newton's simple second law:

$$F = ma.$$

Place your apple on the ground—notice that it's distance from the center of the Earth is, R_E. Let's calculate the force on that little apple with mass m due to the big Earth, with mass M_E. Newton taught us that the force between them is

$$F = G\frac{M_E m}{R_E^2}$$

Now isolate the little m outside of the other terms:

$$F = m\left(G\frac{M_E}{R_E^2}\right) = ma$$

and can you see that we've discovered an acceleration buried in the middle term by recognizing $F = ma$ in it:

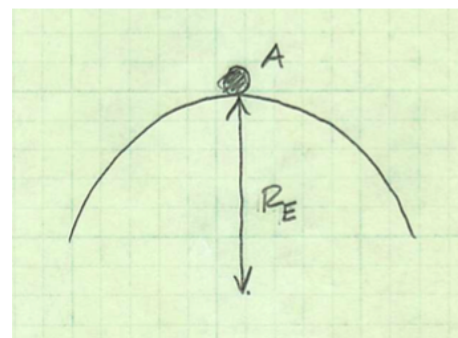
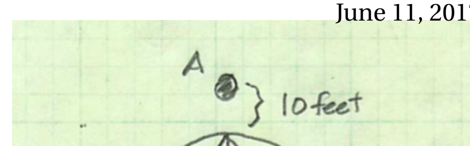


Figure 10.15: An apple sitting on the ground a distance R_E from the center of the Earth.



$$a = G\frac{M_E}{R_E^2}.$$

Since this situation is an apple on the surface of the Earth, what we've really found is a derivation for Galileo's g! So we can just identify:

$$g = G\frac{M_E}{R_E^2}. \tag{10.6}$$

You Do It 10.1. Calcuating g

or copy the solution

Using the following parameters

- $G = 6.67 \times 10^{-11}$ m^{3}kg^{-1}m^{-2}.
- $R_E = 6.37 \times 10^6$ m.
- $M_E = 5.97 \times 10^{24}$ kg.

and using Eq. 10.6, calculate g.

Did you get 9.8 m/s^2? Look familiar? So, that's where our weight comes from. The Earth attracts us with a force that's $F = mg$, which is a constant—on the surface of the Earth. When you step on a scale, it pushes back and is calibrated to read back how much spring-force is required to balance your weight.

Example 10.1

Constant g?

Question : Let's calculate the acceleration due to the Earth's attraction for an apple in a tree, 10 feet above the Earth's surface as shown in Fig. 10.16.

Solution:

$$a(\text{tree}) = G\frac{M}{(R_E + 10\text{ feet})^2}.$$

Ask Mr. Google and you'll find that that R_E is more than 17 million feet and you can readily see that the little 10 foot addition is minuscule. This calls for the use of the approximations that we listed in Section 2.6, in particular Eq. 2.20 which looks like our function with $10/R_E$ playing the role of x. Here's how this works. Manipulate the equation above so that the denominator is like $1+$ something...and we do that by dividing out R_E. Then "something" is a very small number and we can use Eq. 2.20. So let's do that: Now just use the first two terms of the approximation from Eq. 2.20:

$$\frac{1}{(R_E+10)^2} = \frac{1}{(1+10/R_E)^2}\frac{1}{R_E^2} \text{ now the approximation:}$$

$$\approx \frac{1}{R_E^2}[1-2(10/R_E)] = \frac{1}{R_E^2}(1-20/17,000,000)$$

$$\approx \frac{1}{R_E^2}(1-0.0000012) \text{ So, from this the acceleration at 10 feet:}$$

$$a(\text{tree}) = GM\frac{1}{R_E^2}(0.99999882) = 0.99999882g \tag{10.7}$$

So for all practical purposes $a(\text{tree}) = g$.

Teaching Moment!

Wait. *But we've been saying that Galileo showed that the acceleration due to gravity is a constant. Now you're saying that it depends on how far away one is? Which is it?*

Glad you asked. *Yes. Galileo's g is really not a constant, but it varies very little...even for large distances above the Earth. So for all practical purposes, we can consider it to be a constant. In fact, let's calculate that for the highest (above the Earth) that you've ever been:*

Here we have a situation that's going to repeat itself over and over in the history of physics. Galileo said that the acceleration due to gravity was constant. Then along came Newton who showed that this wasn't right in the strictest sense: that the acceleration due to gravity varies as you move away from the gravitating object. Over and over we'll have to grapple with a question that in this case, looks like:

Was Galileo wrong?

For almost half a century, Galileo's discovery was considered a fact of nature. But then it was shown to be the case only in a restricted domain...in this case, when you're close to the surface of the Earth. As you'll see in the next steps, there's really no circumstance that you or any of us (except for a handful of astronauts) will ever experience in which Galileo was incorrect.

The scenario runs like this: First, Theory A explains a feature of the world and establishes a fact of nature and a mathematical Model that uses it. Then along comes Theory B that shows that the facts and the models of Theory A are not strictly correct. Yet if the facts of Theory A and the models in Theory A are included in the facts and the models of Theory B *within a domain of experience that's smaller than the domain that Theory A describes*, then we'd say two things: First, Theory B is more inclusive than Theory A . It explains more about the universe. And second, Theory A is still the case when applied to the restricted domain of experience that's a subset of the domain of Theory B.

In this case, Newton's theory (B) explains gravitation everywhere. Galileo's theory explains gravitation only in the region near the surface of the Earth (A). We still happily—and reliably—use a constant g in the design of any structure or vehicle, for example. Keep this notion in mind, since Newton's gravitation law...will become a "Theory A" in a few hundred years at the hand of Albert Einstein!

You Do It 10.2. title

or copy the solution

Let's say that an airplane is 5 miles above the surface of the Earth. If I drop my delicious snack on the floor of the cabin, what acceleration due to the Earth's gravity would it experience compared with if I had dropped it on the ground? The radius of the Earth, which is $R_E = 3960$ miles.

10.3 Three Problems for Newton

The successes of Newton's model for gravitation were many and astounding. It's sometimes said that the Enlightenment was a direct result of the success of the naturalistic approach to explaining the world. Here are some of what his theory of gravitation demonstrated:

- He showed that the inverse-square rule for gravitation explained Kepler's Laws, that they would accommodate circular, elliptical, and parabolic orbits. Famously, Hayley's Comet was discovered and the predictions that Newton's friend made were based on Newton's rules. He simply assumed that the comet's path was elliptical (but squashed) around the Sun as its focus and could then use Newtons' Gravitation law. He was right...Halley's Comet's path takes it all the way past Neptune before it starts coming back towards the Sun. It's a 76 year round trip.
- The Earth's axis wobbles a tiny bit and Newton explained that, the precision of the equinoxes.
- He explained the tides as a feature of the Moon's attraction for the ocean water closest to it as opposed to the water on the other side of the Earth from the Moon.
- The Earth should not be a perfect sphere since it's not an absolutely rigid mass. Because it rotates material closest to the axis through the poles (near the poles) would feel a different gravitational force due to the material inside of its radius from material furthest away from the axis of rotation (equator). So there should be a measurable difference in the gravitational attraction at different longitudes and this stimulated heroic teams of explorers who traveled very far north with pendulums to make measurements of g everywhere they could. Newton's explanation worked.
- And of course his model explained all of the observed orbital motions of the known planets, a concept that was not even thought possible, or even desirable while Newton was a child. He determined the relative masses of the planets and the Sun.

Of course in addition, he had other unparalleled (including to this day) achievements:

- He properly conceived of the idea of momentum and completely describe motion and dynamics.
- He correctly conceived of the theory of colors as mixing together to make white, in contradiction to the prevailing views led by Descartes.
- He invented and the pioneered the use of calculus.

By the time he died in 1726, magic was gone. Subservience to Aristotle was gone. Everyone believed that...everything could be known. The very essence of the Enlightenment period in western history.

But there were issues that were more philosophical that required his attention.

10.3.1 Action At A Distance

Two distinct camps developed in physics. While a dominant belief in naturalism now reigned in Europe, the French followed the lead of Descartes while the British remained loyal to Newton. But everyone agreed that the actual mechanism of gravity was problematic. In a letter he wrote:

"It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact...that one body may act upon another at a distance through a vacuum, without the mediation of anything else...is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it."

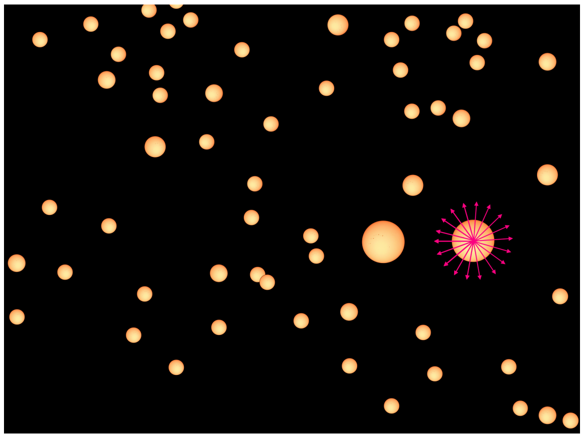


Figure 10.17: newtoninfiniteU

Everyone agreed but two camps developed on proper process. The Continental view was that until you can enunciate the mechanism of gravity and then reason deductively from that, you're not doing acceptable science. By contrast, there was the British view—and the one that we all follow today—that what's important is that if it works, that's good enough. In many ways, Newton differentiates the contrast between **why** a phenomenon occurs and **how** it occurs...and an answer to *how* can be a mathematical model. He famously said: "I feign no hypotheses." which even is its own Latin catch-phrase, "*hypotheses non fingo*" (Google it!).

The bar to making progress that Descartes set up (the Continental view) is too high. One should "hypothesize" (I'd say model-build) and deduce empirical observables, test them and then refine your model. Then you've turned science into a Process that improves on its conclusions. Eventually—and gravity is a good example—one might find an acceptable why...but until that, how is good enough and makes progress possible.

10.3.2 Stability of the Universe—Cosmology

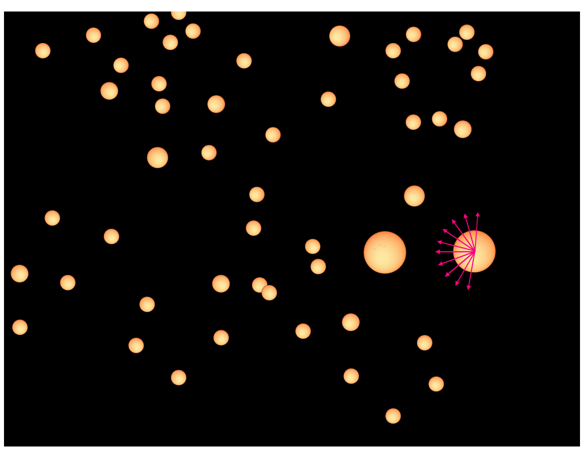


Figure 10.18: newtonfiniteU

Newton wasn't shy about how to apply his model. As described above, there were plenty of terrestrial and astronomical applications that were predicted and tested positive. But what about the whole enchilada? What about the whole Universe?

He recognized quickly that he faced a puzzle even more prickly than Action at a Distance. He couldn't explain the improbability of why we're here at all. Here's the problem, which I can form as a question:

Is the Universe finite or infinite? His theory seems to suggest that the Universe must be infinite, with an infinite number of stars (all anyone knew about were planets and stars...no galaxies). Imagine this enormous space filled with stars, each of which is attracting every other object in it, and is in turn being attracted by every other object. Figure 10.17 is a cartoon of such a situation. That one star is being pulled on by everyone...and the fact that the Gravitation law varies like $1/R^2$ means that there is an influence from all objects, all the way to infinity.

If the Universe had an edge, then Fig. 10.18 would crudely be the story. Notice now that our target star is being pulled to the left and there's no balancing set of forces to the right. That should start our star accelerating which would then pull on other stars differently as it moves and they'd start to accelerate—the end result would be a huge collapse of everything on top of itself. Since we're here, this hasn't happened and so the Universe is infinite. That's the argument, but it's flawed...or at least highly improbable.

Suppose the Universe is infinite and this incredibly delicate balance is at work. A butterfly could cause the whole thing to collapse, much less Jupiter orbiting the Sun. That is, the nature of his Gravitational law is such that the delicate balance that holds everything just right...has to be absolutely perfect. That seems improbable.

Newton had a famous correspondence with the leading theologian in Britain, Richard Bentley in 1692. Bentley was erudite and familiar with science and Newton took him seriously. He wrote to the reverend:

> "As to your first query, it seems to me that if the matter of our sun and planets and all the matter in the universe were evenly scattered throughout all the heavens, and every particle had an innate gravity toward all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of the space would, by its gravity, tend toward all the matter on the inside, and by consequence, fall down into the middle of the whole space and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space. And thus might the sun and fixed stars be formed, supposing the matter were of a lucid nature. But how the matter should divide itself into two sorts, and that part of it which is fit to compose a shining body should fall down into one mass and make a sun and the rest which is fit to compose an opaque body should coalesce, not into one great body, like the shining matter, but into many little ones; or if the sun at first were an opaque body like the planets, or the planets lucid bodies like the sun, how he alone would be changed into a shining body whilst all they continue opaque, or all they be changed into opaque ones whilst he remains unchanged, I do not think explicable by mere natural causes, but am forced to ascribe it to the counsel and contrivance of a voluntary Agent."

And again,

> The reason why matter evenly scattered through a finite space would convene in the midst you conceive the same with me, but that there should be a central particle so accurately placed in the middle as to be always equally attracted on all sides, and thereby continue without motion, seems to me a supposition as fully as hard as to make the sharpest needle stand upright on its point upon a looking glass. For if the very mathematical center of the central particle be not accurately in the very mathematical center of the attractive power of the whole mass, the particle will not be attracted equally on both sides. And much harder it is to suppose all the particles in an infinite space should be so accurately poised one among another as to stand still in a perfect equilibrium. For I reckon this as hard as to make, not one needle only, but an infinite number of them (so many as there are particles in an infinite space) stand accurately poised upon their points. Yet I grant it possible, at least by a divine power; and if they were once to be placed, I agree with you that they would continue in that posture without motion forever, unless put into new motion by the same power. When, therefore. I said that matter evenly spread through all space would convene by its gravity into one or more great masses, I understand it of matter not resting in an accurate poise.

> … a mathematician will tell you that if a body stood in equilibrio between any two equal and contrary attracting infinite forces, and if to either of these forces you add any new finite attracting force, that new force, howsoever little, will destroy their equilibrium and put the body into the same motion into which it would put it were those two contrary equal forces but finite or even none at all; so that in this case the two equal infinities, by the addition of a finite to either of them, become unequal in our ways of reckoning; and after these ways we must reckon, if from the considerations of infinities we would always draw true conclusions.

Newton's solution to this delicate balance was an appeal to God. God's job is holding everything in place. While today we don't do that in science, we'll see this particular problem come back for our other full-blown Scientific Hero and we'll find that Einstein provided a different explanation.

Cosmology

This is an important point in the history of physics. It's the beginning of quantitative, predictive science of the universe. This subfield of physics and astrophysics is called Cosmology: the study of the whole universe.[15]

[15] Cosmogony is another similar word that describes the origin of the universe. Formerly, thought to be outside of the provence of science. Now...a regular part of physics. We tend to have expanded the word Cosmology to include origins.

definition Cosmology

definition Cosmogony

While Kepler came close, after all, he provided a formula that was descriptive of how the planets move. We understand Kepler's law now as a logical (meaning: algebraic) consequence of Newton's Gravitation...s0 it was eventually appreciated to be derivative.

Here we have Newton using an abstract (meaning: using a mathematical formula) explanation to describe the entire universe. His equation's form insists that a gravitational force only goes to zero at infinity, so no matter how far away two objects with mass are situated, they will still attract one another. This presented a problem that needed explanation: his model was predictive, but not complete. His approach to this level of incompletion was to give up and require a deity. Our approach is to leave it open as a problem remaining to be solved. In that sense, we're more Newtonian ("no hypothesis") than he was!

10.3.3 Absolute Space and Time

Newton's mechanics led to big questions that required speculation about space and time...that is, **S**pace and **T**ime! He asked himself questions like this (although not this particular one). Suppose the universe consists of only four objects: you, your friend, a rope, and a knife. You and your friend are connected by a rope. Are you stationary or are you *rotating* around the center point of the rope? Remember, the universe is empty but for you two. How could you tell?

This is sticky matter of relative motion. If I were to ask if you were moving linearly with respect to your friend, you could tell me that because you'd see your friend approach, pass you, and recede. Which would be really sad. (By the way, your friend would see exactly the same thing except in the other direction.) So you might not agree about who is moving and who is stationary, but you'd have no trouble believing that relative motion exists between you.

But rotation is a different matter and this question is specifically about an accelerated "frame of reference" since in order to rotate about that center point, a centripetal force through the rope would be required and so an acceleration is at work. Well, one of you has a knife and if you cut the rope one of two things might happen. Nothing! In which case you'd conclude that you were not rotating because the other thing that might happen could be that you'd immediately begin to separate meaning that: you had each been orbiting the center and that when the rope no longer connected you, you'd start straight line motion in accordance with Newton's First law.

The question is...if you are rotating in this situation, *with respect to what* are you rotating? Newton felt that he needed an absolute measure for inertia and acceleration and he chose Space, with a capital S. To Newton, space was a thing. Take everything out of the universe and space will still be there acting as an absolute coordinate system. All motion, constant velocity and accelerated, can be described mathematically with respect to this absolute coordinate system. So he said. Newton also insisted that there is an absolute clock...absolute Time, with a capital T.

Needless to say, there was also the Continental point of view, championed by Leibnitz who said that space was defined by the relative positions of things. Take away the things and there is no space...it's completely relative...to stuff.

This argument is going to come back to haunt us a few more times before we reach the 21st century! But the important thing is that nobody talked like this; nobody theorized scientifically about the universe before Newton.

10.4 Gravitational Energy

I

Chapter 11

Charges and Magnets

Charge It!

William Gilbert, 1591, by unknown artist, Colchester and Ipswich Museum Service: Colchester Collection.

William Gilbert, 1544-1603

““In the discovery of hidden things and the investigation of hidden causes, stronger reasons are obtained from sure experiments and demonstrated arguments than from probable conjectures and the opinions of philosophical speculators of the common sort....”” *De Magnete (1600)*

Arguably, a rival as one of the first experimental physicists , was the physician for Queen Elizabeth I and briefly James I, a contemporary of Galileo’s and Kepler’s. Gilbert was a civil servant in service to the Royal Navy and the monarchy as physician, but also privately practiced research using a recognizable scientific method. He was the first to systematically study magnetism and determine that the Earth was a magnet. He also studied static electricity and was a Copernican, applying his understanding of magnetic force as a possible source of the attraction that the Sun presumably applied to the Earth. His major work, *De Magnete*, was read all over Europe and had great influence for two centuries.

11.1 Goals of this chapter:

- Understand:
  - How to calculate forces on charges using Coulomb's Law
  - How to determine the direction of magnetic force due to a current
  - How to calculate the numbers of charged particles corresponding to a total charge in Coulombs
- Appreciate:
  - The effects that parallel currents have on one another
  - The similarities between a loop of current and a magnet
  - That the Earth is a magnet and what are the north and south magnetic poles
- Be familiar with:
  - The role of Gilbert in the history of physics
  - The contributions of Franklin to electricity and magnetism

11.2 A Little of Gilbert

William Gilbert was born into a middle class family during a period of relative calm following the turbulence of Henry VIII's reign and the future civil war in the early 17th century.[1] During that century, Britain reorganized its entire society with the confiscation of Church properties, secularization of public education, and subsequent increase in general wealth and a middle class. London quadrupled in size and became an international center of trade and hub for British exploration of the Americas and East India. Gilbert's father was a benefactor, serving as a prominent lawyer and benefited from these changes. Gilbert's life was spent in civil service as both physician and a private scientist, outside of a university. His was an illustrious career, subsequently serving as President of the Royal College of Physicians.

[1] "Relative calm" might be a "relative" phrase. There was the small issue of the Armada and the attempted invasion of England by Spain in 1588.

He was born a year after Copernicus died (coinciding with the publication of *De Revolutionibus*) and a year before the Catholic Church's Council of Trent was seated to try to recover after the simultaneous loss of influence in England (with Henry VIII) and Northern Europe (with Luther). Little is known of Gilbert's early years, but for his education in his family's home town of Colchester and subsequent study at St John's College, Cambridge. There he took his medical training and absorbed, like everyone, the Aristotelianism that passed for advanced learning...and like his famous contemporaries, he rejected it in order to make intellectual room for a more reliable understanding of nature. Like Galileo, he taught Aristotle for a while as a "Master" at Cambridge, but like Galileo after him, he became convinced that the only route to understanding nature was through observation and experiment—not through appeal to authority and ancient writings. Unlike Galileo, Britain was now free of the Church and its authority over what could and could not be written about.

Gilbert died in 1603 of bubonic plague and many of his personal effects were burned so we lack a detailed understanding of his life and research. We know that he never married and of course know of his professional success. His definitive book, *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure* (On the Magnet, Magnetic Bodies, and the Great Magnet of the Earth, 1600, aka *De Magnete*) is enlightening and shows perhaps the first true experimental scientist at work. A relative (his brother?) later collected unpublished notes into a postumous book, *De Mundo*, which shows a thinker, like Copernicus, with one foot in Medieval traditions and the other in modern times. Ironically, fire during the London "Great" plague of 1665 destroyed Glibert's London home as well as the library of the Royal College of Physicians where Gilbert had bequeathed many of his papers, so the second recurrence of that terrible disease in London spawned Newton's imaginative scientific genius but destroyed the last bits of Gilbert's history.

At some point, Gilbert became interested in both magnetism and electricity. Both were similarly magical and both magnetic and electrical effects seem to create forces without any intervening medium—without contact. Little was known but much mysticism was invested in magnetism, especially. Rather than relying on hocus-pocus descriptions or ancient Greek authority, Gilbert set about to systematically study what materials were specifically influenced by magnets and what by static electricity. The Greeks and Chinese had discovered "Lodestones" which we know today to be a naturally magnetized[2] mineral and both cultures had found that small iron magnets, when floated on water, would align themselves in a particular direction: North.

Gilbert's book is a remarkable document for its orderly discussion of heretofore unstudied phenomena. He described past views on the phenomena and criticized them, often on the basis of his own experiments. He concluded that magnetism and electricity were different phenomena and categorized materials by their magnetic and electrical properties. How did he do that? He found that heat would dissipate electrical attraction but not magnetic. This is a conclusion, then, based on controlled experimentation.[3]

What he's most famous for having figured out was that the Earth is a magnet, suggesting that its core was made of iron. He came up with this idea and then tested it in a variety of ways. He actually constructed spheres of lodestone and then did experiments on them with tiny magnets, determining how the angle of dip of a compass needle would vary with longitude on his test spheres, comparing that with actual compasses on the Earth at various locations. He even carved indentations into his test spheres to simulate valleys and mountains on Earth and subsequent changes to the compass needles.

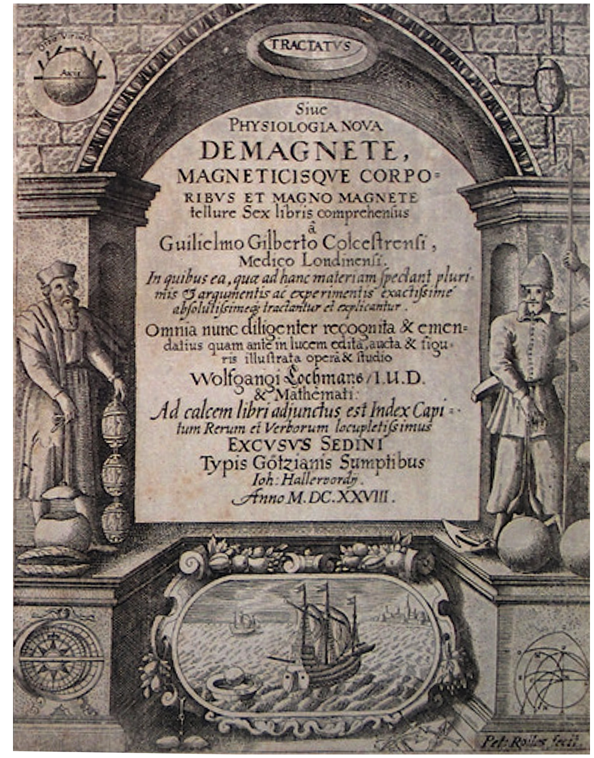


Figure 11.1: demagnete1628

[2] In fact, just how lodestones are magnetized has been a matter of considerable research. The Earth's inherent magnetic field is too small to have caused it. Rather, the generally accepted explanation is that lightening in early times was responsible, a suggested supported by the fact that lodestones are to be found near the surface of the Earth.

[3] Unknown to him, from the late 19th century we understand that very high heat can also cause magnetized iron to lose that property.

Figure 11.3: littleguymagnet

Figure 11.2: earthmag

Having in hand an invisible force associated with the Earth, he then anticipated Kepler and proposed that it was magnetism that was responsible for the attraction of the Earth to the Sun—he was a Copernican and because he was in England, he felt little antagonism for these beliefs. Galileo was criticized by the Inquisition for having praised the "perverse and quibbling heretic" and he himself did experiments on lodestones with Gilbert's work as a guide.

While Gilbert taught mathematics, he held an entirely different view of mathematics and natural philosophy. Copernicus' book was too complicated for him and he maintained that mathematicians' jobs were to describe the phenomena of the heavens and not to reach conclusions about physical cause. He explicitly excluded magnetism as a topic worthy of mathematical explanation and was largely silent on the subjects of motion, let alone a mathematical description of the sort that Galileo championed.

And Gilbert also worked in electricity, especially on the "amber" effect—static electricity. . . he knew that if one rubbed an amber rod, that it would then acquire the ability to attract paper. He invented the electroscope, which had useful scientific applications into the 20th century. Finally, it is to Gilbert that we owe the word "electricity." A very modern physicist for his time.

> As for the causes of magnetic movements, referred to in the schools of philosophers to the four elements and to prime qualities, these we leave for roaches and moths to prey upon. Gilbert, *De Magnete*, Book II, Chapter 3.

11.3 Magnetism

Puzzles regarding electricity and magnetism dominated the first two decades of the 19th century and it was in resolving this exciting confusion that led us to radio, TV, cell phones, and all that's good about life. The notion that one's hair might stand on end when combed, or that certain stones might move one another without touching had been disconcerting for millennia. Indeed, by about 1600, magnetism especially was thought to be magic and associated with illnesses and odd cures. But this period marked the beginning of thoughtful reasoning and experimentation in the natural sciences and we saw that William

Gilbert literally "wrote the book" on magnetism, both describing his own researches and specifically debunking one-by-one the various ways in which magnets were thought to be helpful or harmful.

11.3.1 Magnets

Let's review the simple facts about magnets that we've all known since childhood. Magnets come in one "package" a bar of typically iron in which the two ends are special: one is by convention called "North" and the other "South," inheriting their names from a compass's use as a pointer to the north geographic pole. The pole of the magnet that points to the north geographic pole is dubbed "North." The other fact that we all know is that you can cut a magnet into pieces all day long and you'll never isolate a separate north or south magnetic "monopole." "Never" is a tough word to use in science and even though an atomic explanation for magnetism in bulk is now accepted, there are theories of cosmology which require the creation of separate, magnetic "charges" which have not been discovered yet.

If you come near to a target north pole with a second magnet's north pole, the target will be forced away. Likewise, if you approach with the opposite, south pole the magnets will be attracted. Endless fun. But what's actually going on and how to characterize the phenomenon? Clearly there are forces created between magnetic poles with no material connection between them and this had been a puzzle since the Greeks created the first refrigerator communication device. (Okay. I made that up, but the Greeks first noted magnetism and named it for one of two towns called Magnesia—there is dispute about which.) In any case, the order is clear:

Magnetized objects have both "north" and "south" poles which always seem to come in pairs and cannot be isolated. *Key Observation 10*

Magnets exert forces on Magnets: Like poles repell and opposite poles attract. *Key Observation 11*

11.4 Electricity: Poor Richard

Figure 11.4: Benjamin Franklin, 1785 at 79 years old (portrait by Joseph-Siffred Duplessis in the National Portrait Gallery in Washington, D.C.).

There were various attempts to study both electric and magnetic forces, although the idea that there might be a connection between them had not occurred to many people, it was clear to those who came after Newton that some sort of Action at a Distance seemed to be going on. In 1747, Ben Franklin (1706-1790) took time out of his busy social life in London and Paris to experiment and set the course for our understanding of electricity. This man traversed the Atlantic Ocean eight times in his life[4] and his wit and romantic appeal was familiar in Britain and especially France. While he came to it naturally, the Enlightenment fever which by that time had overtaken Paris certainly influenced both Franklin's politics of individualism and self-determination, but also his innate inquisitiveness to try to understand the "how's" of the natural world. We all know him in scientific lore for a foolhardy experiment in the rain. But he studied many things and it was for electricity where he had his most significant impact in physics.

[4] Do you suppose he got frequent-shipper miles? Instead of a berth below the water's surface, maybe he upgraded. Surely he didn't stand in line to board.

Figure 11.5: benkite

Franklin's electrical experiments began in 1747 with a glass rod and silk cloth and after enough experimentation he was inspired enough to write a book of his numerous experiments in electricity four years later. It was here that he guessed that lightening was a manifestation of electricity and he proposed a method to prove it.

While not necessarily due entirely to Franklin, by the time he was done with his scientific work, it was generally acknowledged that:

Electrical Charges exert forces on Electrical Charges: like charges repel and opposite charges attract. Electrical charges can be isolated from one another. *Key Observation 12*

11.4.1 When Your Tool is a Hammer...

...then everything is a nail. Isn't that how the saying goes?

Franklin's Model for electricity held sway for nearly a century: he asserted that electricity was a single fluid, just like everyone thought heat was a fluid. Neither was a bad guess, although both were wrong, but one works with the tools that you have and the mathematics of fluid flow and thermodynamics were being developed at that time and became quite sophisticated by the middle of the 19th century. It's interesting that while electric and magnetic fields are not fluids, the mathematics used to describe them is very similar to that of fluids.[5] So barking up the wrong tree can sometimes still be a worthwhile experience (unless, of course, you're an actual dog).

[5] This is because both are time-dependent vector flows.

Box 11.1 The Kite

So did he actually fly that kite? Probably not the way the legend holds, certainly not as the Currier and Ives etching in Fig. 11.5 suggests. First, it had already been done, so he wasn't first. And, second, he would have died had he done this in an actual lightening storm as happened to Georg Wilhelm Richmann in St. Petersburg in 1753. No, what Franklin probably did in his demonstration in Philadelphia in 1752 was fly a kite high enough to collect some charge along the wet kite string from a storm cloud and then suggest that lightening's current was from the same source. And then he surely waddled for cover.

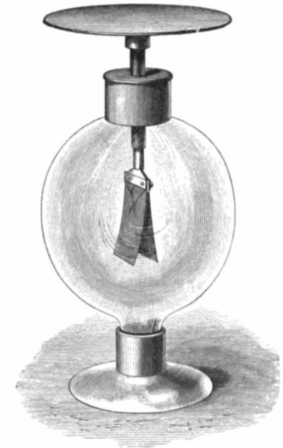


Figure 11.6: An electroscope is a delicate instrument that can detect the presence of electrically charged objects. By touching the object to the plate on the top, the gold leaf strips separate. The more separation, the more the charge and so quantitative measurements could be made. (Clipart courtesy FCIT)

Franklin's "fluid" fit into the porous volumes of all substances. If two bodies had the same amount of fluid, they were neutral. But the fluid could ooze from one to the other and a body which had an excess of the electrical fluid would repel another similarly "full" body, and attract a one who's porous material contained less. How this fluid exerted its force was not understood and collectively people threw up their hands and attributed it to the same mysterious Action at a Distance attributed to Newton's Gravitational Law. And like in gravity, the force was instantly felt across otherwise empty space.

Because of this to and fro flowing, he was led to postulate a conservation law, that the total amount of electricity was conserved and only just moved from one place to another. This imaginative and in pictorial process describes one of the most important principles in Particle Physics: that electrical charge is conserved. Remember our definition of Conserved Quantities on page 6.3:

Net Electrical Charge is Conserved. *Key Concept 21*

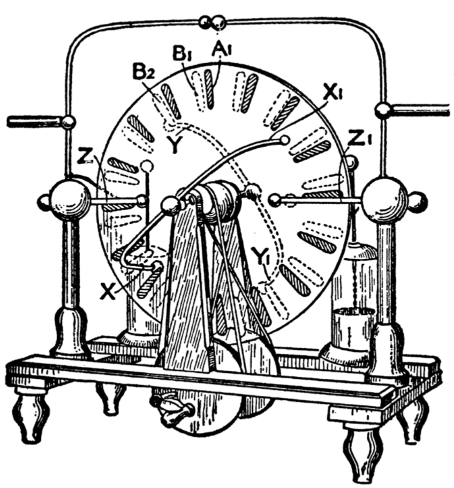


Figure 11.7: A Wimshurst machine is an example of a 19th century frictional charging device. Notice the Leyden Jar in the background for accumulating the developed charge. (Clipart courtesy FCIT)

Just like our discussion of the conservation of energy and momentum, this conservation law is related to a symmetry, but a complicated quantum mechanical symmetry for which there is no classical description. It is one of the most fundamental laws of Nature.

This movement of charge he called a *current*, the word we use today, in homage to that fluid-inspiration. Of course, we also speak of current *flowing* from one terminal to another, which in turn implies a *direction* for current. From Franklin we also get the language of "positive" and "negative" charges. He thought that the body which had been *relieved* of its fluid was then "negative" and that the fluid itself was an *addition*, hence "positive" in a body in which it's accumulated. Of course, today we understand that flow in circuits to be the nearly free flow of electrons within a copper matrix

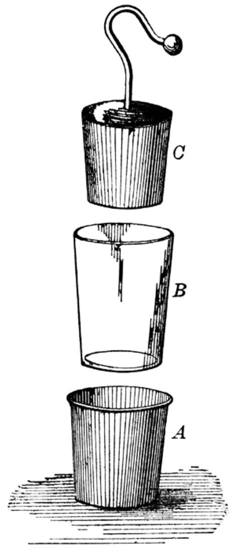


Figure 11.8: A Leyden Jar is essentially two metal cans which fit into one another separated by an insulating can, like glass. This was a forerunner of a capacitor and can be used to store electrical charge. (Clipart courtesy FCIT)

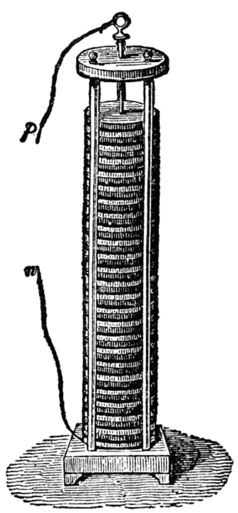


Figure 11.9: An early Voltaic pile—"battery" to you and me.

which only loosely binds them to a particular nucleus. Today we describe the flow in Franklin's terms, from positive to negative, but the more dominant actual charge motion is from the negative terminal to the positive.

Franklin named things and electrical matters were no exception, "I feel a Want of Terms here and doubt much whether I shall be able to make this intelligible." He was inventing a science. In addition to his terms "positive" and "negative," we use "plus" and "minus" and the symbols + and – and the words "charge" and "battery."

11.4.2 Science Follows Technology

Franklin's research and that of others who quickly followed was due to their imaginations, but also some major technological inventions. One of the first was the improvement of the electroscope as shown in Fig. 11.6. Using a nearly evacuated bell jar and thin gold-leaf tabs, small amounts of charge could be deposited and relative amounts determined by the separation of the leaves. This was the primary device for measuring electrical charge throughout the 18th and 19th centuries. But, it's hard to generate measurable amounts of electric charges when all you've got is muscle-power, glass rods, and silk and furthermore, and it's delicate work since it's hard to keep charges isolated as they will quickly bleed away into even slightly humid air.

But this situation was made much easier when in 1745, at the University of Leyden, a glass jar was surrounded by a metal can with another metal can on the inside: a cylindrical, metal-glass-metal sandwich as shown in Fig. 11.8. It was found that this "Leyden jar" could be "charged" by successively adding increments of charge which would stay on the cans and grow to even dangerous amounts. Today we call such an arrangement a "capacitor" after the fact that a capacity of charge can be stored on it. The Leyden Jar made it possible for researchers to store and then use charges in their investigations without having to create them over and over. Franklin was "filling" a Leyden jar in his kite experiment (which you can see beside his foot in Fig. 11.5). But charge still needed to be created by frictional means—either by rubbing two materials together by hand, or by more efficient Wimshurst devices, large circular plates with brushes which could be turned by hand as in Fig. 11.7

Eel Be Sorry

One of the most important devices is derived from a fish story. Well. Not exactly a fish. Eels were imported to Europe from South American and Africa as a delicacy in the mid-1700s and their strange talent of generating enough voltage to charge a Leyden jar made for endless fascination in 18th century Europe. In 1786, the physician Luigi Galvani (1737-1798) was dining in a Bologna restaurant as a thunderstorm was approaching and he noticed that frogs legs hanging near a iron railing were involuntarily twitching. He drew the conclusion that this was the same effect as that of the eel, and studied it. By him and others, the biological effects of this "Galvanism" were ghoulishly investigated by discharging Leyden jars in the bodies of guillotined prisoners and their severed limbs. You can only imagine how startling it must have been to see the dead move and public demonstrations made the phenomenon a popular topic. (This "cutting edge" research is known as one of the inspirations for Mary Shelley's creation of *Frankenstein*.) Another Italian, Galvani's friend Alessandro Volta noticed that two dissimilar metals in salt water would produce the eel's effect inorganically and in 1800 eventually constructed a sandwich of copper and iron/zinc disks separated by cloth soaked in salt solution. This "Voltaic Pile," shown in Fig. 11.9, the forerunner of the battery, made it possible for natural scientists to create large amounts of charge, which they could store in their Leyden Jars and to further study their motion—currents.

11.5 Charge It!

Amounts of charge were very difficult to determine and of course their actual nature was unknown. Charles Coulomb was a French mathematician-military engineer who was exceedingly proficient with instruments and in 1784 succeeded in building a very delicate "torsional balance" of precisely machined metal balls. These he would attach across a lightweight beam, which in turn was suspended from at its center by a thin wire. Gently pushing the balls in a plane around the center of the brace would twist the vertical wire and that twist could be precisely measured and converted into the force of the push. By charging the balls and charging fixed, external balls to act as the push (or pull), Coulomb was able to determine the amount of force exerted by given amounts of charge, and the degree to which this force changed when the charges were separated. Rather surprisingly, he found this force has a familiar form:

$$F_e = k\frac{Q_1 Q_2}{R^2} \tag{11.1}$$

which today we call Coulomb's Law, the force between two charge collections Q_1 and Q_2 which are separated by a distance, R as shown in Fig. 11.1.[6] The value of the constant in front, k depends on the units that are used. Using Coulombs as the unit for charge, along with meters for the distance, k can be determined in MKS units to be:

$$k = 9 \times 10^9 \frac{\text{N m}^2}{\text{C}^2} \tag{11.2}$$

Equation: Coulomb's law.
$F_e = k\frac{Q_1 Q_2}{R^2}$

[6] The capital letter Q is standardly used for electrical charge. I'll sometimes use q to refer to the charge of a sub-atomic particle when it matters.

Constant of nature: Coulomb's constant.
$k = 9 \times 10^9$ N m^2 C^{-2}

Definition: Coulomb.
The standard unit of electrical charge, C. A single C is an enormous amount of electricity.

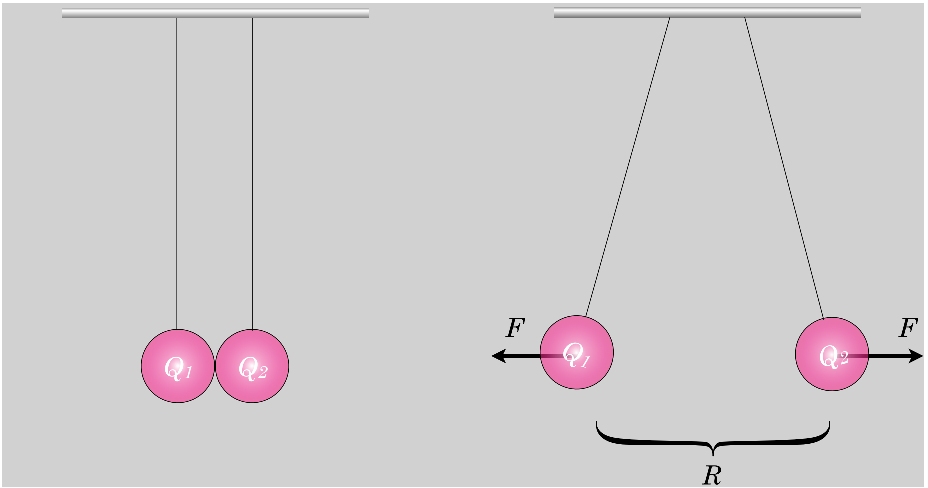


Figure 11.10: The repulsion of two identically charged pith balls can be described by Coulomb's Law.

Notice that Equation 11.1 suggests that the electrical force depends directly on the magnitudes of the charges and the inverse square of the distance that charges are separated. The constant of proportionality, k is arbitrary until the units of charge and distance are specified. Notice that the sign of charges matters here. Maintaining Franklin's designation of plus and minus, if we make one of the charges plus and the other minus, then the force in the equation is algebraically negative. So, this means we have to be very careful about what force is on "whom." Here's how to think of it: If we draw an arrow from charge Q_1 to Q_2, then Equation 11.1 represents the force on Q_1 due to Q_2 and if the charges are opposite in sign, then they are attracted to one another and the negative sign in the force equation is just telling you that the force is in the opposite direction of the arrow you just drew. The unit for charge that's commonly used is the *Coulomb*.

It's actually defined in terms of the unit of current, the Ampère (A): that is, the amount of charge flowing in a current of 1 A during 1 second is a Coulomb. In fact, this is the definition of current: $I = Q/t$. We'll meet M. Ampère (A) below.

A Coulomb of charge is an ENORMOUS amount of charge. The definition of it is historical and has almost no bearing on real-life. But we're stuck with it. A couple of examples will make this clear.

Example 11.1

A Single Coulomb.

Question : How much force is there between two positive charges, each of 1 C, separated by 1 meter?

Solution: This is not a hard problem, but it sets the scale for just how outrageous a Coulomb of charge actually is. Of course we'll use Coulomb's law, namely Eq. 13.1:

$$F = k\frac{QQ}{R^2} = 9 \times 10^9 \frac{1 \times 1}{1^2} = 9 \times 10^9 \text{ Newtons} \tag{11.3}$$

Are you kidding me? This is 2 billion (with a B) pounds. If this were the weight of some object, it would have a mass of $W/g = 9 \times 10^9/9.8 = 918$ billion kg. It is the weight of 10 Nimitz-class super-aircraft carriers—the attraction of the Earth to 10 aircraft carriers stacked one on top of the other.

Just let that sink in...with the emphasis on "sink." Let's make it worse. How many Coulombs is an electrically charged elementary particle?

11.6 Electromagnetism

There was a lot of patient electrical experimentation during the early 1800s. And, as sometimes happens, the historic breakthrough came by accident. Accounts differ, but those of us who lecture for a living like to think this version is true! As the story goes, the Danish natural scientist Hans Christian Oersted (1777-1851) was giving a public lecture in 1819 about the heat given off by electrical currents when he made a discovery. (Wires get hot, as you know from playing with batteries.) He generated his currents using a Volta's Pile and for some reason had a compass on his lecture bench which was pointing to the North as compasses will do. His current-carrying wire was above it, along the North-South direction as well. When he turned on his current, the compass needle jumped and pointed to the West! With nothing up his sleeve, he had demonstrated a brand new connection between currents (charge) and magnetism:

Currents exert a force on Magnets perpendicular to the current. *Key Observation 13*

Figure 11.11: Hans Christian Oersted (1777-1851)

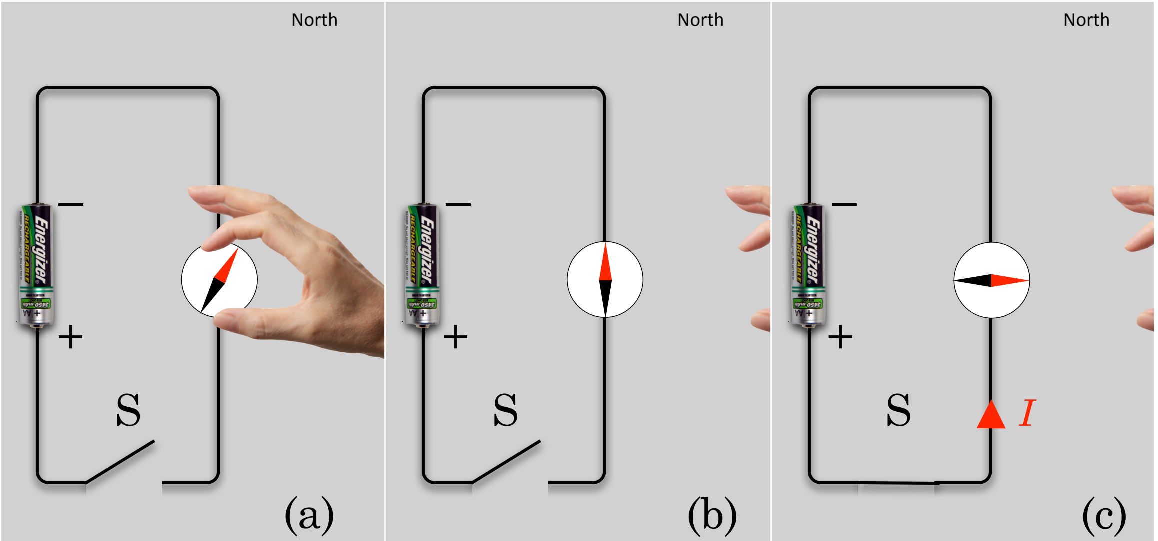


Figure 11.12: Shown is a circuit with a switch (S) and a battery with the positive pole down. When the switch would be closed, current would flow counterclockwise. In (a) a disembodied hand brings a compass near to the wire. In (b) the compass recognizes that magnetic North is to the top of the figure and it points that way...as is the job of a compass. Then in (c) S is closed and current flows as shown and Oersted's discovery was that the compass "forgets" all about the puny Earth's magnetic field and responds to the current.

in

Figure 11.13: Mariè-Andre Ampère (1775-1836)

I don't know how composed he remained during the demonstration because this would have been quite a shock (no pun intended!). He finished his lecture and then went into feverish experiment-mode studying the effect during the ensuing weeks. He found a number of surprising results. For example, the compass was not attracted *to* the wire. Newton's gravitational attraction would have led one to expect that two objects which are the source of a force should be attracted directly towards one another—like the gravitational force. No, the compass needle did something unusual: it twisted.

Oersted interpreted this as a magnetic influence of the same nature as that of the earth (after all, it's a compass) that was radiating outward from the wire. Here, he was wrong and a more careful examination of the effect—which is hard, because the force is very weak even for very large currents—shows that the magnetic influence is not radial, *but circular, around the wire* as shown in Fig. 11.14. When the current flows, it's there. Turn off the current, it disappears. Reverse the direction of the current, and the other pole of the compass is attracted. There is a rule-of-thumb (again, no pun intended!) on how to identify the direction of the magnetic influence from a wire by using your right hand: the first of a handful (sorry! no pun again!) of "Right Hand Rules."

With your right thumb in the direction of a current, your fingers will curl in the direction of the magnetic influence. Key Concept 22

Oersted wrote about this effect and went on tour, demonstrating it around Europe causing an enormous stir among natural scientists. That there was a connection between electricity and magnetism was now undeniable and this led to a number of speculations about how and under what conditions these connections might hold. The idea came more naturally to Oersted than to others because he had a particular religious belief, "*Naturphilosophie*," that held that all of Nature is connected and so he was open to unification of all natural phenomena. Since the traditional way to make magnets move is with another magnet, it was apparent that what Oersted had done is demonstrate that:

Currents create a magnetic force, in a circular pattern around the current. Key Observation 14

Oersted's phenomenon was demonstrated at the *Académie des Sciences* in Paris on September 11, 1820 and in the audience was Mariè Ampère, a troubled, French mathematician. He was precocious in mathematics (and many other things)—especially calculus and the refinements of Newton's physics. But a melancholy man, he was sad most of his life. His father had been executed during the French Revolution and although Ampère was happily married he lost his wife in 1803 while he was away at a new teaching position. Separation from her and their young son had been especially hard, as she'd already been ill when he departed and so his guilt sent him into a gloom which stayed with him for the rest of his life. He married again, disastrously, but separated from his wife after only a year and a half with their daughter under his custody.

During the week that followed that momentous autumn lecture, He managed to work out and measure Orested's effect, but he went further. He quickly constructed a delicate, current-carrying coil which could be suspended with the axis of the coil horizontal. When current flowed and a bar magnet was brought near, the coil pivoted and was attracted to or repelled from the magnet. The coil has a left and right side, like poles—you can again use your right hand and wrap your fingers in the direction of the current and your thumb will now point in the direction of a "North" pole-like direction. So the coil of current behaves as if it were itself a bar magnet! From this he quickly hypothesized that all magnetism is due to "molecular" current coils and that there must be circular currents in the center of the Earth to produce the Earth's magnetic field.

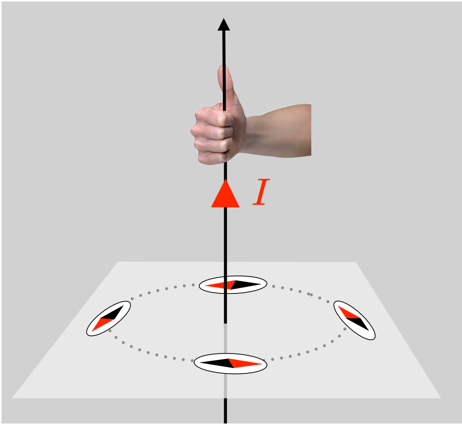


Figure 11.14: Oersted's very careful experimentation demonstrated that a compass needle responds in a circular pattern around a wire carrying current. The direction of the north pole of the compass needle can be found by using one of many "Right Hand Rules." Here if your thumb points in the direction of the current, then your fingers curl around the wire and point in the direction of the north pole of the compass. We"ll reinterpret this in terms of the Magnetic Field in the next chapter.

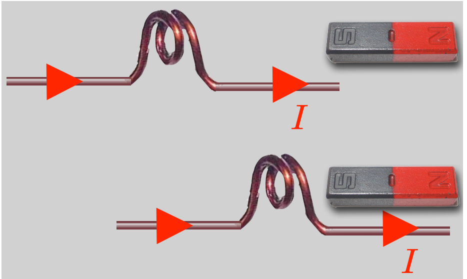


Figure 11.15: A current carrying coil behaves like a magnet and is attracted or repelled depending on the direction of the current.

A coil of wire carrying a current behaves like a bar magnet. *Key Observation 15*

Not content, he went further: he further reasoned that if currents cause magnetism, then one current ought to influence another current, just like magnets influence other magnets.in[0] By producing very high currents he showed that two parallel wires experienced a force between them—attractive when the currents are parallel and repulsive when opposite:

[0] We now define the unit of current, the Ampère, or "Amp" (A) by the amount of force between two wires. "The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2 x 10-7 newton per meter of length." (from the National Institute of Standards and Technology)

Currents exert a magnetic force on other parallel currents causing them to attract (if the currents are aligned) or separate (if the currents are anti-aligned) . *Key Observation 16*

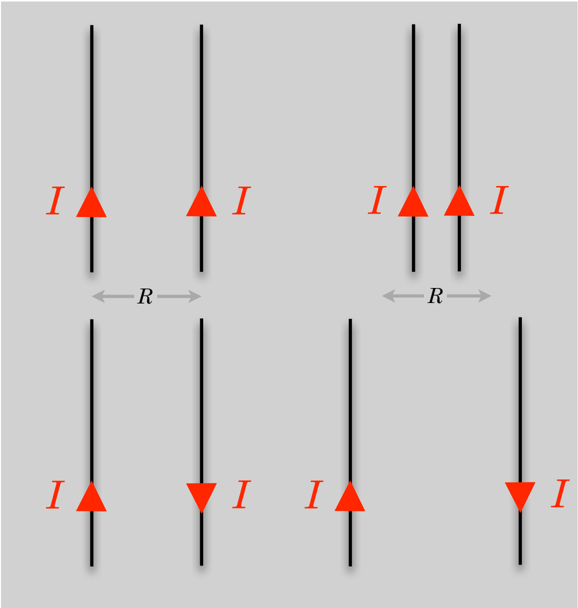


Figure 11.16: Currents in parallel wires will create an attractive force (left) when the currents are parallel and repulsive force (right) when antiparallel.

With his mathematical skills, he was able to work out the force on one wire due to another, which is related to what we call Ampère's Law. He did this work extraordinarily fast and used a procedure that's not generally used today. By November, he had a complete Model for magnetism: magnetism as due to the collective action of little, molecular currents. He came to this through experimentation as well as his mathematics.[7]

[7] This is quite remarkable and close to the modern interpretation of magnetism as due to the collective effects of the magnetism inherent in the electrons in ferromagnetic materials. This is a quantum mechanical phenomenon, but Ampère's guess was inspired.

Ampère's reputation was secure and he was able to create a course in "Electrodynamics" in which he furthered his researches in the relationship between electricity and magnetism with both inspired mathematics and impressively precise experimentation. While his place in the history of physics was secure, his relationships with his surviving children were unpleasant. Consistent with his dark outlook, his son, a successful historian, and he were always at violent odds and his daughter was an alcoholic who lived with her husband in Ampère's home. This was a life not made in heaven, one of constant turmoil, including frequent visits from the police.

11.6.1 Modern Ideas

Let's remind ourselves of modern interpretations of these ideas—things you all know. The historical discussion above was about regular-sized objects that were electrically charged or magnetized. The idea of atoms was still highly speculative, much less electrically charged particles.

The sub-atomic objects in nature are either: positively charged, negatively charged, or neutral (zero charge). As we will soon see, the total electrical charge of any macroscopic object (your hair in winter and the comb that charged it?) can't be just any value, but instead is a multiple of an apparently special amount.

When electricity was first studied—Franklin's time—there was no idea of an electron or a proton and their particular charges. Rather "stuff" was charged: everyday-sized things that charge of which might be pretty large. The unit of charge is the "Coulomb," (abbreviation is "C") named for our Charles Coulomb of previous pages. He studied regular sized objects and so they had lots of charge. But when the 20th century came around and people started to study atomic-sized objects it became apparent that the most natural value of charge was not Coulombs...but nano-nano Coulombs: the fundamental electrical charge is that of single electrons and protons. It was a couple of decades into the 20th century when the value of the discrete, fundamental charge was finally measured as the charge of the proton, which in magnitude is the same as the charge of the electron. Here it is. Lot's of zeros:

$$e = 1.602176487(40) \times 10^{-19} \text{ C}$$

Constant of nature: **Fundamental Unit of Charge.**

$e = 1.602176487(40) \times 10^{-19}$ C

Let's look at this number. There's a lot of information in it—beyond just its value. The first thing you notice is that it's tiny, as I warned: $e = 0.0000000000000000001602176487(40)$ C. The second thing you notice is that there are lots of non-zero numbers to the right of the decimal point! That is, it's measured to very high precision which should suggest to you years and years of patient, careful measurements by lots of hard working people. The third thing you notice is the (40) at the end of that string of numbers which are the only digits in the string of 12 that are considered to be slightly uncertain. So the precision to which this number has been measured is $0.00000000040/1.60217648740 = 2.49 \times 10^{-8}$ percent! You would correctly conclude that knowing this number must be a pretty important thing. It's precision is much higher than that of the other fundamental physical constant that we saw earlier, the Gravitational Constant, G.

Electrical charge appears as multiples of a fixed, fundamental amount called e. *Key Concept 23*

We use the symbol e to represent the value of this fundamental amount of charge (so we don't have to carry around all of the decimal points!). So in this way, we would say that the electrical charge of an electron is: $Q_e = -e$ C and the electrical charge of a proton is: $Q_p = +e$ C. You might wonder why they are the same value? We don't know. Franklin had a 50-50 chance of guessing which direction current flowed, and he got it wrong since the objects that move in a wire—that are an electric current—are negative electrons. So we're stuck with his assignment. [8]

Definition: Charge of a proton.
$Q_p = +e$ C

[8] Notice that here a "Key Concept" is also a "Key Question"!

Why is there a fundamental unit of electrical charge? *Key Question 9*

We'd love to understand this and it's an object of speculation and research.

Well, guess what. Just when we thought we'd nailed down the fundamental unit of charge in Nature...quarks happened. We'll talk a lot about these fundamental entities later, but they have the surprising property of electrical charges which are not multiples of e, but rational fractions, namely $Q_{\text{quarks}} \pm 2/3e$ or $\pm 1/3e$. Quarks go together to make many of the particles that we understand to be composite by adding their fractional charges to make ...whole multiples of $\pm e$. We take e as the fundamental quantity primarily out of habit and because the electron has this fundamental, whole unit.

Definition: Charge of an electron.
$Q_e = -e$ C

Definition: Charge of a quark.
$Q_q = \pm 1/3e$ or $\pm 2/3e$ C

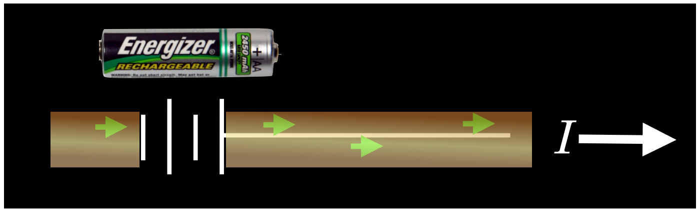


Figure 11.17: A little segment of a circuit. The arrows represent the opposite direction of the actual electrons!

Earlier I noted that we define the unit of electrical current as that necessary to cause a particular force to separate two parallel currents. But most of us think in practical terms, that an electrical current is a measure of how much charge passes a given point, during a given time.

Equation: Electrical current.
$I = \Delta Q/\Delta t$

$$I \equiv \frac{\Delta Q}{\Delta t} \tag{11.4}$$

So 1A = 1C/1s. You've probably all played with batteries and wires in school and you know that if you set up a little circuit with a AA battery some wire and a bulb that after some trial and error you can make it light. We say that the current flows from the positive terminal of the battery to the negative terminal, but that's Ben's mistake: it's actually the electrons that move in a wire and so the current is in the opposite direction from the actual charge motion. We've learned to live with that and we call positive current *from* the positive terminal of a battery.

An electrical current is rate in time at which electrical charge moves. *Key Concept 24*

Example 11.2

An elementary particle's charge.

Question : Okay. So a Coulomb creates a lot of force. Just how much electricity is there in an electron? Or a proton? That is, what is the charge of an electron in Coulombs and how many electrons are there in a single Coulomb?

Solution: The elementary unit of charge is the magnitude of the charge of an electron which is identical to that of a proton...they only differ by their sign. A proton is positive and an electron is negative. That elementary unit is called e and we'll make considerable use of it later. For now,

$$e = 1.6 \times 10^{-19} \text{ C}$$

Not very much. So now the number of electrons?

$$\text{number of electrons in 1 C} = \frac{1}{1.6 \times 10^{-19} \text{ C/electron}} = 6.25 \times 10^{18} \text{electrons/C.}$$

That's more interesting, but maybe distressing: There are about 10^{28} electrons in an average human body so how many Coulombs of charge are there in the average human body? An easy calculation:

$$\text{number of Coulombs in an average human} = \frac{10^{28}\ \text{electrons}}{6.25 \times 10^{18}\ \text{electrons/C}} = 16 \times 10^{9}\ \text{C}.$$

...which is a lot. Given our aircraft carrier example, why aren't we each crushed by our mutual electrical forces? It's because we're neutral. For every electron in our body—and every other element—there is a proton. All of those little $e - p$ pairs cancel and the net result is that we don't repel or attract one another.[9] In fact, it appears to be a law of nature that all electrical charge comes ready-made with something of the opposite electrical charge. There aren't positively charged people, nor trees, nor stars, nor galaxies. This is a really important observation about the universe and we'll talk about it later.

[9] . . . that way.

The work of Coulomb, Ampere, Volta, Oersted and many others created a body of experimental knowledge and considerable confusion about the nature of electricity. Magnetism had been perplexing for centuries and while Ampere's guess as to its nature was close to the truth, the real explanation was more than a century out of reach in the early 1800s.

What was required was someone to look at the evidence with fresh eyes. Someone with a natural instinct for experimental technique and a modern dedication to uncovering Nature's secrets. What was required was a young man born in the extreme poverty of newly industrializing Britain with almost no education. Michael Faraday came along just at the right time to sort it all out and to intuit the strange solution. Almost miraculously, the young James Clerk Maxwell overlapped the mature Faraday and developed the mathematical power required to complete Faraday's insight. One of the most incredible scientific duos in history brought about the beginning of Modern Physics.

Chapter 12

Faraday's Lines of Force

Modern Physics Begins

Michael Faraday, by Thomas Phillips, *circa*1841.

Michael Faraday, 1791-1867

""The [natural] philosopher should be a man willing to listen to every suggestion, but determined to judge for himself. He should not be biased by appearances; have no favourite hypothesis; be of no school; and in doctrine have no master. He should not be a respecter of persons, but of things. Truth should be his primary object. If to these qualities be added industry, he may indeed hope to walk within the veil of the temple of nature. "" *Fifth Lecture to the City Philosophical Society, 1816.*

His story is almost out of Charles Dickens ...with whom he became friendly as an adult. While he was not exactly like Pip, Michael Faraday's story is a warm, threadbare-to-success tale. There will never be another Faraday. Single-handedly he linked electricity with magnetism, and back again. He essentially invented the field of electro-chemistry, the motor, the generator, and created the basis of numerous industrial processes. All because he was inquisitive.

12.0.1 Goals of this chapter:

- Understand:
  - Faraday's Law of Induction
  - Lines of Force (see the Diagrammatica chapter following)
  - Importance of the Ether
- Appreciate:
  - The general idea of a motor
  - The general idea of a generator
- Be familiar with:
  - Faraday's life

Figure 12.1: Young Michael Faraday

12.1 A Little Bit of Faraday

Faraday grew up in an exceedingly poor family in suburban London at the beginning of the Industrial Revolution. His blacksmith father moved the Faraday family to the city to try to find work and as a result, Michael and his three siblings had little formal education. But his luck was stunning and his enterprise was impressive. He was apprenticed at the age of 14 to a generous bookbinder and encouraged to read many of the books that he worked on. He found himself infatuated with chemistry texts, which were all the rage and was even permitted a little chemical lab in the book shop. By the time he was in his late teens, he was beginning to attend public educational events in the city and in 1812 a customer who was a member of the Royal Institution presented him with tickets to the farewell public lectures of the preeminent scientist in Britain. And so electric (pun intended!) with excitement, with notebook in hand, young Faraday set off to attend the series on chemistry by Sir Humphrey Davy of the Royal Institution. Davy was a flamboyant pioneer in many aspects of chemistry, but most notably electrolysis in which electrical currents are used to dissociate chemical compounds into their separate constituents. A one-man industry of chemical analysis, he gave regular public talks and demonstrations, often using his own body as a part of the show dangerously inhaling noxious elements of one kind or another.

With pencil in hand, young Michael faithfully attended the lectures in the gallery, wrote out and bound them, and presumptively sent them to Davy. So imagine Faraday's astonishment when one day a summons arrived for him to visit the Great Man who had fired one of his laboratory assistants for fighting and out of the blue, offered the job to Michael. His apprenticeship had ended, he was at loose ends, and so the timing was remarkably fortuitous. He accepted and after a short introduction in the Laboratory to his amazement he was bundled up with Davy's wife for a year and a half scientific and educational tour throughout Europe as Davy's "philosophical assistant."

Faraday had never been more than a dozen miles from London and perched atop Lady Davy's carriage (yes, he was made to ride on top of her coach!), the city boy delighted in letters home at the French countryside. The little scientific entourage embarked from Plymouth in October of 1813. Although the 20 year-long war with France was still raging the Great Man carried special credentials from Napoleon them to travel through enemy territory. And so, in spite of grumbling from London's conservative press, Davy pressed forward with safe passage to Paris, then to Montpelier, Genoa, Florence, Rome, Naples, Geneva, Munich, and dozens of other towns in between. During their trip, Napoleon's army was defeated, the Emperor was exiled, escaped, and hostilities renewed. These ominous events led them to cut their trip short, avoid France on their return to England in April of 1815.

This excellent European adventure was Faraday's alternative college education—a continuous "study abroad" experience which brought him into direct contact with all of the scientific luminaries on the continent, some of who he corresponded with for years. Throughout, Davy did experiments with Faraday's assistance. While exhilarating, this experience also made him very aware of his low station in life and how he appeared to others. To make matters worse, Lady Davy's self-appointed role seemed to be the reinforcement of their apparent class distinction which led to many despondent letters home about ill treatment at her hands. Faraday was equal parts gratified for the education and miserable. He wasn't just Davy's scientific assistant but also expected to be Dir Humphrey's valet. Faraday determined to change his manner and his speech and when they returned took elocution lessons and joined reading groups which he attended his whole life. "Self-made man" seems a label designed for Michael Faraday.

With their return, Michael spent the rest of his life working within the Royal Institution, eventually with his wife in provided apartments. Davy later commented that his most famous discovery was "Michael Faraday," but the good feelings didn't last, as much later the temperamental Davy wrongly accused Faraday of plagiarism and unsuccessfully tried to block his election to the Royal Academy of Science...a sorry chapter in an otherwise heart-warming relationship.

Davy quickly came to totally depend on Faraday. He chafed under the assistant's role, but broke through—by 1820 he started his own researches and naturally took up electrolysis as a study and methodically characterized both his procedures and his results. It is to him that we owe the terms "ion" and "electrode" among others.[1]

While an extraordinary experimenter—imaginative as well as skillful, he was notoriously mathematically illiterate. He knew it, everyone knew it. Yet without any formal training, his contributions were often guided by a natural and highly developed mathematical intuition. He "thought" mathematically,

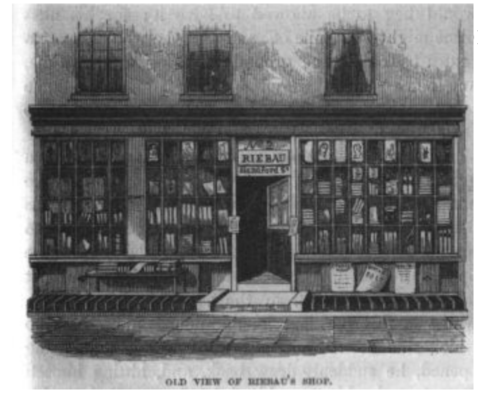
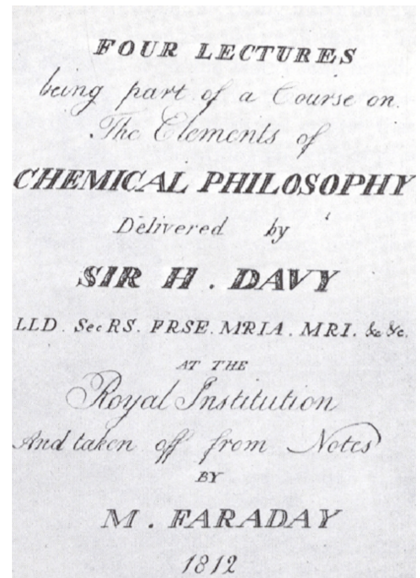


Figure 12.2: George Reibau's book bindery.



FOUR LECTURES
being part of a Course on
The Elements of
CHEMICAL PHILOSOPHY
Delivered by
SIR H. DAVY
LLD. Sec RS. FRSE. MRIA. MRI. &c &c.
AT THE
Royal Institution
And taken off from Notes
BY
M. FARADAY
1812

Figure 12.3: The book of Davy lectures compiled by 19 year old Michael Faraday.

[1] In his life-long self-improvement project, Faraday met and befriended many classical scholars and consulted them when he felt the need to name something. He often wanted to use a word with an appropriately stately pedigree in Latin or Greek.

he just couldn't express his ideas in that language. Later when James Clerk Maxwell codified his work in sophisticated mathematical form, he marveled at Faraday's intuitive and pictorial sense of the phenomena. In fact, arguably it was precisely his lack of formal training that freed him to make heretical suggestions which were quite outside of the standard wisdom. . . and which were often right. As his fame grew, it didn't immunize him from scathing criticism from his more sophisticated colleagues who would object on grounds that he was out of his league. What these colleagues couldn't appreciate was that not only was Faraday out of his league—he'd invented a whole new sport: Michael Faraday and James Clerk Maxwell were the first modern physicists.

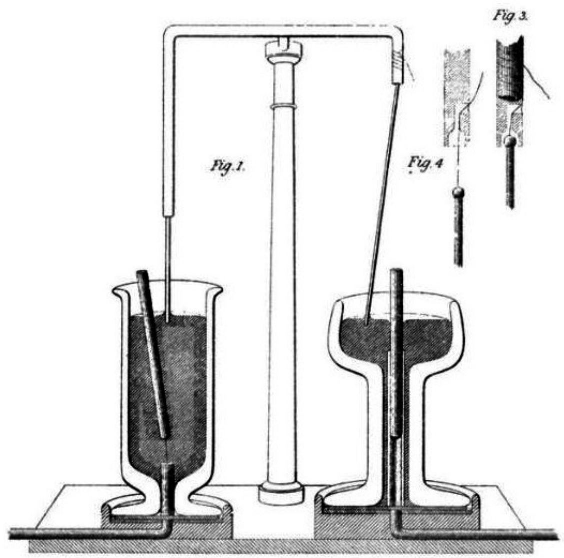


Figure 12.4: The arrangement of a wire free to rotate around a vertical axis (right) while maintaining electrical contact with the Mercury and the opposite situation on the left where a magnet rotates around a fixed wire also connected to the Mercury-circuit was proof of their reciprocal natures.

12.2 Faraday's Experiments in Electricity and Magnetism

By the time Faraday was working on his own, he knew all of chemistry. Davy had been a pioneer and was devoted to an atomistic-molecular view which rubbed off on Faraday who tended to think in terms of particulate matter all of his working life. His skills as an analytic chemist were second to none and he was in considerable demand for industrial consultation, fitting right into the goals of the Royal Institution. He created compounds of chlorine (discovered by Davy) and carbon, he isolated Benzine (and other organic molecules), liquefied chlorine, and discovered paramagnetism. He made important advances in alloying of steel. He was the go-to man in Britain for industrial chemistry innovation. Our concern is with this research into electricity and magnetism. Let's consider each of his major discoveries in their historical order.

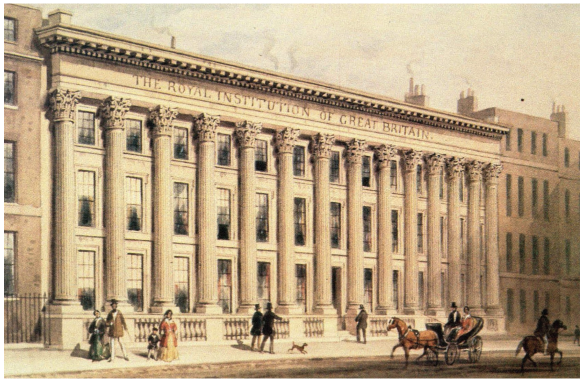


Figure 12.5: The Royal Institute of Great Britain which evolved into three roles. First, practical application of science for the good of industrial Britain. Second, a home for basic research— often at odds with its first mission. Finally a home for public science education. This latter role served as a fund-raising mechanism and was enhanced under Faraday's leadership.

12.2.1 The Motor

Oersted's results in 1820 electrified the Davy-Faraday laboratory.[2] Their lab was proficient in electrochemistry and so they must have had many of the materials required in order to repeat Oersted's experiment, and they did in great detail. It was Faraday who imagined that Oersted's compass demonstrating a circular relationship around a current-carrying wire might imply that magnets might themselves feel circular force around a wire. He had the clever idea to construct such a device, making use of the fact that mercury is a good conductor of electricity, while still being a fluid. Look at the left side of Fig. 12.4. The beaker of mercury has a bar magnet attached on a swivel at the bottom and at the top of the pool of liquid, a wire just breaks the surface. The wire running off the bottom to the left is attached to a battery and the vertical wire (ignoring the right hand part of the picture) is attached to the other terminal of the battery. When current flows through the mercury and the wire, the magnet swivels around its base in a circle around the upper wire.

[2] Get it?

The right hand picture shows what he did first. Ignoring the left hand side, now the wire comes in and attaches to a straight wire that's free to swivel around the top. Now a bar magnet is fixed vertically in the pool of mercury and is attached to the other end of the wire and battery. When current flows, the hanging, straight wire moves in a circle around the magnet. In each case (left and right) the circuit is completed through the mercury and the vertical wires.

Faraday connected the two experiments together as a show-and-tell stunt demonstrating the complimentary aspects of the same phenomenon: electrical energy is converted into mechanical motion, which is...the first motor.

Electrical energy can induce circular mechanical motion in a circuit. *Key Observation 17*

Faraday knew that this was a significant discovery. In his notebook he wrote, "Very satisfactory, but make a more sensible apparatus." But his 14 year old brother-in-law, who was in the lab with Faraday's wife, Sarah noted later that they all danced around the apparatus and then went to the circus to celebrate.[3] (He even prepared small hand-held versions of the Faraday Motor that he gave to colleagues for their amusement.) Faraday had none of the "Newton upbringing" and so his mind created different pictures. This idea of circular forces didn't fit into a Newtonian force-world and in this regard, Faraday was completely on his own.

[3] Michael and Sarah Faraday had no children, but was an active uncle with his nieces and nephews often playing in his laboratory. They later remembered his reactions to successes and failures and their descriptions of an excited uncle contrasted with his always sober lab notebook recording of his results.

At this point, Faraday is still a young researcher—he was 29 years old—and he made a rookie mistake: he published his results without properly acknowledging people who had supported him and done preliminary work themselves. One of those was Davy, who appeared to never forgive him raising the issue to the level of a complaint of plagiarism. It was a gross overstatement and everyone understood that to be the case. Except, apparently, his former mentor who tried to block Faraday's election to the Royal Society.

In 1821, young Faraday was asked to write an article which would summarize the all of the known electrical and magnetic phenomena. By this point, he was an extraordinarily careful experimenter and exceedingly clever with his notebooks, inventing an indexing system that he'd use his entire career. In-

Faraday's record-keeping was remarkable. Very early in his career he fell into the habit of numbering every paragraph in his logbook. Then periodically he would add to an index so he could efficiently refer back to his notes. He produced his massive reports on Electricity and Magnetism later in life which summarized his careful experiments, as well as his far-reaching conclusions. He maintained tens of thousands of numbered paragraphs in his career!

stead of just reviewing the literature, he decided to repeat all of the experiments that had ever been done on electricity which for a genius like Faraday, led him into new territory.

His Day Job

By 1825 Faraday was the director of the Royal Institute and inherited fund raising and financial affairs which were a considerable burden, as the place was nearly broke. Its purpose was originally to assist British industry with solutions which could be sold or which would increase efficiency in the nation's factories. That contrasted with Faraday's own personal desire to continue to do fundamental research, undirected by practical application. He found time to do so, but it was less than he would have liked. It was not until 1831—the year that Davy died—that he was able to return to basic research in electricity and magnetism. But he was always working on matters of public good and industrial progress. Basic research was inserted when he had time.

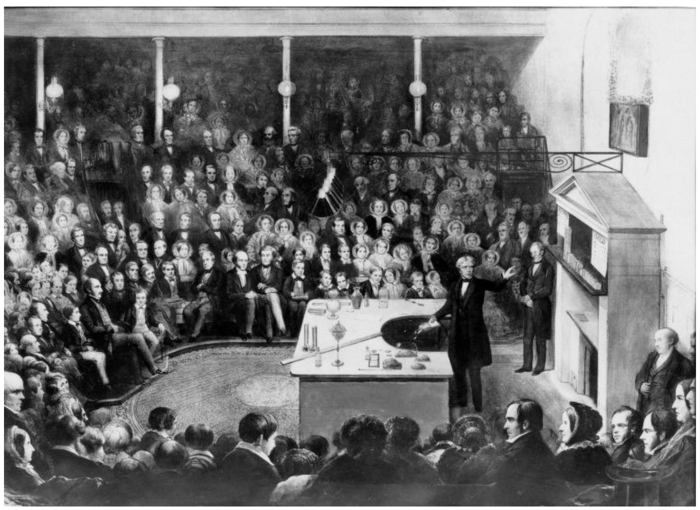


Figure 12.6: Faraday at the 1856 Christmas Lectures, one of his last. He gave 19 of them over his career—one of his earliest was on the chemistry of flame which he turned into *The Chemical History of a Candle*, a book for a non-expert reader.

But one thing he had time for was pay-back. Remember his good fortune in life had come with an opportunity of a chance public lecture. Faraday devoted himself, throughout his life, as an engaging and entertaining presenter of science to the public. As director, he instituted a nearly weekly series on Fridays—always exactly one hour—and he created a Christmas series for children. The Royal Institution Christmas Lectures have been given every year (except during WWII) from a virtual "who's who" of British scientists. Look at that overflow crowd in one of his last lectures Fig. 12.6! Michael Faraday was the world's first "Mr Wizard."

12.2.2 Induction

By 1831, he managed to clear his research decks sufficiently to take on a problem that had plagued him, as well as others. Clearly there were two ways to produce electricity: one could use friction—the silk and fur rubbed on glass and amber (your socks on the carpet) or use chemistry—a battery. And, Oersted and Ampère had shown that electricity—in the form of a current—could produce magnetism.[4] So surely magnetism should be able to produce electricity.

[4] Ampère had proposed that magnetism was caused by molecular electricity—little loops of current. So the challenge to Ampère was then how to tease magnetism out of matter. Faraday, while an atomist himself, was not enamored of this speculation.

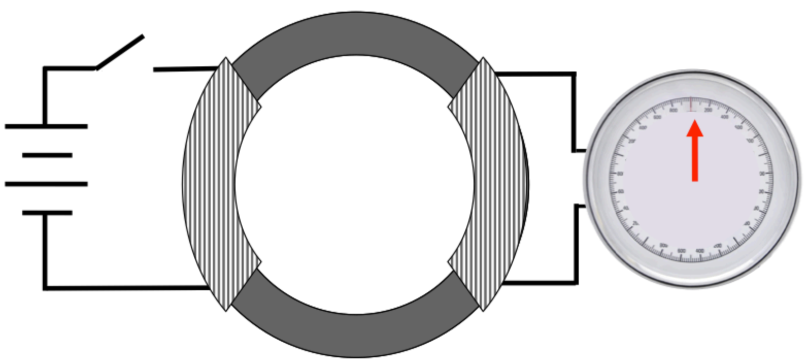


Figure 12.7: When the switch on the left is closed, then current flows through the circuit encircling one part of the iron core. The right-hand circuit contains no current source, but yet the needle on the galvanometer moved indicating that a current had been produced. Faraday interpreted the source to be a changing magnetic field, enhanced and contained within the iron toroid.

Figure 12.8: One of Faraday's original coils. Now called a "transformer."

Faraday went into a furious series of experiments trying all manner of materials and conductors which mostly consisted of trying to take adjacent circuits and get one to cause a current in another. His inspiration came when he took two long wires and wrapped them individually in paper and then wove them together around a core of a circular iron ring. The paper insulated the wires from touching, but they were all very close to one another. He hooked one set to a battery and looped the other circuit over a sensitive magnetic needle as shown in Fig. 12.7. If the needle moved when the current flowed in the first circuit, then it would have been induced by the interwound, but insulated wires.

Multiple tries didn't succeed until he noticed that the needle moved when the circuit was closed and then when it opened...and was stationary when the current just flowed. That was the key which everyone had missed, including after seven years of effort by Faraday: currents don't induce magnetism, but *changing* currents do!

Subsequently, he found another demonstration of a similar sort, but more directly and forcefully a magnet inducing electricity. Figure 12.9 is just a loop of wire connected to a Galvanometer.[5] No battery. No currents. Pointed at the loop is a bar magnet. That's it. Hold the magnet still, nothing happens. But, move the magnet toward or away from the loop and a current flows in it! Here, the changing magnetic influence is caused by a mechanical motion of the magnet. Current is created by a changing magnet, which is the principle behind a Generator. Keep this little experiment in mind.

[5] A Galvanometer is a sensitive meter that detects the presence of very small currents. If they go one way, a needle moves. If the currents go in the opposite direction, the needle moves the other way.

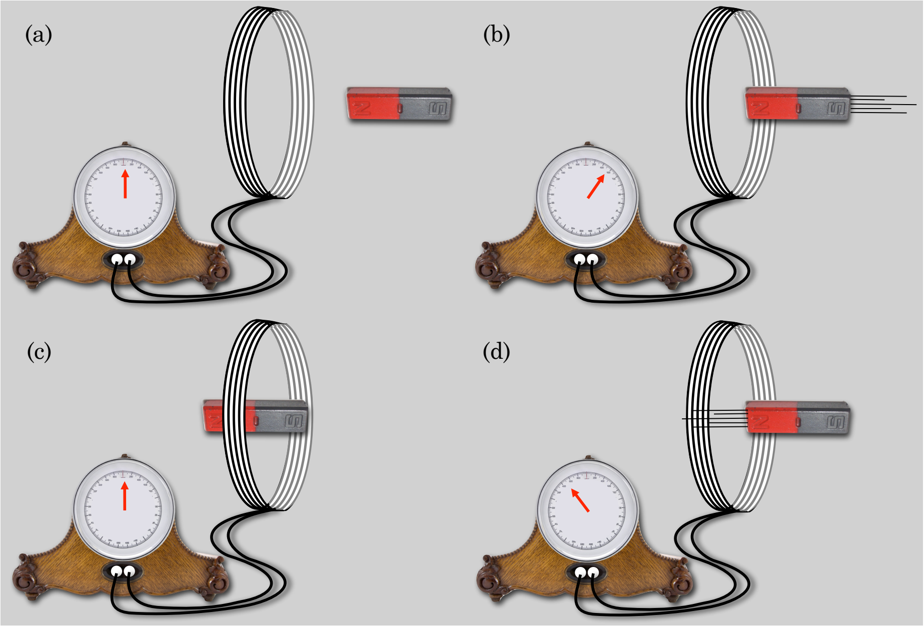


Figure 12.9: This is a standard demonstration of induction, but as we'll see, it is also a fundamental demonstration of the need for Einstein's Theory of Relativity. You wait. In (a) a magnet is stationary outside of a coil of wire. The wire in each stage is connected to a galvanometer and here we see that no current is flowing. ("Nothing up my sleeve" here means: there is no battery—no source of current.) In (b) the magnet just starts to move towards the coil and current flows. In (c) the magnet is inside the coil, and motionless and no current registers. And finally in (d) the magnet is just started to be removed from the coil and current flows, but in the opposite direction from (b).'

The ability for current to flow seems to be inside of the wire and by changing a magnetic field in the vicinity of the wire, that current is *induced* to flow. This phenomenon is called Electrical Induction and it's the principle behind all generators—creating current out of magnetic motion—and motors, its inverse—creating motion out of currents.

A changing magnetic influence creates a current. *Key Observation 18*

What's responsible for all of these various phenomena? Faraday had the standard repugnance for Action at a Distance, and knew that that wasn't an explanation of anything anyway. He wasn't so bound to the idea as were his classically trained colleagues, and so he thought about it in his own, fresh mind.

Faraday announced his discovery to the world at a meeting of the British Royal Society on November 24, 1831. That led to a priority struggle, as the publication of the effect was in early 1832, by which time the effect had been repeated in France and Italy...and where popular press reported that Faraday had confirmed the effect, rather than discovered it. He never again announced a discovery before formally publishing it in an international journal. The French and Italian scientists all acknowledged his priority and so trouble was averted. By now "Professor" Faraday—the uneducated printers apprentice—was world-renown and what he said and did mattered.

Out of Action

Faraday handled mercury and other dangerous materials as a matter of course. In his middle age he suffered a serious lack of memory and bouts of dizziness that took him out of action from about 1839 until 1845. It must have been terrifying. He was forced into months of seclusion and was absent from research only giving a few public lectures in the 1840s. Speculation is that he had poisoned himself. Can you imagine someone as energetic and inspired as Faraday, unable to work? Neither could Sarah, who enforced strict social access to her husband throughout his convalescence. He eventually recovered and resumed his activities. The challenge that seemed to energize him was his evolving view of space and the vacuum.

12.3 Lines of Force

Look carefully at Figs. 12.10 and 12.11, the proverbial 1000-words'-worth kind of pictures. The first is from Faraday's notebook, and so a sketch made by him sometime in the early 1830's. The second is a photograph from our lecture hall. What you see is a bar magnet surrounded by little bits of iron filings, each a little magnet of its own. By tapping the surface one frees them from any friction with the surface and they respond to an unseen presence—what Faraday called "lines of force."

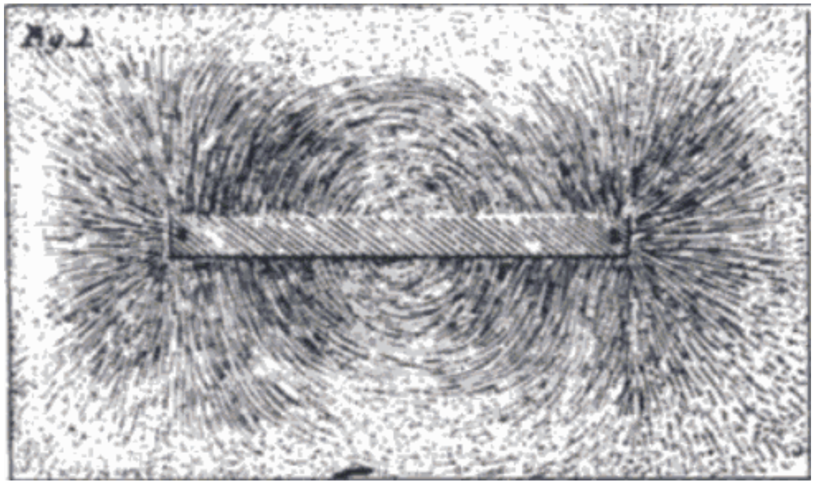


Figure 12.10: A patient sketch from Faraday's notebook showing the lines of force due to a bar magnetic on the accumulated action on little iron filings.

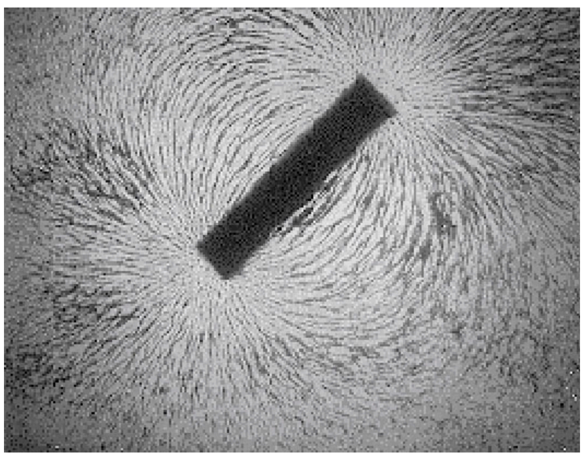


Figure 12.11: A modern photograph of a bar magnet, repeating Faraday's experiment.

[6] Of course, these tiny magnets are laying on a flat piece of paper and so they sample a plane-slice of the actual three dimensional magnetic influence.

He went so far as to induce currents from the tiny magnetic influence of the Earth and he began to speculate the gravity was of the same nature. In fact he tried unsuccessfully for years afterwards to try to induce a current by dropping wires in his lab.

One really cannot help but be mesmerized by this picture when you think about what it suggests. At first Faraday used the image of "lines of force" as a visualization...but by 1831 (in his diary) and 1845 in public and in print, he began to speak of the reality of what he coined a "field." While a compass points north, responding to an apparently invisible quality, these pictures revealed to Faraday that there's *something there*, hidden from view. Like a ghost materializing out of thin air in a Halloween cartoon, these lines of force reveal themselves by their spooky influence on the iron filings.

Faraday, like everyone, had trouble with Newton's action at a distance—that one body could reach out and influence another body with nothing in-between. But the scientific community—especially those in Newton's Britain—and come to peace with instantaneous, Action at a Distance. But his simple experiment shows that there's *something there*. Faraday played with this idea for years but slowly became committed to the reality of the lines of force—that empty space was not empty at all, but full of this almost material substance that seemed to propagate magnetic influence from one place to another. [6]

Faraday was aware that these were not going to be popular notions. First these lines of force were not Newtonian in character—they *curved*. They were *circular*. Maybe even worse: they seemed to *endow space itself with something to do*. In the Newtonian way of thinking, all space did was establish "place." It just defined separations between objects and staked out occupied regions. But Faraday's lines of force, if real, filled all of space. The effects of magnetism required space-filling lines of force. Or so his ideas went. But he didn't publish these ideas.

Instead, for years he amassed his evidence. Look at Fig. 12.12, again from his notebooks. Here we see the cross section of a wire (top) and two wires (bottom) all carrying currents. Iron filings were sprinkled on the sheet of paper punctured by the wires and again, by their orientation and arrangements, the filings signal the presence of magnetic lines of force. From the Oersted experiment, he knew that a current creates a magnetic effect and here he's shown it to be of the same sort of disturbance as from a bar magnet. It was time to come clean.

Finally, in 1837 he read what was the 11th of a series of *Experimental Researches in Electricity* to the Royal Society and he proposed that rather than being just a mental crutch, perhaps the lines of force were real. This was a frontal assault on (British) scientific belief and the reaction was not pleasant. In particular in the audience was the young William Thomson (the future Lord Kelvin), who was disgusted with the idea. Until he worked on it. What Thomson did was to apply the mathematics of heat conduction to the "streamlines" of Faraday's lines of force. He found that it hung together and by 1845, when he first actually met Faraday, he proposed that magnetic forces might actually affect the passage of light through materials.

Faraday set out to try to observe Thomson's effect, obviously encouraged that someone who was of the traditional scientific establishment, and a mathematical person to boot, would take his ideas seriously. He worked feverishly without results for almost a week, until he found the effect, which is today called the Faraday Effect: the presence of a magnet affects the polarization of light in some materials.

Aha. A magnetic disturbance affecting the propagation of light...were they connected?

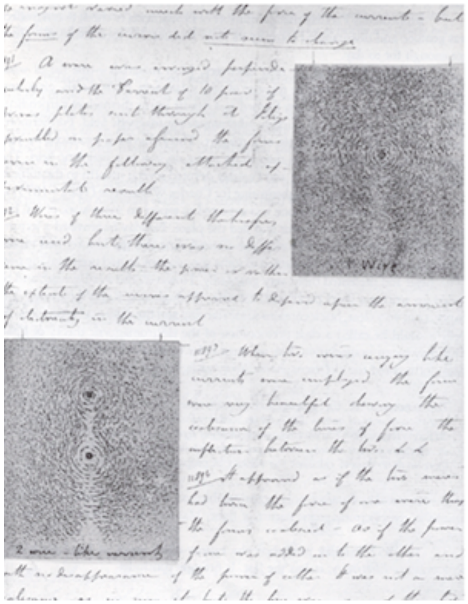


Figure 12.12: Another sketch from Faraday's notebook showing the arrangement of iron filings surrounding a current in a wire (upper right) and two wires (lower right). Notice that they are circular and surround the wires.

12.3.1 The Ether

Light is an undulatory phenomenon—it's a wave. For his colleagues the idea of light propagating through a vacuum was ludicrous. There had to be something "waving" for the light beams to undulate. In the classic sense of naming something being satisfying, but not really explaining, this "something" was called the Ether...specifically, the "Luminiferous Ether" for the Ether that propagated light (there were thought to be other ethers that propagated electricity and magnetism).

The Luminiferous Ether was a strange beast: It was everywhere. It didn't impede the motion of the Earth. It couldn't be seen, felt, tasted, or nudged and yet it could delicately react to light which stretched and compressed it as it propagated from one place to the other. Yet, since the speed of light was known to be very fast, the ether would have to be stiffer than steel to vibrate at that rate! The idea is very strange.

Too much so for Faraday whose views on space were extreme. He had no time for this unrealistic, invisible ether. Rather than imaging space to be just a big container that keeps everything from being all in the same place(!), he imagined a space that was full of electric and magnetic lines of force.

Figure 12.13: A daguerreotype of a mature Faraday from the 1840s.

For example his induction experiment could be explained this way. Look again at Fig. 12.9. As the magnet approaches the patiently waiting, inert coil of wire, more and more lines of flux are "cut" (his word) by the wire and it's the cutting of the "flux" (his word) that starts the current to flow. In fact, he even imagined that this magnetic flux would take time to reach the wire, so it was not instantaneous. Were a wire to be parallel to the lines of force, there would not be any current. Only when the lines of force cut a wire, would current flow.

The degree to which it is close to our quantum field ideas is almost spooky, as we'll see. To him what magnets, currents, and charges do is "pluck" the force lines into action...into wave-like action to transmit energy from one place to another at a finite speed.

Box 12.1 Filling In

Here's a story of how he publicly spilled the beans. One Friday evening in 1846 as he and the Royal Institution"s public speaker (Charles Wheatstone, of the "Wheatstone Bridge" fame) were walking to the hall, Charles saw that a renowned heckler was in attendance and being pathologically shy, beat a hasty exit leaving a full auditorium and nobody to speak. So Faraday took over...spending a little time trying to explain what the vanished lecturer would have told them. Then for the rest of the hour (because the lectures were always exactly 1 hour) he described his idea of a light, quite off the cuff.

"The view which I am so bold to put forth considers, therefore, radiation as a kind of species of vibration in the lines of force which are known to connect particles and also masses of matter together. It endeavors to dismiss the aether, but not the vibration."

Throw out the medium, but keep the wave. He began to write of a space that's full of magnetic, electric, and gravitational lines of force.

Figure 12.14: faradaylab

After that public exposure, Faraday tried to be more specific in a letter to Philosophical Magazine.

> "The propagation of light and therefore probably of all radiant action, occupies time; and, that a vibration of the line of force should account for the phaenomena [sic] of radiation, it is necessary that such a vibration should occupy time also...I think it is likely that I have made many mistakes in the preceding pages, for even to myself, my ideas on this point appear only as the shadow of a speculation, or as one of those impressions on the mind which are allowable for a time as guides to thought and research. He who labours in experimental inquiries knows how numerous these are, and how often their apparent fitness and beauty vanish before the progress and development of real natural truth."

He was roundly criticized for this speculation. He chalked it up to his learned colleagues being overly cautious and hide-bound to their preconceptions. His colleagues basically indicated that Faraday was a brilliant experimenter, but out of his depth when he ventured into speculative ideas about nature and mathematics. Faraday needed a hero and we'll meet his champion in the next chapter.

Faraday's health was a constant concern in his 50s and later. He might have suffered from Mercury poisoning. He was terribly worried about his memory losses and found that only trips to the country and a heavily-enforced isolation from anything scientific would restore him to working form. However, he had to exit himself for a couple of years to recover from a particularly bad episode and it never quite left him alone after that.

Faraday continued to experiment and unravel a number of mysteries in both chemistry and physics. He never forgot his modest education and worked hard to perfect a speaking and demonstration ability, giving many public talks in London through his senior years.

What we take from Faraday's work is of course the list of phenomena that he demonstrated. But, as important, or maybe even more so since other natural scientists would have come upon these same events. It was rather that mathematical intuition which when combined with his naivety about how things were "supposed" to be, that is his enduring contribution. Those lines of force, which he carefully mapped and measured in a number of electrical and magnetic configurations were the direct inspiration to arguably the most accomplished mathematical physicist apart from Newton and Einstein.

He died at the age of 73. Increasingly aware of his inability to remember and function, he resigned from the Royal Institution. His last lecture was given on a Friday and his notes bear some scorch marks where apparently they got too close to an open flame. He announced his retirement at that lecture to what must have been a stunned audience.

Figure 12.15: Michael and Sarah Faraday, 1851.

> "It is with the deepest feeling that I address you.
>
> I entered the Royal Institution in March 1813, nearly forty-nine years ago, and, with exception of a comparatively short period, during which I was absent on the Continent with Sir Humphry Davy, have been with you ever since.
>
> During that time I have been most happy in your kindness, and in the fostering care which the Royal Institution has bestowed upon me. I am very thankful to you, and your predecessors for the unswerving encouragement and support which you have given me during that period. My life has been a happy one and all I desired. During its progress I have tried to make a fitting return for it to the Royal Institution and through it to Science.
>
> But the progress of years (now amounting in number to threescore and ten) having brought forth first the period of development, and then that of maturity, have ultimately produced for me that of gentle decay. This has taken place in such a manner as to render the evening of life a blessing:—for whilst increasing physical weakness occurs, a full share of health free from pain is granted with it; and whilst memory and certain other faculties of the mind diminish, my good spirits and cheerfulness do not diminish with them.
>
> Still I am not able to do as I have done. I am not competent to perform as I wish, the delightful duty of teaching in the Theatre of the Royal Institution, and I now ask you (in consideration for me) to accept my resignation of the Juvenile lectures... I may truly say, that such has been the pleasure of the occupation to me, that my regret must be greater than yours need or can be."

He and his wife, Sarah, never had children but they were very content with one another, as evident in a letter her on one of his last trips,

> "My head is full, and my heart also, but my recollection rapidly fails, even as regards the friends that are in the room with me. You will have to resume your old function of being a pillow to my mind, and a rest, a happy-making wife."

He referred to himself as "altogether a very tottering and helpless thing, and requested a small funeral, attended by only his family. He died in 1867 and at the ceremony planned for family, friends and colleagues "came out from the shrubbery" to say goodbye.

The *Times of London* obituary said in part,

> The Late Professor Faraday
>
> "The world of science lost on Sunday one of its most assiduous and enthusiastic members. The life of Michael Faraday had been spent from early manhood in the single pursuit of scientific discovery, and through his years extended to 73, he preserved to the end the freshness and vivacity of youth in the exposition of his favourite subjects, coupled with a measure of simplicity which youth never attains...as a man of science he was gifted with the rarest of felicity of experimenting...It was this peculiar combination which made his lectures attractive to crowded audiences in Albemarle-street for so many years, and which brought, Christmas upon Christmas, troops of young people to attend his expositions of scientific processes and scientific discovery with as much zest as is usually displayed in following lighter amusements...

Faraday was beloved around the world and the *Times* listed a few of his honors:

> "Oxford conferred on him an honorary degree...He was raised from the position of Corresponding Member to be one of the eight foreign Associates of the Academy of Sciences. He was an officer of the Legion of Honour, and Prussia and Italy decorated him with the crosses of different Orders. The Royal Society conferred on him its own medal and the Romford medal. In 1858 the Queen most graciously alloted to him a residence at Hampton Court, between which Albemarle-street where he spent the last years of his life, and where he peaceably died on Sunday...No man was ever more entirely unselfish, or more entirely beloved. Modest, truthful, candid, he had the true spirit of a philosopher and of a Christian...
>
> The cause of science would meet with fewer enemies, its discoveries would command a more ready assent, were all its votaries imbued with the humility of Michael Faraday."

Chapter 13

Electromagnetism

Fields of Dreams

James Clerk Maxwell, at age 24 *circa*1855.

James Clerk Maxwell, 1831-1879

"All the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers." *On Faraday's Lines of Force (1856)*

Here's some high praise for a young Scots student, a little unrefined for Cambridge Trinity College, but expected to do well: "He (Hopkins) was talking to me this evening about Maxwell. He says he is unquestionably the most extraordinary man he has met with in the whole range of his experience; he says it appears impossible for Maxwell to think incorrectly on physical subjects; that in his analysis, however, he is far more deficient; he looks upon him as a great genius, with all its eccentricities, and prophesies that one day he will shine as a liable in physical science, a prophecy in which all his fellow-students strenuously unite."

13.0.1 Goals of this chapter:

- Understand:
  - goal1
  - goal2.
  - goal3
- Appreciate:
  - goal1
  - goal2.
  - goal3
- Be familiar with:
  - goal1
  - goal2.
  - goal3

13.1 A Little Bit of Maxwell

The William Hopkins referred to in the previous paragraph was a mathematics professor at Trinity College, Cambridge University (Newton's home) where he had a reputation for training the students for the terrible mathematics examinations call the Tripos. The first prize winner each year was given the title "Senior Wrangler" and Hopkins was the "Wrangler-maker." Maxwell came in Second Wrangler.

The "analysis" that Hopkins says is not as well developed refers to the fact that while Maxwell was powerfully ahead of his contemporaries in geometric thinking, his algebraic "analysis" of problems took second place. However, he correctly assessed Maxwell's skill. What James eventually accomplished was to combine these two ways of thinking—the geometric and algebraic—to a degree that not even Newton, nor Leibniz, nor Euler, nor Ampere...not any mathematical physicist had achieved. In many ways his accomplishments were only surpassed by Einstein in the whole history of physics in how geometrical intuition led the way for "regular" algebraic approaches to be deployed for the rest of us. It was Maxwell who discovered Electromagnetism and fleshed out Faraday's prescient concept of the Field and in doing so, revolutionized all of physics.

That Clerk Maxwell turned out well would not have been a surprise to any of his family or friends, or especially his teachers. He was always unusually inquisitive, even published poetry and a mathematical proof before entering college. That he was as funny and companionable, as well as considerate as a supervisor rounds out a picture of one of the Big Three (Newton, Maxwell, and Einstein) as being the most normal and highest quality individual among them, by far. In these ways, he was perfectly matched for collaboration with Faraday, about whom (largely) nobody would say unpleasant things either—except about his presumed atrocious ideas. Maxwell was indeed, Faraday's hero.

13.2 The Field

We tend to label 20th Century physics as "modern" and everything that came before, as "classical"...at least that's how we divide up a physics curriculum. I disagree with that classification for reasons that will become apparent as we continue. The field concept—the sole ownership of which belongs to Faraday and Maxwell—is so foreign to any mathematical or physical mechanism before it, that it is modern in every respect: especially technically and conceptually. It's impossible to underestimate the importance of Maxwell's Theory of Electromagnetism. It's a sophisticated theory, meant for experts, and so we'll limit ourselves to descriptive methods...with only a few well-meaning attempts at giving you the flavor of the calculations without requiring three years of college calculus! Let's get right to it in modern terms. The path to Electromagnetism was long and passed through many mathematical notational idiosyncrasies for which there's no benefit of reviewing.[1]

[1] For aficionados, it was the British engineer, Oliver Heaviside who put them into the modern vector-calculus form that thousands of physics and engineering students study today.

13.2.1 The Electric Field

The first field was the Magnetic Field, which, as we've seen, was forcefully suggested by iron filing patterns around a permanent magnet or steady current. But for our general introduction, the Electric Field is easier to understand, so that's where we'll start.

Faraday believed strongly in a unified nature in which the laws would be related and so given his conclusion about magnetic lines of force and his speculation about electric lines of force, he also imagined a Gravitational field in the same spirit...performing many failed experiments to try to detect a gravitational influence on currents and magnets. He was limited to guessing about electricity since a visual demonstration for charges analogous to the iron filings experiments was beyond his lab's capability. Nonetheless, he speculated about the existence electric lines of force and Maxwell baked that idea into his theory of both Electric and Magnetic fields.

To some its reality is an uninteresting question! A standard textbook in Electricity and Magnetism says, "Perhaps you still want to ask, what is an electric field? Is it something real, or is it merely a name for a factor in an equation which has to be multiplied by something else to give the numerical value of the force we measure in an experiment?...First, since it works, it doesn't make any difference...." To Edward Purcell, a completely pragmatic view is sufficient.

We've already dealt with Coulomb's law, which is the force of attraction or repulsion between two electrically charged objects. It's so important, let me reprise it here (Q_1 and Q_2, separated by a distance, R):

$$F = k\frac{Q_1 Q_2}{R^2}. \tag{13.1}$$

Remember that the Q's have an algebraic sign, + for a positively charged object and − for a negatively charged object—and then the multiplication results in an overall sign for F: if positive (like two positive or two negative Q's) then the force is repulsive and if negative (where one Q is positive and the other is negative), then the force is attractive.[2]

[2] We can just accept this as a rule, but a little thought and a careful definition of what the force acts on will create a vector result that will point to the other charge or away from it.

Into the middle 1800s, everyone assumed that whatever effects were felt by charges in Coulomb's law, masses in Newton's Gravitational law, and magnets were instantaneous and facilitated through motions in an ether.

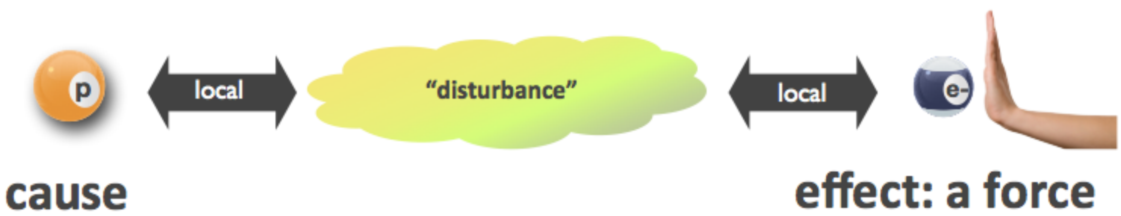


Figure 13.1: The metaphor of the Maxwellian electric field showing the reaction of the negative electron as due to the local field and not the distant Q.

Faraday felt otherwise: his fields would propagate between objects at a finite speed and were themselves "a thing" not requiring any intermediate substance. By contrast, Maxwell thought that his theory was a model of the ether—that the propagation of electric and magnetic fields were disturbances in it. It wasn't until the 20th century that the ether idea was abandoned. This strange substance maybe one of the longest-believed, mathematically sophisticated (lots of still useful mathematics was developed in trying to describe the nature of the ether), and *wholly false* models in the history of physics! So we'll describe Maxwell's theory differently from how he would have. Ours is an ether-free-zone.[3]

[3] It's always been interesting that one can take all of Maxwell's model... and simply remove the ether references and it remains intact.

The modern idea of the field as conceived by Faraday—who was right!—is shown in the cartoon of Fig. 13.1 which imagines the electrostatic attraction of an electron and a proton. A mathematical metaphor, if you will. In this view there are three aspects to the field:

- The source ("cause"). An electric charge, a magnetic pole or current, or mass create a field in its vicinity.
- The sink ("effect"). An electric charge, a magnetic pole or a current, or a mass detects a field in its vicinity.
- The disturbance ("field"). The intermediate space is filled by the field, which propagates at a finite velocity from source, to sink.

Of course, one person's source is another persons effect. That is, each acts on the other through the field.

13.2.2 A Quantitative Notion of a Field

The mathematical idea of a field is familiar to anyone who's looked at a weather map. It's nothing more than a distribution of some quantity in space (and time) with a value— a number—associated with every point in space. If it's weather, then any map that shows the distribution of temperature is a perfect example of a Temperature Field. You could imagine a million little weather-people all armed with thermometers and GPS transmitters who patiently take the temperature of the air in front of them and report it back continuously to Weather Central which displays it on a map. You'd expect that the values of the temperatures would be continuously varying between any two correspondents and such continuity is an important feature of a field.

Scalar Field

Figure 13.2 shows such a map. Continuity is manifest in the weather map in that the colors are not speckled like a pointillist painting, but continuous (transitions between the colors are continuous also...look at the scale at the top). Largely, the blues, greens, and yellows are connected and the colors indicate a continuous change of temperature across the country. This assures that fields can be described by smooth, mathematical functions.

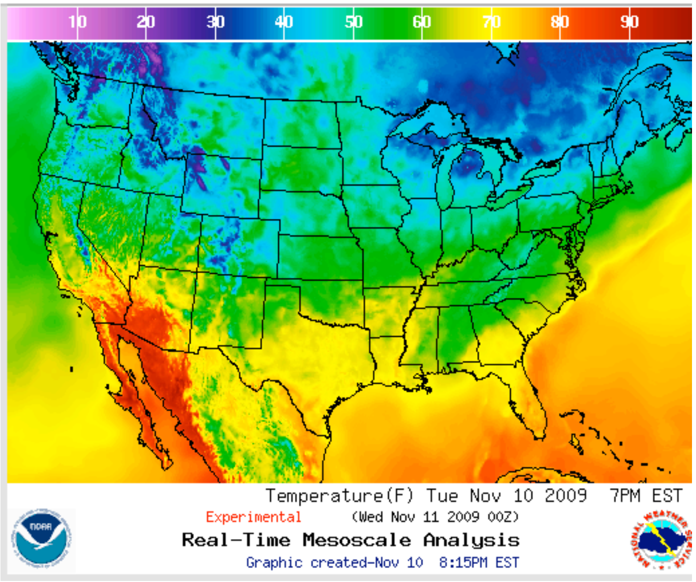


Figure 13.2: A weather map from the National Oceanic and Atmospheric Administration (NOAA) which is colorized to show the regions of common temperature values.

Another feature of a field, which will become important is that if you're holding thermometers in each hand, you expect that the temperature of the right hand thermometer only depends on the actual temperature in the vicinity of the right hand—not from the temperature across the room, or down the street. Likewise, the temperature of the left hand thermometer only depends on what's near the left hand. This is the idea of *locality*...that you can describe the effect by only the local conditions. If you're standing in front of a fire, for example, you might want to say that the fire is the cause of the warmth on your hand. But in reality, it's only the air exactly touching your skin that is the source of warmth, not the fire directly. In this way, the field is an intermediate carrier of some condition. If we have a theory that's correct about whatever that condition is (heat propagation) we describe the cause (fire) as creating the condition (the temperature field) which in turn, causes the effect (your warm hand).

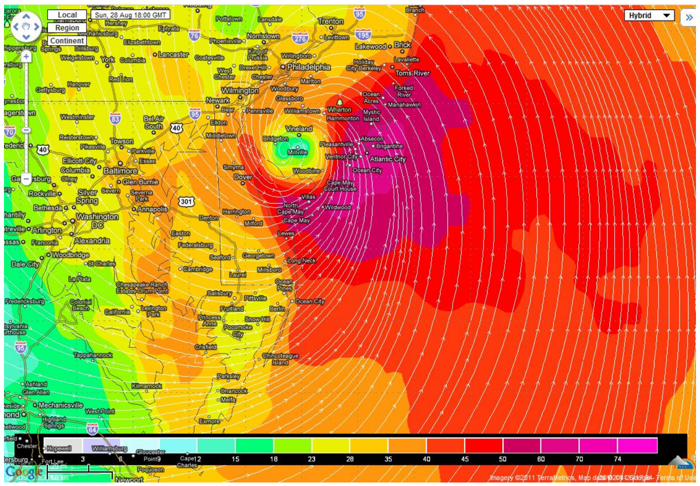


Figure 13.3: A Euro model forecast of wind speeds on the Jersey Shore during the height of 2011 Hurricane Irene: www.wunderground.com/wundermap

Vector Field

Let's take the field idea a step further: What's the direction of 70 degrees Fahrenheit? That's a nonsense question, right? Temperature, like speed or mass, is a scalar quantity, not a vector quantity. But what about the distribution of wind on a weather map, such as a hurricane? There, as is the case for all vector quantities, you care a lot about the speed—the magnitude of a hurricane's wind velocity—and its direction, which in the case of a North American hurricane, is counter clockwise, so often coming at you from the northeast, if you're on the east coast nervously watching a hurricane just coming ashore from the Atlantic Ocean (What direction is the wind if the hurricane eye has passed by you?) So wind velocity is an example of another kind of field—a vector field.

While a complicated mathematical subject, vector fields are easy to think about if you keep the wind-velocity idea in your head.[4] Figure 13.3 shows another weather map, this time a model for wind velocity over the NYC region during the 2011 Hurricane Irene. Electric and Magnetic Fields are Vector Fields with magnitudes and directions both required in order to characterize them.

[4] Our little weather-people might also constantly monintor the wind direction and speed and have little yellow-stickies that point in the right direction with arrose of lengths that indicate speed.

How the fields change in time depends on the physics being modeled (heat? sound? mechanical vibration? electromagnetism?). A model of the particular phenomenon would consist of a set of "field equations" which would be the calculation-machinery that would lead to predictions and encompass the physics of the particular fields in question. We model the field, and predict the behavior of the sources and sinks and if our predictions are confirmed, we accept that as evidence that the field is real and understood and the model is correct.

We'll need to understand field patterns for various configurations of electric charges and currents. Just like Faraday's magnetic field of force picture, we can do something similar for electric fields. In the spirit that a picture is worth 10^3 words, Fig. 13.4 is a picture of an electrode of a positive charge—a macroscopic version of a point electric charge. The green lines are little specks (sometimes of pollen) that are themselves influenced by electricity and align themselves in a clearly visible pattern. The pollen specks can become differentially electrically charged and so respond to the electric field in exact analogy to little iron filings responding to a magnetic field. "Something's there"! That was the conclusion that Faraday became convinced of for magnetic fields. Here you should have the same feeling about the "reality" of an electric field in staring at Fig. 13.4!

Figure 13.4: A photograph of little dielectric bits which orient themselves in an electric field created by the charge in the center of the photograph.

13.2.3 How To Detect An Electric Field

An electric charge needn't be of a point or an elementary particle—indeed in Faraday and Maxwell's time, such a notion was not even imagined. Rather their subjects were macroscopically sized objects like your finger when you've generated a spark from walking across the carpet— or like the silly, charged vegetable in Fig. 13.5. In this figure I've imagined a large piece of charged vegetable and a field emanating outward from it, just like in Fig. 13.4. How do I know that it's actually there? I can't see it or taste it or hear it.

We have to *interact with it.* This is our first example in physics where the measurer is an integral part of the definition of a physical phenomenon! That is, in order to "see" that an electric field is present, you must introduce another charge and watch what happens to it. That's what's pictured in Fig. 13.5. The broccoli is sitting there minding its own business and we bring a little, tiny charge, $+q$ and place it at that point shown. If there's an electric field there, our little test charge will feel a force. It the picture, we see that—both are positive charges so that force points away from $+Q$.

If we carefully note the direction of the force and the magnitude (which we can determine from its acceleration) then we can declare that an otherwise invisible electric field value is non-zero and is right...there. If the little test charge does *not* experience a force, then either that region is field-free—or, there are multiple fields present that just happen to cancel one another at that point.[5]

[5] It's a vector, remember?

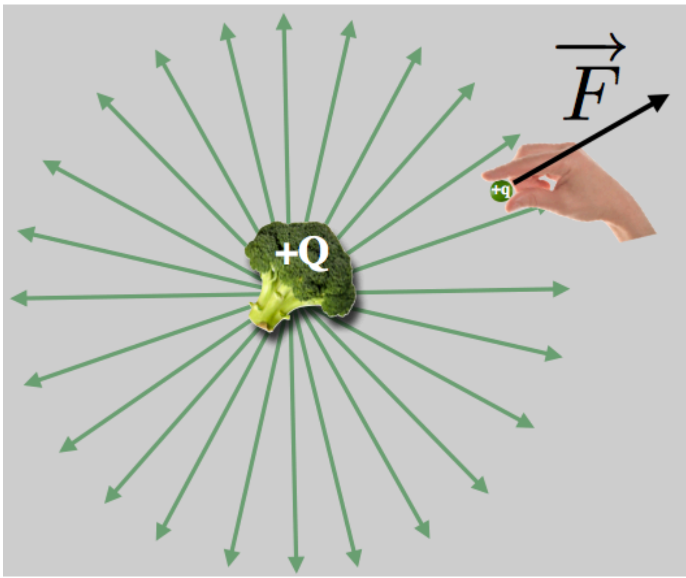


Figure 13.5: An Electric Field in the vicinity of a large charge distribution, represented by the broccoli. The force F on a little charge q is shown and since both are positive, that force acts to separate them: it's repulsive.

Our model is that of Coulomb and so we can predict the force on $+q$ with Eq. 13.1 as our guide, but now between $+Q$ and $+q$. Let's think about it in terms of Fig. 13.5 where we have a large positive charge, $+Q$ and a smaller positive "test" charge $+q$. According to Coulomb's Law, if we release the little charge in the presence of the charged broccoli, it will feel a force of repulsion and by Newton's rules, begin to accelerate away from it:

$$F_q = k\frac{Qq}{R^2}. \tag{13.2}$$

How do we know that a field is there? Well, we introduce little charges...little q's...and we see whether they accelerate. You do this all of the time with your car radio and with your cell phone. The little q's are the conduction charges (electrons) in the wire of the antennas that are built into all radios and now phones. When there's an electric field in the vicinity, these little charges feel a force and start to move and that motion is a current, which is suitably sampled and turned into Mom calling to find out where you are.

Notice that Coulomb's Law depends on both the big charge ($+Q$) and the little charge. Here's an ontological[6] commitment: we treat the field as if it's there, even when we're not testing it with test charges.

[6] Ontology is the branch of philosophy that deals with *being*. So my "ontological commitment" is a statement about reality.

Electric and magnetic fields are real. *Key Concept 25*

To that extent, we interpret the field as entirely due to Q and it should be defined in terms of only that charge. But of course, we are disturbing it a tiny bit with little q. That's new.

This is an unusual definition for a physical thing. We *presume* it's there, but in order to be sure we have to probe it with something...in this case, little q. How it responds tells us about the field. The "little" adjective for q means that we want to interpret our results as the field generated by "big" Q and not the effects of little q added in. On the one hand, we are really never observing the unadulterated field of Q. But on the other hand, charges are really, really small. I've avoided examples in electricity until now. Let's see just how much charge we're talking about here before worrying too much about our inability to perfectly measure the pristine field of a charge.[7]

[7] This is a practical statement, right? There's a more sophisticated point to make. When we get to quantum mechanics and relativistic quantum field theory, we'll see that this whole field-thing is radically modified.

So back to our charged broccoli. It's pretty easy to imagine a little test charge in the presence of any sort of charged object that we'd ever produce in a lab. So our need to not disturb the field is pretty easy. Maybe you'll see this demonstration in a class—if not, ask Mr Google for a video of charging a "pith ball." You'll find that a charge that you can reasonably put on a little ball is about a micro-Coulomb, 1×10^{-6} C. So if we don't want to disturb the field around such a little object, we'd have to use a test charge of much less than this...say 0.1% of that? If so, then the amount of charge that we'd get away with using as just a test would be $0.001 \times 1 \times 10^{-6} = 1 \times 10^{-9}$ C. But that's still a lot of electrons-worth of charge so if we detected our field with, say, conduction electrons in a wire? We could indeed get away with this.

In order to not change the field much from what it was before we "looked" at it with the little charge. If this is bothering to you, don't worry. You're correct to be bothered and when we talk about Quantum Mechanics we'll dig deeper for an interpretation.

Pencil 13.1.

But this workable metaphor suggested in Fig. 13.1of Q producing a field which accelerates a probing q leads to a convenient, if not subtle definition of the Electric Field...just take out the little q from the force equation, Eq. 13.2 :

$$E = \frac{F}{q} \tag{13.3}$$

Equation 13.3 doesn't tell the whole story. The force that q feels is a vector and so the field is also leading us to a vector definition of the electric field, **E**:

$$\mathbf{E} = \frac{\mathbf{F}}{q}. \tag{13.4}$$

I know. You're asking how can the field depend on whatever little q we stick into it? The definition actually accounts for that. The force depends on the product Qq so if we put in some other little charge (or even a big one), say $p = 2q$ then the magnitude of the force that p would feel is

$$F_p = k\frac{Qp}{R^2}.$$

and we see that $F_p = 2F_q$, but when I calculate the field (which is still due to Q in this narrative), I get:

$$E = \frac{F_p}{p} = \frac{2F_q}{2q}, \tag{13.5}$$

or the same thing as when q was the guinea pig.

From Coulomb's Law directly, we can then write the relation for the magnitude of the Electric Field due to a point charge of value Q at a distance R:

$$E = k\frac{Q}{R^2}. \tag{13.6}$$

The force lines in Fig. 13.5 are now replaced by the Electric Field Lines in Fig. 13.6. The lines get farther apart and that's the visual way in which we interpret the field's strength getting smaller and smaller as we move away from Q.

Notice in Fig. 13.6 that the field lines point away from the positive Q. This is a convention and coincides with the sign of the force that a positive charge would feel due to that field as in Eq. 13.4.

The signs work out for all of the possible combinations. Remembering that two charges of the same sign (either both positive or both negative) would repel one another and that two opposite charges (one positive and one negative) would be attracted.

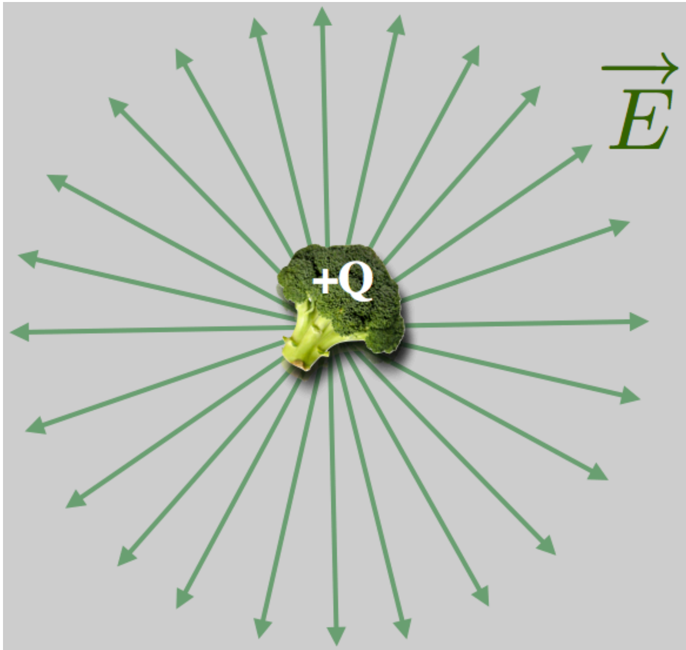


Figure 13.6: The Electric Field Lines due to a charge Q. Notice that they are pointed *away* from the positive charge.

Looking at Eq. 13.4 where only the field is shown, not the charge that causes the field, if q is positive, but Q is negative, then we get the field to cause a force that points towards Q by making **E** point the other way than the example above. So if our broccoli were negatively charged, then the arrows on Fig. 13.6 would point *towards* Q, rather than away from it.

Remembering that Faraday had the original idea and that he did not like the idea (nobody did!) of Action at a Distance, it's a small step from this discussion for electric fields back to the discussion of Action at a Distance from Newton's gravitational theory. Nobody liked it! But both time and Newton's huge reputation meant that an instantaneous influence across space for two masses was pretty much the accepted norm. Faraday's electric lines of force were not particularly well received and it was nearly a century before people were willing to overthrow the originally distasteful Action at a Distance for something more sophisticated.

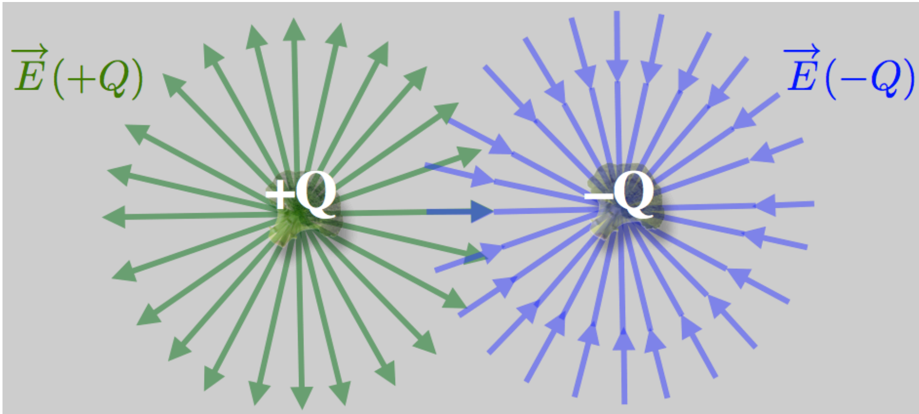


Figure 13.7: The Electric Field lines for two charges of opposite charge combine by adding their vectors at all points in space.

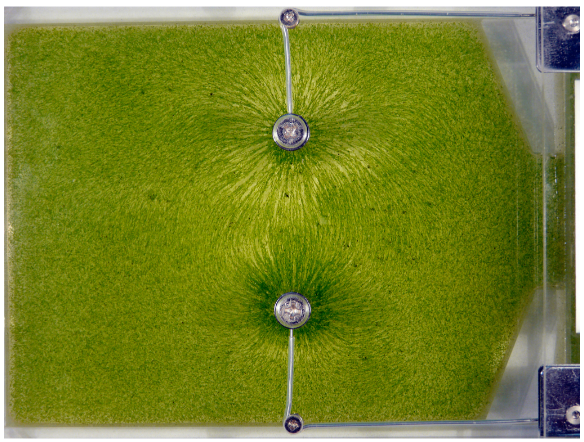


Figure 13.8: Two electric poles, one positive and one negative, which cause little dielectric grains to align with the electric field produced.

13.2.4 Electric Field for Other Configurations

The Electric Field for a concentrated charge at a point gives rise to the inverse-squared strength of the above discussion. But what about other distributions of electric charge, for example that of Fig. 13.8? Now there are two charges of opposite sign where obviously it's apparent that there is a much different force distribution from a single charge and of course, a different field shape. It's not too hard to think about how this is constructed: just overlay the field of a positive charge as for Fig. 13.6 with that of a negative charge (which we've already described as having all of the arrows pointing in) and add the vectors for $\mathbf{E}(+Q)+\mathbf{E}(-Q)$. This is suggested in Fig. 13.7 where for example, between the charges the fields would add in parallel toward the negative charge and everywhere else they would combine into a pattern that curves around the space between (and external to) the charges.

Suppose rather than a set of concentrated charges, we have a sheet of charge. Let's imagine a metal plate, infinitesimally thin, on which some charge has been added. As soon as that happens, they would scurry away from one another (since they would be repelled by their like charges) and since the metal sheet is a conductor, that "scurrying" would distribute the charges evenly all over the surface.

Now think about each charge, carefully and warily stationary because of all of the balanced forces from all of the other charges. Each would have the same Electric Field shape as in Fig. 13.6, but now there are cancellations. Suppose we have three people (A, B, and C) standing at attention side by side, but with their arms pointed diagonally up and their legs parted so that they all look like "X"'s. Person B's right arm, pointing up and towards Person C, would be overlapping with Person C's left arm which is pointing towards her and up. If their arms are vectors then the component of B's right arm pointing towards C would cancel with Person C's left arm pointing back at B. But the vertical component of each of their arms would continue to point straight up and at the same length. This cancelation along the horizontal and adding along the vertical would happen for each person in the line.

If we imagine that rather than people with their arms outstretched, we've got positive charges with their Electric Vectors "outstretched" then the result of adding together all of them would be vectors pointing perpendicular to the sheet from it surface up and also underneath, from that surface down. Figure 13.9 is a rendering of what this would look like. The result is a *constant* electric field pointing perpendicular to the sheet, undiminished through all of space.

If the charge dropped onto the sheet were negative, rather than positive the same redistribution would occur and the result would be another constant Electric Field, only this time the vectors would all be point *towards* the sheet.

Now let's make a sandwich of two such sheets which have been prepared with the same magnitude of charge, but with one positive and one negative. The "meat" part of the sandwich we'll take to be empty space. If the positive sheet is on top and the negative sheet on the bottom then the vectors from the top point down between the sheets (away from the top plate) while the vectors from the bottom (negative) plate point *up* away from it: they point in the same directions and so the two vectors add together with the resultant Electric Field having a value that's twice that of either of the plates.

Such a configuration is a very common one and is called a "capacitor," which is a way to store charges in a stable way. Figure 13.10 shows a drawing of the field distribution while Fig. 13.11 shows a photograph of such a device.

Charge distributions like this will be of interest when we turn to particle detectors and accelerators.

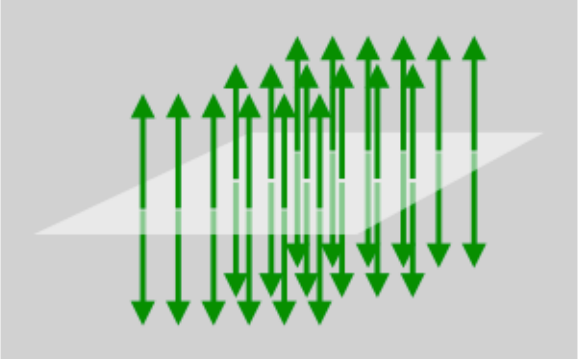


Figure 13.9: The electric field of a positive sheet of charge.

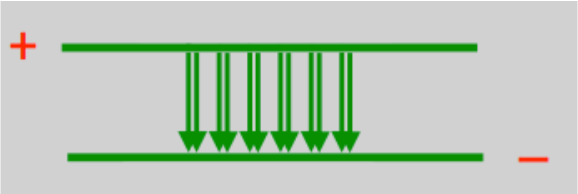


Figure 13.10: A "parallel plate capacitor."

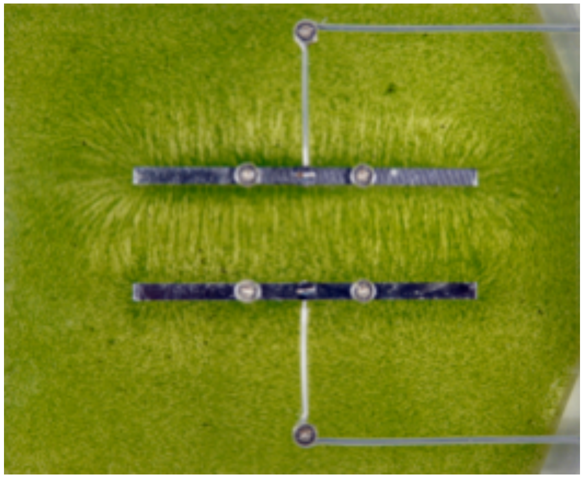


Figure 13.11: A parallel plate capacitor. Notice that the lines of force (the Electric Field) are parallel and uniform between the plates. Outside they all but cancel.

13.2.5 Energy In Electric Fields

Fields are the thing. We've established that a positive charge (our test charge) will be repelled by a positively charged vegetable (or, anything, fruit or vegetable), but now it's really time to drop the language that suggests any direct contact between charges. What matters is only how a charge is influenced by the *fields* in its vicinity—without regard to how those fields come about. We could create simple field shapes or complex ones, of course by arranging charges or shaping conductors with excess charges in them. We'll go for "simple" since we can do a lot with that!

The point-charge distribution, leading to **E** fields radially out from a center (caused by a net positive charge) or radially in (a negative charge), but that's actually relatively complicated. The field is diluted through a constant sized area as you go further away. The **E** field due to a localized positive electric charge is diluted as you go further away. Let's envision this.

You Do It 13.1. Bandaid

or copy the solution

Draw a positive charge (a + sign will do) and **draw an imaginary sphere around it** (make it half of the height of the empty space) and draw a band-aid on its surface. Now draw about 10 arrows representing the **E** from the charge, piercing the sphere, but make about 4 of them go through the band-aid.

Now draw a bigger sphere around the first one and let's draw another identical band-aid on it, in line with the first one, back to the charge. **Extend the lines that you drew originally** through that bigger sphere. Are there more, fewer, or the same number piercing our second band-aid?

What you should have seen is the arrows all piercing both spheres, but fewer of them going through the band-aid on the big sphere than through its partner on the smaller one. We think of the Intensity of the field to be proportional to the "lines" of the field. Please note! There aren't any "lines"...this is a visualization, but it's a near-perfect analogy for intensity, or strength of a field, so keep that in mind.

Now let's construct a perfectly uniform **E** field, which is very different from our point example. We've already talked about how to do that with a capacitor. Figure 13.12 shows such a circumstance. This is an idealized situation in which the field is perfectly uniform and confined to only reside between the plates (a real situation like this would also have some fringing of the field around the edges of a finite-sized capacitor. But then, finite sized anything is all we can really build.) Notice that we've put a positive charge ("p") and a negative charge ("e") in between the plates...what will happen?

Well, we expect a force on each and here the force is given by a slight manipulation of Eq. 13.4:

$$\mathbf{F} = q\mathbf{E}. \tag{13.7}$$

Let's pretend that for our situation, the postive charge is $q = +2$ and the negative charge is $q = -4$ (forgetting units for a moment). Now we can sing a familiar little song: when there's a force, there's an acceleration. When there's an acceleration, the velocity changes. If the velocity changes, then the kinetic energy changes. Not very catching, but you see the point. The positive charge will experience a force that is oriented along the electric field direction...that's because the charge is $+q$ and so the sign of the $\mathbf{F}$ is positive and hence, parallel to the direction of the $\mathbf{E}$. That is, $\mathbf{F} = +2\mathbf{E}$ Likewise, the negative charge uses its negative sign as a geometrical tool. So it experiences: $\mathbf{F} = -4\mathbf{E}$.

So, which charge will gain the most kinetic energy? Obviously here the negative charge does. Same field for both, but the size of the negative charge is bigger so it feels the bigger force and gains more kinetic energy. How much more? Glad you asked:

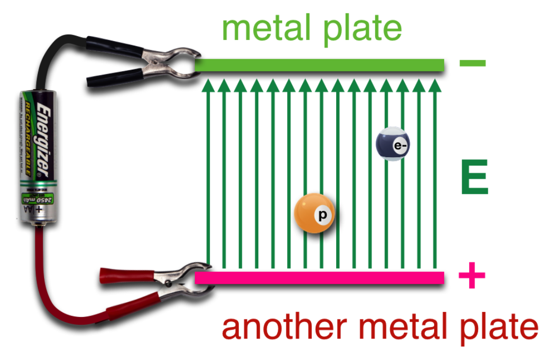


Figure 13.12: capcharges

Example 13.1

Kinetic Energy of Electric Charges In An Electric Field

Question : How much more kinetic energy does the negative charge experience as compared with the positive charge? Assume that they are both inserted into the field at the center so they start from rest. Also assume that they don't interact with each other. Finally, assume that the kinetic energies are compared after they have both moved a distance, d from the center.

Solution: Here's how to think about this: Equation 13.7 gives us a constant force. From Newton's Second Law, we can calculate the constant acceleration that would result...in terms of the mass of each charge. From that acceleration, using the last equation in Eq. 3.7, we can calculate the speed-squared. Then from there, it's a simple move to calculate the kinetic energy. Okay? Here we go:

String together Newton's Second Law with our electric field force, and solve for the acceleration, a in terms of the mass of the positive charge, m_p:

$$F = m_p a = qE$$
$$a = \frac{qE}{m_p}$$

Now find the speed squared from Eq. 3.7 after it's gone through a distance, d:

$$v^2 = 2ax = 2\frac{qE}{m_p}d \tag{13.8}$$

And now determine the kinetic energy

$$K = 1/2mv^2 = \frac{1}{2}m_p\left(2\frac{qE}{m_p}\right)d$$
$$K = qEd \tag{13.9}$$

Since the positive charge is half in magnitude of the negative charge, then the kinetic energy of the negative charge will be twice that of the positive.

This is an important conclusion. The kinetic energy of the charges comes from the force, but the force comes from the field. We conclude that fields carry energy...we could say that

Electric fields store energy and can do work on electric charges. *Key Concept 26*

This brings us to some familiar terms, but in a new context. For example, the term "voltage" comes in relating the work done on a charge in an electric field. When you deploy a battery in your flashlight, you're arming it to supply energy to electrons to force them through the circuit and the higher the voltage rating, the more current you can supply.

First of all, it's more useful to think about the energy that could be expended in moving a charge, and that's just the potential energy, U. So Eq. 13.9 can be more conventionally recast as:

$$U = qEd \tag{13.10}$$

A particular arrangement of electric charge will change how U relates to E, but there's a single relationship that always works and that leads to the definition of "voltage" which I'll introduce as

$$U = qV. \tag{13.11}$$

Voltage is the potential (energy) per unit charge. So the units of this quantity are Volts (V), or Joules/Coulomb. So

$$1\text{V} = 1\ \text{J/C}.$$

For this particular arrangement of charges in parallel plates,

$$V = Ed \tag{13.12}$$

which leads us to the more practical measure of Electric Field of "Volts/meter," which is pretty much universally used in engineering. (But from the original definition of the field, $E = F/q$, "Newton's per Coulomb" is another, perfectly acceptable unit for the field.

Let's put this together

Source	Electric Field Strength, V/m	Comments
atmosphere	100-150	near the surface of the Earth
home background	<100	typical home
inside your grandparents' TV tube	40,000	
near an electric blanket	<1,000	in typical use
near a microwave oven	600	
near a cell phone	<50	<0.1 inside your body, just below surface
below a 500kV power distribution line	10,000/m	some states restrict 3-5 KV/m at edge of a lane (These are the huge transmission lines that you see going across country.)

Table 13.1: Typical electric field strengths. Keep in mind that in order to cause a spark in dry air, an electric field strength of 10,000 V/**cm** is required.

Example 13.2

Your Old TV.

Question : What is the energy of a single electron accelerated through your old TV picture tube?

Solution: We'll use the field strength from Table 13.1 and assume that the tube is 0.5 meters long. Here's how we'll figure this out.

The kinetic energy, as we've seen, depends on the field, but now with knowing the voltage, we can use $U = qV$ to calculate it. So we need the voltage itself, but the table gives us the field. But we know the length of the tube, so we can calculate the voltage:

$$V = Ed = (40,000)(0.5) = 20,000 \text{ V}$$

The charge is that of a single electron, so $q = e = 1.6 \times 10^{-19}$ C and we find:

$$U = qV = eV = (1.6 \times 10^{-19} \text{ C})(20,000 \text{ J/C}) = 3.2 \times 10^{-15} \text{ J}$$

"Power tends to corrupt and absolute power(s of 10) corrupts absolutely."[8]

[8] Almost what Lord Acton said...

Electron Volts

All of these powers of 10 are really irritating. More importantly, they represent mistakes just waiting to happen. Not to fear! We have a very useful unit of energy that works very nicely for atomic physics, nuclear physics, and particle physics: the "electron-volt" aka, "eV." Here's how it works.

Suppose we have another capacitor that has a 1 V battery connected to it. What's the energy that a single electron (or proton, for that matter) would acquire as it's accelerated through that 1 Volt difference? It would be of course

$$U = qV = (1.6 \times 10^{-19} \text{ C})(1 \text{ J/C}) = 1.6 \times 10^{-19} \text{ J}$$

This is really useful and is the definition of a single electron volt:

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}. \tag{13.13}$$

13.2.6 Magnetic Fields

By this point, it's no surprise that the field idea could be applied to magnetic configurations, but with a twist. We know of no magnetic charges which would be called "magnetic monopoles." Not for lack of trying! Many theories of the beginning of the universe demand that they exist. But just as when you might chop a bar magnet in two, and then chop one of the pieces in two, and then again, and again...you will never find a separate North and South pole! Only N-S pole pairs seem to exist, all the way to the atomic level.

The field due to a bar magnet also follows the lines of force and start on the North pole and stop on the South. Ampere's guess that magnetism was due to little circular currents is compatible with the distribution of field lines as can be seen by comparing the iron filings from the bar magnet as sketched by Faraday in Fig. 12.10 with those of the straight wire in Fig. 12.12. However, some reorientation is required. First let's introduce our first "Right-Hand-Rule."

The mathematics due originally to Ampere and then as sketched by Faraday (without any mathematics) shows that the lines of force are circular, and idea that was not at all well received by the Newton Fan Club. Forces were supposed to be straight. Ampere's Law states that a current of magnitude I will produce a magnetic field in concentric circles around the wire in a direction that you can predict with your right hand: put your thumb in the direction of the current and unless you are built very strangely, your fingers will curl around the direction of the magnetic field, $\vec{B}$.

Figure 13.13 shows this configuration. The value of the field diminishes the further one is from the wire, but unlike Coulomb's Law, the rate of decrease is inversely proportional to the distance:

$$B = k' \frac{I}{2\pi R}$$

(k' is a constant that depends on the material outside of the wire.)

Remember Oersted's discovery? A compass would align itself around a wire when brought near a current. A compass is nothing but a little bar magnet and his discovery was just the statement that the magnet aligns itself with the magnetic field with the north pole following the **B** field direction. Of course that's all a compass is doing as a navigational device, since there is a tiny magnetic field due to molten currents in the core of the earth. It's following the Earth's **B** field pointing to the geographical North pole, which is the magnetic South pole.

Now let's take the wire and bend it into a circle. The field is still concentric around the wire, but look at how its field manifests itself in Fig. 13.14. Inside of the circle, the field is concentrated where all of the field lines add together (imagine wrapping your fingers around the wire, all around the wire). The field rises out of the plane of the circle, traverses around and returns through the loop from below. It's exactly

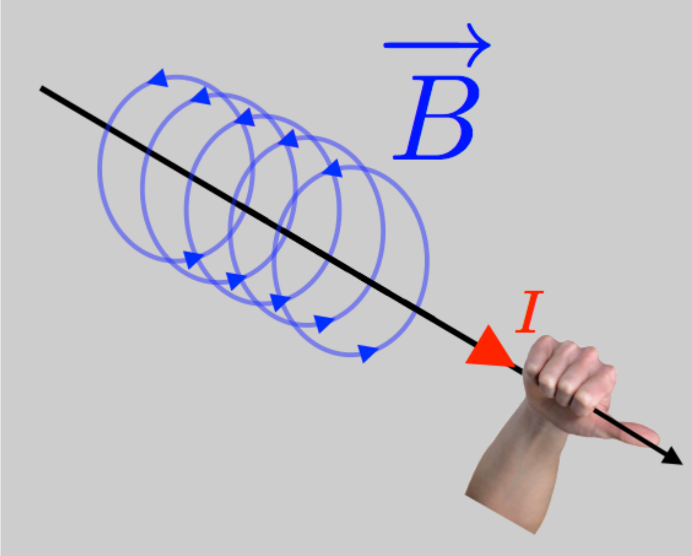


Figure 13.13: The magnetic field, $\vec{B}$ due to a wire carrying a current forms concentric circles around the wire in a direction related to your right hand as shown.

the form of the field of a bar magnet. Now Ampere didn't know this, but his imagination was such that he got it right.

If we make a tube of current circles like a Slinky, the field lines continue to add inside and there result is a useful circuit element called a Solenoid (also useful as a part of your car's starter circuitry.)

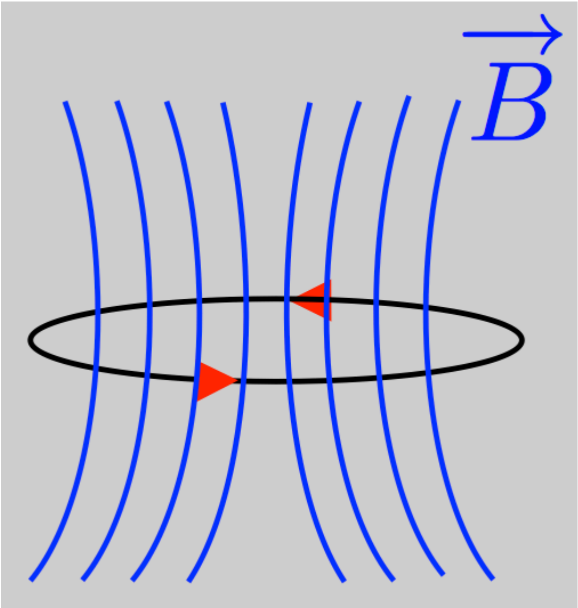


Figure 13.14: The magnetic field of a ring of current is shown.

Part II Physics and Cosmology of My Parent's Generation

Chapter 14

Special Relativity, 1905

space and time aren't what they used to be

Albert Einstein during his time at the Swiss Patent Office.

Albert Einstein (1642-1727)

"Now to the term 'relativity theory.' I admit that it is unfortunate, and has given occasion to philosophical misunderstandings." *To E. Zschimmer, September 30,1921.*

Can you think of a more recognizable face than that of Albert Einstein (1879–1955)? Even in our culture of being famous for being famous, *Time Magazine* named him the *Person of the Century* in its December 31, 1999 issue. The *century*! Einstein's scientific career was as much or more remarkable than Newton's and together, they complete an exclusive club of two.

14.1 A Little Bit of Einstein

In one year Albert Einstein had three breakthroughs of pure thought and simple mathematics, any one of which would put him in textbooks forever.[1] Before 1920 he had at least four more theoretical discoveries that were again, all Prize-worthy. He basically invented five different fields of physics and changed the way humans look at themselves and our universe forever.

[1] Yet when it was time to award a Nobel prize–he received only one–it was held up for a year because of antisemitism in the Nobel Committee, and among many of his then-German colleagues. There was no Nobel Physics Prize in 1920 and then two were awarded in 1921, including his.

Einstein was a complicated man. He was not plagued with the sort of insecurities that blinded Newton, or the self-destructive combativeness of Galileo. He sometimes showed a shocking inability to empathize with individuals, while simultaneously demonstrating great feeling for mankind and professing a highly principled view of world affairs. He had great, life-long friendships and a childlike sense of humor that showed through in photographs and interviews. People around the world loved him and, while he was not prepared for the notoriety that he would achieve—bursting on him overnight on November 7th, 1919—he clearly grew to enjoy the spotlight. The myths surrounding him are, as we've seen, an inevitable consequence of being larger than life.

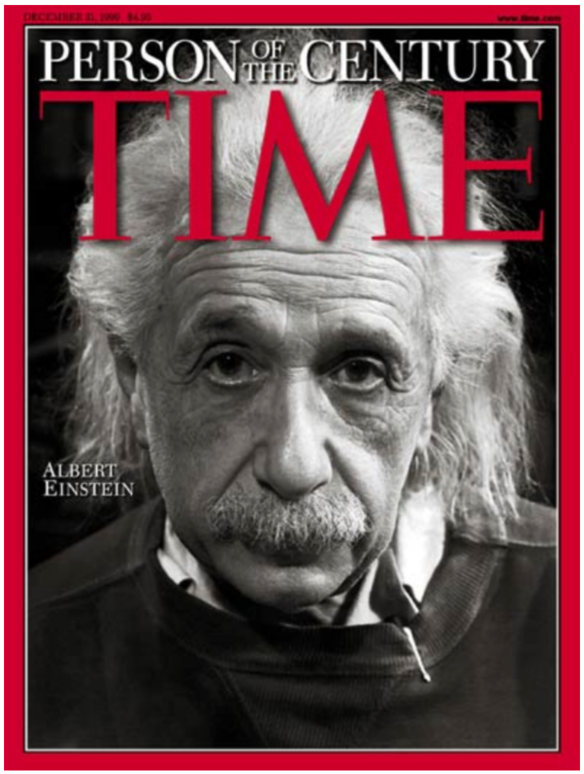


Figure 14.1: The cover of *Time Magazine*, December 31, 1999.

He died in 1955 (when I was five years old) in Princeton, New Jersey where he and his second wife escaped from the Nazis two decades previously. I think he looked older than his 76 years, perhaps a reflection of a stressful life following his scientific discoveries. Their Mercer Street home is still a private residence and he requested that it not become a museum (as his apartment in Bern, Switzerland is now). Those wishes were respected, although it is listed on the National Register of Historic Places.

14.1.1 Education

Einstein was so smart that he was born at a very early age. :) He grew up in Munich in a comfortable household with a younger sister he adored. His father and uncle had a successful business of electrifying German municipalities. But it failed when he was a teenager and the family moved to Italy to start over. But Einstein was famous for his independent streak and even at 15 years of age he remained in Munich by himself. The Catholic school in which his Jewish parents had enrolled him was not a good fit, but then most schools weren't.[2]

[2] This wasn't a match made in heaven and when he got fed up with the regimental manner of education and eventually followed his family south. After he left one of his teachers publicly expressed a sigh of relief. Albert was not shy to challenge teachers as a youngster and certainly as a college student.

Wait. *Didn't Einstein flunk math?*

Glad you asked. *That's one of those things that "everyone knows" about him. But it's not true, in fact it's the opposite of true. He mastered differential and integral calculus by the time he was 15 and precociously learned algebra and geometry as a very young child. He was mathematically gifted.*

Einstein was famously lazy as a student in spite of being brilliant. After he left high school, he took entrance exams for the Swiss Federal Polytechnic[3] in Zürich two years early, but while his scores were exceptional in mathematics and physics, they were unacceptable in other subjects and so he lived with a family near Zurich and attended a private school for a year in order improve his chances. That worked and he entered the Swiss Polytechnic in 1896 to study physics when he was 17 years old.[4] Also enrolled was the only woman in his class of six students, Mileva Marić. They became inseparable and a love-affair built around their joint studies bloomed. Much of their correspondence still exists.

In 1900 he graduated with top marks[5] (and a bad reputation) receiving a high school physics teaching degree. Mileva did not graduate, because of low mathematics scores—she tried a second time and again failed. For the next two years, Einstein unsuccessfully searched for a permanent teaching position and during that period Mileva went back to Serbia to have their out-of-wedlock child, a daughter. The baby's eventual fate is unknown and Einstein never saw her. They were married in a small civil ceremony a year later[6] and eventually had two sons.[7] Einstein had left such a bad taste in the mouths of the Polytechnic faculty that he was the only one of the graduating class to not be offered a continuing research position. He even had reason to believe that the professor who first supported him was by his graduation actively disparaging his former student to other prospective employers.

Marcel Grossmann was a college friend and talented mathematician who figured in Einstein's life multiple times. He was often a source of lecture-notes for when Albert skipped class—which was frequently.[8] And he was from an influential family and prevailed upon his father to help his friend, by securing a job for Albert at the Swiss Patent Office as an examiner in 1902.[9] So he and Mileva moved to the capital city of Bern where he settled in as a middle class clerk.[10] He managed to begin a research project with a professor at the University of Zurich simultaneously and in 1906 received his doctor's degree.
The year before he graduated? He changed the world.

14.1.2 Bern Years

Einstein's work at the Patent Office was not demanding and he could not only pursue his graduate degree part time, but also work on his own, outside of an academic environment.[11] He, Mileva, and Hans Albert lived in a second-floor apartment[12] a few blocks away from work and he would pass through the famous Bern Clock Tower (the *Zytglogge*) twice a day. That iconic, medieval structure and its famous performing clock figures into his later descriptions of how he came to Special Relativity. Let's go boating.

[3] Now called *Eidgenössische Technische Hochschule*, ETH.

[4] In the process, he also renounced his German citizenship in order to avoid his obligation to go into the military. From a very early age he was a pacifist and ardently opposed to the militarism that was Germany at the turn of the century.

[5] "Oh, that Einstein, always skipping lectures. . . " A remark by one of his professors, Hermann Minkowski, who later put Special Relativity on a firm mathematical foundation.

[6] Einstein's mother was especially disapproving of this marriage.

[7] One, Hans Albert, became a professor of Chemical Engineering at the University of California, Berkeley. The other, Eduard, had a breakdown at the age of 20 and spend the rest of his life in and out of mental health wards with schizophrenia. Einstein expressed affection for a former lover while married to Meliva and eventually married his cousin, Elsa Löwenthal. The divorce from Meliva was very unpleasant and included strict instructions on her handling of their children and behavior towards him and the proceeds from his expected Nobel Prize. Elsa died three years after they moved to Princeton in 1933.

[8] He taught himself Maxwell's theory of electromagnetism out of class, disapproving of the German theory that was taught at the time. The Electromagnetic course was taught by that professor who became an enemy and Einstein had mocked him for teaching only old material rather than the more modern Maxwell.

[9] We'll see Grossmann return a decade later as Einstein's tutor in the advanced mathematics he would need for his General Theory of Relativity.

[10] "That secular cloister, where I hatched my most beautiful ideas and where we had such good times together." From correspondence to his Patent Office, and Olympia Academy pal, Michele Besso.

[11] "Whenever anybody would come by, I would cram my notes into my desk drawer and pretend to work on my office work."

14.2 Frames of Reference

You've all had the sensation. You're on a train and next to your car is another one that suddenly begins to move. Or do you move?. . . you can't quite tell the difference. Galileo asked whether there was any way to distinguish moving from not moving, and said "no" and of course, Aristotle had said "yes." In order to figure out who was right, let's set up a thought experiment.

Figure 14.3 depicts Galileo (G) and Aristotle (A) each standing on the deck of a boat (B) that's moving to the left at a constant speed relative to the shore (S), where Newton (N) watches them go by. Kepler (K) is at the top of the mast holding an apple[13] and loses his grip and it falls. Being good scientists, G, A, and K each predict the path that the apple will make on the way to the deck. . . where will it land?

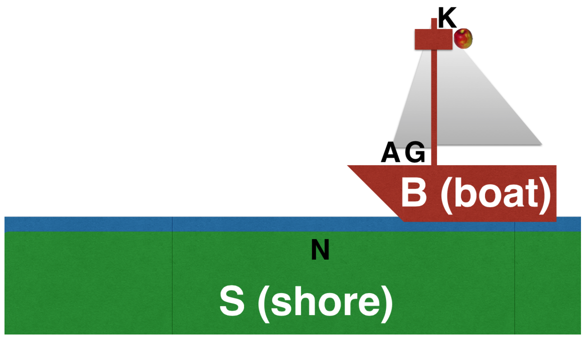


Figure 14.3: boat1

[13] Fruit, again, play important roles in the history of physics! Not.

- Of course, A answers that the ship will have moved out from under the apple as it falls and so it would land near the stern, away from the mast. With that, he's repeating one of the classical arguments against a moving Earth: that the atmosphere would be left behind, that birds would fail to reach their destinations if they were flying in the same direction as the Earth, and so on. So the Earth must be still: Earth at the center of Aristotle's universe is absolutely stationary.
- G would tell A, "No!" The apple would land at the foot of the mast, falling straight down.
- K would look at the apple's fall from up above and also agree that it would fall straight down, directly below him.
- Newton (N), on the shore would watch the whole spectacle and when asked where the apple would land, he'd say, "with respect to what?"

Newton's question sets up an important idea that we'll use repeatedly: the notion of a **Frame of Reference**. In this story, there are two main frames of reference: The **shore, S**, and the **boat, B**. Newton would say that B is moving relative to S and that S, with him, is stationary.[14] What would Newton actually see relative to his fixed-Earth frame, S?

[14] He'd go further. Remember our discussion of Newton's Absolute Space. This unique, definitely stationary reference would be the

- While still in Kepler's hand, the apple would have the motion of the boat, to the left. When he drops it, the apple *still* has that horizontal motion, but now it starts to acquire the downward acceleration due to gravity as shown in Fig. 14.4. What N would observe is the same parabolic trajectory that Galileo discovered when he rolled a ball off the table edge. Here the boat's motion is providing that horizontal speed.

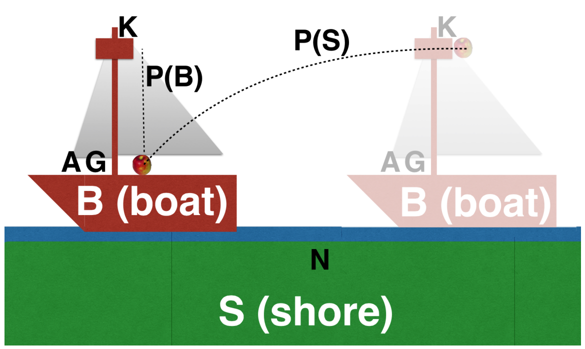


Figure 14.4: boat2

For Galileo and Newton a moving Earth is no problem. Just like the apple, birds and the atmosphere all share the Earth's speed (albeit a rotational speed. which we can pretend to be linear because of the large radius of the Earth) and are dragged along with it.[15]

[15] When you fly across the ocean East or West, you're not helped or hindered by the Earth's rotational speed and direction. You're flying through and relative to the atmosphere which is being dragged along with the Earth.

What would G, A, and K say about the shore, S?

- A would state that S is stationary, as it's part of the unmoving Earth, and that he and B are moving with respect to it.
- G and K would be more sophisticated, and Galileo in particular worried about this situation. They would say that B (on which they are passengers) appears to be stationary, and that S seems to be moving at a constant speed the other direction from N's assertion.

Frame of Reference. A fixed coordinate system in space and time. *Key Concept 27*

14.3 Galilean Relativity

Let's get technical and instrument our frames with their own space and time measuring devices: Figure 14.5 is a little spacetime measuring kit consisting of a ruler[16] and a clock. We can imagine that every moving object can be so equipped and that events anywhere can be measured in space and time from within any frame.

[16] We'll be in no more than two space dimensions and our relative motions will all be in one dimensions.

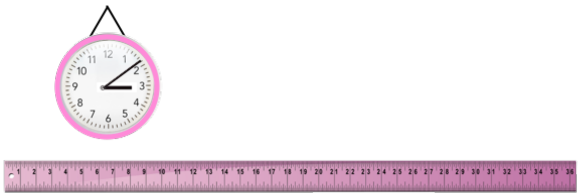


Figure 14.5: A little spacetime tool-kit, suitable for any frame of reference. Blue is a part of the AF equipment and pink is for the HF.

Let's recap what we've learned so far:

- Every observer is in his own Frame of Reference and at rest relative to it.
- If an observer moves by you at a constant speed, you'd say she's moving and you're not.
- Likewise, she would report that you're moving and that she's not.

Now, let's kick it up a notch.

Let's imagine that Newton and Galileo are playing catch[17] on the shore.[18] Each learns how to throw a given distance and since they're tired and they know the rules of projectiles and force, they rig pitching and catching machines to play their game for them. With a few trials, they can calibrate their machines to accurately pitch and catch, back and forth.

[17] Maybe with an apple.

[18] In these physics stories, you don't ask why the characters do what they do.

Now they take their machines to the deck of the moving boat.[19] Without making any adjustments at all, the machines pick right up where they left off and continue the game of catch with the same repetitive success as on the shore. Does this surprise you? I'll bet not.

[19] In this chapter, "moving" always means at a constant speed.

Galileo's and Newton's rules about motion and force work exactly the same in constant speed, co-moving reference frames. Instead of setting up the machines by trial and error, suppose they'd first solved the algebraic equations of motion according to Newton's force and Galileo's motion rules, they would predict the same settings for their machines as they guessed in the trial and error approach. But the exact same equations would work in both the shore and the boat frames. That is, the same rules of physics would be active in both B and S.

Galileo wrote about this in 1632 and it's worth reading. Remember, this is the early days of figuring things out (emphasis, mine):

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), *have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that.* You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still.

What Galileo means in his wordy way is that there would be **no** measurement that you could perform that would tell you either that you were moving or stationary. If there were no windows, you would have to conclude that you are at rest.

Definition: Galilean Relativity..
For relatively moving, constant velocity (inertial) frames of reference, that mechanical rules (only) can be transformed between frames by assuming that time is independent of the frame is called Galilean Relativity.

Einstein called this realization Galilean Relativity: **there is no mechanical measurement that can detect that a frame of reference is moving at a constant speed or at rest with respect to any other frame of reference.** A frame of reference that's at rest with respect to itself is called the **Rest Frame**. This makes sense.

Wait. *Does this always work this way?*

Glad you asked. *What if instead of the linearly, and constantly moving boat, they take their machines to a big playground merry-go-round where they set up on opposite rims across the center. Now they try to play catch with the same settings as the previous situations, and they would find that if the merry-go-round is rotating that every throw would be way off target. A throw straight ahead in the merry-go-round frame, would appear to be curved, rather than straight and a good student of mechanics would recognize that as a demonstration that there was a force at work and that the frame was accelerating.*

Frames of reference that are not accelerating, such as the boat (as view from the ground) and the shore (as viewed from the boat) are called **Intertial Frames of Reference**. An accelerating frame of reference, like the merry-go-round is non-inertial. Let's go to the airport.

Definition: Rest Frame..
A frame of reference that has no relative motion to another is a "rest frame." Every object is in its own rest frame.

Definition: Inertial Frame of Reference.
A frame moving at a constant velocity relative to others is called an Inertial Frame. If accelerating, they are non-inertial.

14.3.1 Coordinate Transformations

We'll usually be concerned with two different frames of reference which move side by side (so one dimension) with respect to one another. In this chapter they will be Inertial Frames and transportation industry analogies will be impossible to resist. Figure 14.6 is our tool-kit (rulers and clocks) which we'll pretend is standard equipment in all frames of reference. The overriding puzzle that we must solve is this: using our tool-kit, how can we describe events in an adjacent, co-moving frame from measurements made from our rest frame, or visa versa. An "event" is something that happens at a particular place (with x, y,and z coordinates and time, t). Events could be "happenings" (like an explosion or a light turning on) or the location of the edges of an extended object (like the coordinates—location—of the two ends of a stick at a particular time).

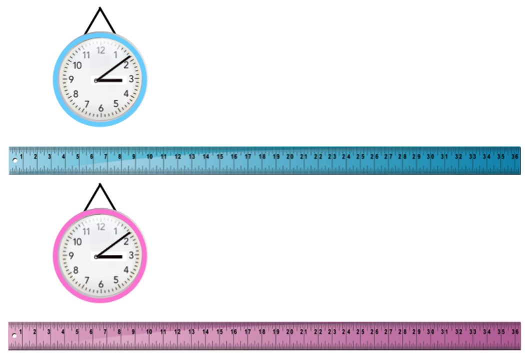


Figure 14.6: rulersclocks

The Moving Sidewalk

You've all been there: the big airport with miles of walking and little time to get from one gate to another. A long time ago airport planners found the solution to your limitations: the moving sidewalk.

Let's define some terms. Instead of calling one frame the "moving frame" and another the "rest frame"[20] we'll refer to the **Home Frame** and the **Away Frame**, or **HF** and **AF**. In this way we avoid any mistakes of language (or physics!). So for our story:

- For G, K, and A, the boat, B, is the Home Frame and the shore, S, is the Away Frame.
- For N, B (the boat) is the AF and S (the shore) is the HF.
- There's no Home Advantage.

[20] Since Galileo, we all agree that "moving" and "rest" don't make any sense.

Figure **??** shows two resting travelers in the airport watching the weary road warrior just standing there on the moving sidewalk that's going at a constant speed with respect to the airport, and hence, the Couch-People.[21] Another bit of terminology: I'll always use **u** to be the velocity of the Away Frame so u is to be the speed of the moving sidewalk.

[21] If we remember our boat, Newton on the shore is like the Couch-People at the airport.

Pencil 14.1.

Solution

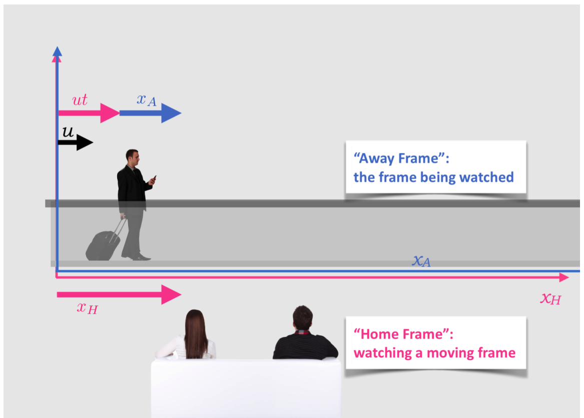


Figure 14.7: From the perspective of the airport (HF) the moving sidewalk (HF) is moving to the right with speed u. The traveler is stationary in the AF. Let's assume that the sidewalk (the AF) moves at a speed of $u = 2$ m/s relative to the airport (the HF). WearyTraveler's stationary foot is 2 meter from the origin of the moving sidewalk's frame, so $x_A = 2$ m. Trick question: after 2 seconds, how far has WearyTraveler's stationary foot moved within the AF? (He's standing still.) Next, the real question: what is x_H after 2 seconds, the position of his foot relative to the airport?

Did you get that $x_H = 6$ m?

What we've just done is called a "coordinate transformation." We've expressed an event that happens in one frame in terms of coordinates in another frame. In order to solve this simple airport problem, you invented a formula in your head:

Pencil 14.2.

$$x_H = x_A + ut \tag{14.1}$$
$$x_H = 2 + (2)(2)$$
$$x_H = 6 \text{ m}$$

Equation 14.1 is the mathematical expression of the **Galilean Transformation**. Actually, there' s one more piece to the Galilean Transformation, and that's the following equation:

$$t_H = t_A \tag{14.2}$$

and therein lies a subtle problem for later.

Definition: Coordinate Transformation. The conversion of the coordinates of an "event" in one inertial frame expressed in terms of another inertial frame.

Wait. *If I'm on a moving sidewalk, I know that I'm moving and that the airport is not.*

Glad you asked. *This is a difficult tool to conceptualize, but it's at the heart of Galileo's original notion, just brought up to modern transportation machines. People are not the crucial factors. Just the laws of mechanics. But you have all been in a moving rest frame and forgotten about it. How about a long airplane trip. Hours at cruising altitude, at night, over the ocean, and everyone's got their windows closed. You're simply in a room and cannot tell that you're moving at hundreds of miles an hour. Drop your pillow? It falls directly to the floor, not behind you. There's nothing you can do to show your motion. In fact, in your room, you're a part of a solar system that's moving more than 500,000 mph and don't know it! But don't pick up that pillow and use it because it's dirty down there.*

Now we get to the nub of it. Let's pretend that there were two Newtons. . . one who lived in the rooms at Cambridge University and the other who lived in a spaceship traveling at a constant speed relative to the British university's campus. They both do the same pendulum experiments and both reach the same conclusions about forces. And the form of the equations that they invent look exactly the same. The Cambridge-Newton uses F to mean force, a to mean acceleration, and m to mean mass, so his Second Law looks like

$$F = ma.$$

The spaceship-Newton uses $\clubsuit$ for force, $\diamondsuit$ for acceleration and $\heartsuit$ for mass and his Second Law looks like

$$\clubsuit = \heartsuit\diamondsuit$$

See that the *form* of the equations are the same? This is an important mathematical idea and the word **Invariance** is used to describe a formula that doesn't change form after some change in coordinates has been made on it. This is another way to say that every experiment performed in either frame would give the same results, and so no experiment could determine a state of motion or of rest. Our two Newtons would agree on the physics because their equations have the same form with the same definitions for the terms.

Definition: Invariance.
Something is invariant with respect to a change of its coordinates if the coordinates are modified and the form of the formula stays the same as before.

We say that **Newtons Laws are Invariant with respect to a Galilean Transformation**. Hold that thought, let's chase a beam of light.

14.4 The Paradoxes of Electromagnetism

When Einstein was a teenager he wondered what it would be like to look at a clock as you move away from it near the speed of light. He knew that light traveled at a large, but finite speed and so presumably as you moved away you'd see the time that the clock *was*. . . when the reflected light from its face bounced off on its way to your eye. What if you were traveling at the speed of light? What would happen then? This strange question stayed with him for a decade.

While he was working at the Patent Office, he and some friends would regularly meet in evenings and discuss matters of interest to them. They were a book-club of sorts[22] and with tongue-in-cheek they dubbed themselves the Olympia Academy. Later he credited the "Academy's" late-night, deep-dives into philosophy, mathematics, and physics as helping him to work through puzzles that bothered him for many years. It was during this time, he changed the world with three pieces of work in one amazing year:

[22] Meliva stayed home with the baby, while Albert played with his friends.

1905, Einstein's *Annus Mirabilis*

Remember that he has no doctorate in 1905... he's got a physics hobby, a regular "9-5" job, and an unused teaching diploma.

One. After a paper of his early doctoral thesis work, he published his second academic paper in the prestigious German journal, *Annalen der Physik*. It was about light and how it interacts with matter. We'll talk about that later when we get to the quantum theory, but suffice to note that this *March 1905 paper creates Quantum Mechanics*.

Two. In May of that year, he sent another paper to the same journal, this time related more closely to his slowly forming PhD thesis. All this *May 1905 paper does is convincingly demonstrate that atoms exist*. 2000 years of dispute, essentially settled in this one simple calculation. That's all.

Three. In June of that same year, he sent yet a third 1905 paper to *Annalen der Physik*. This was on one of his favorite topics, electromagnetism, and it's entitled *On the Electrodynamics of Moving Bodies*. In this *June 1905 paper he shows that Galilean Relativity is not correct*—the birth of Special Relativity.

The contents of this last paper are our focus here. He had never gotten his adolescent image of traveling at the speed of light out of his mind and it motivated him to look more closely at the equations of Maxwell once he had enough experience to understand them thoroughly. Electromagnetism got under his skin as he realized that has mathematical subtleties that reveal actual logical paradoxes.

1905 has ever since been called Einstein's *Annus mirabilis*... his miraculous year.

Racing Light

The speed of light was well-known by 1900 or so. Remember, we reserve the symbol c to represent this very special number, and in scientific units, c just about $c = 3 \times 10^8$ m/s or $c = 671,000,000$ mph, big in any units![23] Every experimental physicist learns that light travels about a foot in 1 nanosecond (1×10^{-9} s), so how long it takes for an electrical signal to go from one spot an electronics rack to another can be estimated by eye.

[23] Actually, it's $c = 299,792,458$ m/s or 670,616,629 mph.

If you look in the mirror a foot away, then the light that reflected from your face to the mirror and back into your eyes is stale—you're seeing at what your face looked like 2 nanoseconds ago.

This finite light speed is a lot more dramatic when we think about astronomical objects. For example, the Earth is so far from the Sun, that it takes 8.3 minutes for the light from its surface to reach us. If the Sun suddenly turned off, we'd have 8.3 minutes of sunlight before everything went black. Astronomers refer to this distance as an "Astronomical Unit" or AU. The Earth is 1 AU from the Sun, while Jupiter is about

5.2 AU away. They have also invented a self-explantory distance as how far light would travel in a year: a "light-year," which is 9.4607×10^{15} m. So a single AU is $1.58128451 \times 10^{-5}$ ly.[24]

The nearest star to our Sun, Proxima Centauri, is in the constellation Centaurus, the Bull and about 4.2 light years from us. I'm writing this in 2015, so just about now any Proxima Centurions with their TVs tuned towards Earth are just seeing the last of Oprah Winfrey's long-running TV show and the first episode of *Game of Thrones*.

Andromeda is nearest galaxy to ours, and one that probably looks most like the Milky Way. When you look at it,[25] you're seeing the image of what it looked like 2.5 million years ago, so it's 2.5 Mly from the Milky Way. The object that holds the record for being furthest from us is affectionately known as z8_GND_5296 which is 13.1 billion light years away and was only recently discovered. The Universe itself has been determined to be 13.7 billion years old, so this is an image from the universe's adolescence. You'll agree that the finite speed of light leads to interesting, if not completely awesome ideas. But light is tricky.

[24] We could also say that the Earth is 8.3 light-minutes from the Sun.

[25] In a dark sky, the Andromeda galaxy is a binoculars object.

Tricky Light

Let's list some common sense ideas on the highway:[26]

- Common Sense #1. Suppose we're on the highway traveling at 50 mph and another car goes by us on our left traveling at 60 mph. If we treat our car as at being at rest (like inside the hold of Galileo's ship), we would conclude that other car is not going 60 mph, but 10 mph relative to us.[27]
- Common Sense #2. Suppose your crazy passenger stands up and throws a 90 mph fastball forward through the sun roof of your car. Relative to the ground, the ball is seen then to be moving at 140 mph—superpowered major league material. But relative to the other car? That pitch is just junior varsity, only 80 mph.
- Common Sense #3. Suppose the speeding car beside you is actually a beam of light. Then you would expect that it would be moving at 671,000,000 minus 50 mph, or 670,999,950 mph relative to you.
- Common Sense #4. Suppose your athletic passenger whips out his laser pointer rather than a baseball and points it straight ahead. You would expect that someone on the ground would measure it to be moving at 671,000,000 plus 50, or 671,000,050 mph.

[26] Still...with the transportation analogies.

[27] This is so obvious as to be tedious.

Einstein found hints that something was strange about light. Here are two:

- If you were moving faster and faster away from the Bern tower clock its hands would appear to stop and freeze at the moment you reached light-speed. Its reflected light could not catch up with you and forever you'd see the time that it *was* when you passed that boundary. Time would appear to stop, which is kind of strange. This is Hint #1 that the speed of light is special.

- Remember that in an electromagnetic wave, it's the changing **E** field that creates a **B** field, and the changing **B** field that creates an **E** field. They mutually produce one another and mutually depend on one another and the propagating wave motion perpendicular to the **E** and **B** vectors is the result. Well, if like our cars you're speeding up alongside of a light wave—and reach the speed of light so you're now alongside of the light beam—the **E** and **B** fields would appear to go up and down in place, but not appear to propagate forward as a wave any more! That is a direct contradiction to the very clear mathematics in Maxwell's Equations. This is Hint #2 that the speed of light threshold is special.

Common Sense items 1 and 2 above are obvious. But these two hints suggest that *moving relative to light beams is strange*. And since the speed of light is somehow "hard-wired" into Maxwell's Equations, there's simply no good way to deal with situations like these. Before working our way out of this, let's make it even worse:

Definition: Lightyear..
The distance that would be traversed by a beam of light in one year. It is 9.5×10^{15} meters, which is approximately 6 trillion moles.

It Gets Worse

If that's not bad enough, let's think about two situations that Einstein actually writes about. Let's suppose that we have two lines of electric charge, one of positive charges and one of negative charges as in "Situation #1" at the top of Fig. 14.8 and place a positive charge, Q next to them. Since there are as many positive as negative charges, Q will feel no electrostatic force. Now let's have both Q and the negative line of charge move to the right with respect to the page (still no electrostatic force) with a constant velocity, v. In its rest frame, Q would see that the positive line of charge is moving to the *left* (along with the book), which is a current in Q's frame. Since a current produces a magnetic field, **B** into the paper at Q's position, and since Q has a velocity to the right, it would feel a force, up.[28]

[28] Remember, that the force of a charge in a magnetic field is $F = QvB$. And your right hand.

Now, let's reverse the situation as in Situation #2 in Fig. 14.8, without seeming to change anything but the relative motion. Leave Q stationary with respect to the page, so it has no velocity in that frame and hence even if there were a magnetic field nearby it would not experience a force. So start the *positive* line of charges moving with speed v to the *left* relative to the page. That creates the same **B** field as before, but since the charge has no velocity, it would not feel anything.

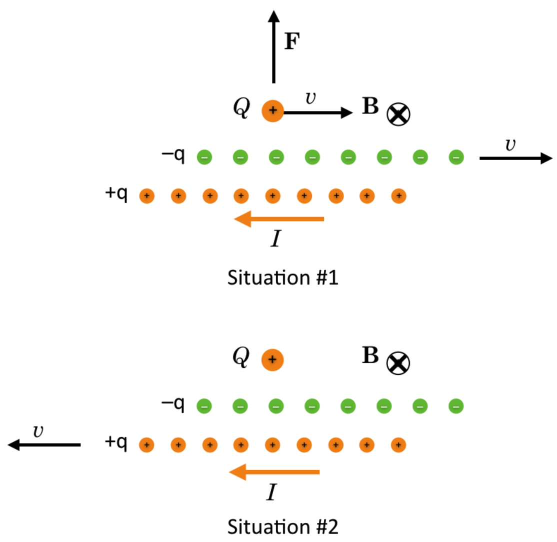


Figure 14.8: chargeparadox

Wait. *The situations aren't really different. In the first one, Q is moving and the positive line is stationary. In the second, Q is stationary and the positive line is moving. They only differ in their reference frame interpretation.*

Glad you asked. *You got it. The two situations only differ by the who is considered moving and who is considered stationary. Such co-moving frames in our Galileo-Newton discussion of mechanical systems didn't lead to any physical differences. But when a magnetic field is involved, Q experiences different outcomes just by interpreting who's Home and who's Away. Yes! It doesn't make any sense!*

Uh Oh.
These are two different physical outcomes for situations that differ only by their relative motion.

Here's another. Remember the magnet and the coil? When we push the magnet through the coil, a current flows in the wire. No battery, just Faraday discovering the generator. Now, suppose we leave the magnet still and move the coil over it. What happens? The same thing happens. The two relatively moving frames give the same physical result. So what's odd about this?

Let's ask what happens *physically* in each situation. When the magnet moves through the coil, the circles of wire capture the changing magnetic field. The changing **B** field creates an **E** field in the wires and that **E** field in the wires causes a force on the charged electrons in the wires, and so they move... which is the current that we see.

In the second situation, the wires—including their electrons—are moving toward the field of the magnet. *Now* the electrons in the wires have a velocity *and* they see a magnetic field, so from $F = Q_e v B$ the electrons experience a force... but this time it's a force due to **v** and **B**, *not an* **E**.

> Uh Oh.
> These are two identical physical outcomes for situations which have entirely different causes arising simply from the interpretation of who's Home and who's Away.

14.5 The Postulates of Relativity

Newton's Laws of mechanics gracefully flit about among co-moving inertial frames without any modification. That made sense. But comparing the consequences between co-moving frames when light is involved either leads to logical absurdities or an abandonment of Maxwell's beautiful formalism since there is no messing with c in Maxwell's Theory. Einstein was having none of that. Others also worried about these matters, but they were unwilling to go as far as our young clerk who was offended that the two best theories that explained all known phenomena should act so differently when viewed from co-moving frames. So he resolved to make them both behave.

He thought that Newton's Laws and Maxwell's Equations should not behave differently when viewed from co-moving inertial frames of reference and he resolved to fix this ugly discrepancy. This had been rumbling around in his mind for years. What he wanted was a way to expand on Galileo's notion that mechanical phenomena are blind to steady motion and promote optical phenemena to that level of consistency as well. In his 1905 paper he enunciated this idea and called it a postulate, the "Relativity Postulate": "...the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good."

It says: we require *all phenomena*, mechanical *and* electromagnetic, to work the same in all co-moving intertial frames. So in our example of Galileo and Newton's pitch-and-catch machine above, we could reasonably add that the microwave oven that they use to prepare popcorn for their spectators works the same on the shore is it does on the boat and that the rules that describe its manufacture would be the same as well.

Before we work out the physical consequences of this, let me make a bigger point about the philosophical consequence of the First Postulate, one that has guided all of science ever since Relativity became accepted. If the only tools you've got are mechanical and electromagnetic, then you can **never** tell whether a frame is stationary or moving, relative to any other frame. Never.

> **Wait.** *So, if there's a fixed frame of reference that has no motion, we could never know it?*
>
> **Glad you asked.** *Yes, that's true, but we need to take that idea and make it into a principle of knowledge—a statement about Reality, itself!*

The importance of the need to make a measurement is now a criterion of science. If you can't make a measurement of some idea, Einstein's work has convinced us all, then you cannot treat that idea as real. What's real in 20th and 21st century science are only those things that can be measured. Say that again:

> **If you can't measure it, it can't be real.** Period.

Let's call this the **Reality Postulate**. The advent of Quantum Mechanics only reinforced this idea and since he was also instrumental in ushering in that most strange theory, Albert Einstein has to be considered as among the foremost *philosophers* of all time, as well as scientists.

The consequences of the Reality Postulate turn into a criterion for what is and what isn't a scientific statement. If your theory contains statements about how things are, but those things cannot be measured, then you're not doing science. Obviously, this hardens the separation between religion and science, but we'll not go any further there.

Now comes the mathematical consequence. Maxwell's Equations say that the speed of light, c, is a single number related to the very properties of empty space which everyone believed was full of the ether. But if you could move relative to the ether and then make a measurement of the speed of light and find it to be different from that predicted by Maxwell, then you'd have detected that you're moving. That would be contrary to the Relativity Postulate, and in fact be responsible for all of the paradoxes that we discussed above.

So this forced him to a second postulate: "...light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. In our language the two postulates are:

1. All laws of physics—mechanical and electromagnetic—are identical in co-moving, inertial frames of reference.
2. The speed of light is the same for all inertial observers.

The import of this is that the Common Sense items #3 and #4 above need to be modified into *un*-Common sense situations:

- *un*-Common Sense #3. Suppose the speeding car beside you is actually a beam of light, and so moving at $671,000,000$ mph relative to the ground. Then you would find that it is also moving at $671,000,000$ mph relative to you also.
- *un*-Common Sense #4. Suppose your athletic passenger whips out his laser pointer rather than a baseball and points it straight ahead so that it would be moving at $671,000,000$ mph relative to your car. Someone on the ground would also measure it to also be moving at $671,000,000$ mph relative to the ground.

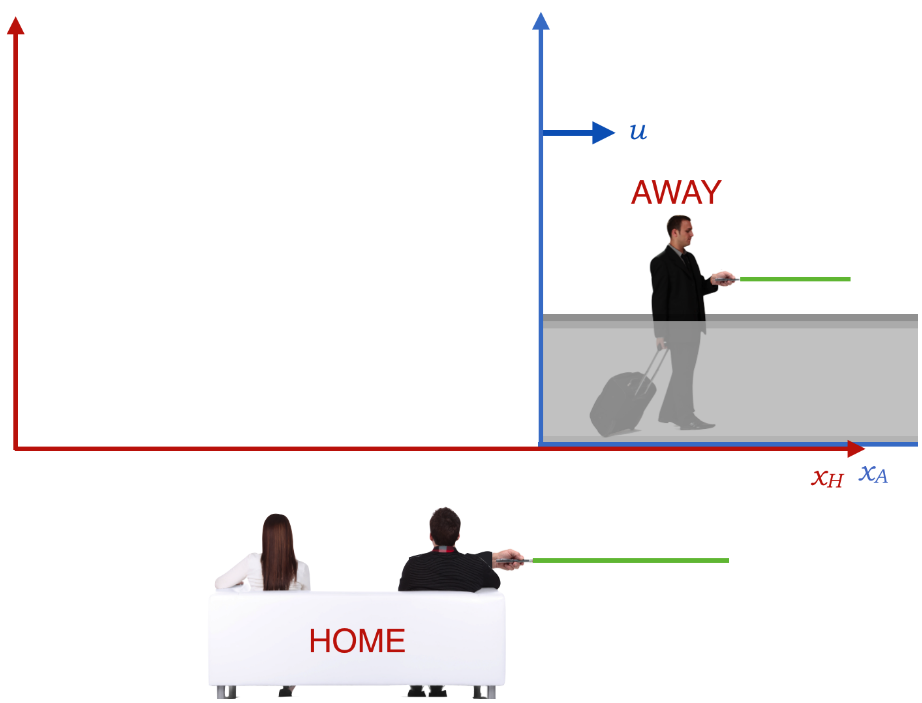


Figure 14.9: lasercar

This is exceedingly strange! What the second postulate requires we'll illustrate in Fig. 14.9 : this shows WearyTraveler playing with his laser pointer, aiming it ahead of him and CouchGuy in the airport with an identical laser pointer. Without bothering to get up, CouchPeople measure the speed of the light from their laser pointer to be c. Meanwhile, WearyTraveler measures his light beam to have the same speed of c, relative to the sidewalk frame.... But CouchPeople can also measure the speed of the light generated by WearyTraveler, moving past them at the sidewalk speed of u. They would *not* measure that WearyTraveler's light as moving at $u+c$...they would determine that WearyTraveler's light is *also traveling at that same value of* c as theirs and that he determines for his!

Let's take Einstein's leap into brand new territory. In the next chapter we'll unwrap the presents that the two Postulates of Relativity delivered to us and in the chapters after that, dramatically end World War II with the most famous equation ever.

Chapter 15

Special Relativity, Consequences

space and time aren't what they used to be

The stern looking Hermann Minkowski.

Hermann Minkowski (1864-1909)

"The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality." *Address to the Assembly of German Natural Scientists and Physicians on September 21, 1908).*

It was a considerable shock for the teacher to find that the pupil that he dismissed as lazy and inconsequential would turn out okay. Better than okay. Herman Minkowski put Special Relativity on a firm mathematical foundation and then gave a memorable speech including flowery phrases that every physicist knows about his newly named...*Spacetime.*

15.1 Goals

15.2 A Little Bit of Minkowski

You've probably never heard of Hermann Minkowski (1864-1909), but his influence on 20th century physics was imaginative and fundamental. He invented a language—plus a brand new kind of geometry—that actually simplified the physics of relativity before Einstein became known to the rest of Europe. While still anonymous, though, Einstein and Minkowski were well-known to one another and it's arguable as to who was more surprised at Einstein's breakthroughs. Einstein never expected to rely on a mathematician and Minkowski certainly thought that Einstein would never amount to anything. A match not made in heaven.

Minkowski was a child mathematical prodigy. His parents emigrated to Germany in 1872, when Hermann was 8 years old—they settled in the university town of Königsberg which provided ample opportunity for his unexpected talents to become apparent and be nurtured. He entered the University of Köningsberg at the age of 16 and received his doctorate in mathematics at 21. As a student he won a prestigious French mathematics competition and then moved up through the German and Swiss university systems as a specialist in the connections between geometry and number theory. The important time for our story is the period between 1896 and 1902. Minkowski began teaching at the Swiss Federal Polytechnic in Zürich in 1896 at the age of 32, the same year that Einstein began his studies there at the age of 17. Einstein registered in many of Minkowski's classes in the next four years but didn't endear himself to his mathematics instructor because of his habit of habitually skipping his classes. In 1949 he later wrote, "...the most fascinating subject at the time that I was a student was Maxwell's theory..." conceding later that

> "I had excellent teachers (for example, [Adolf] Hurwitz, Minkowski), so that I should have been able to obtain a mathematical training in depth. I worked most of the time in physical laboratory, however, fascinated by the direct contact with experience. The balance of the time I used, in the main, in order to study at home the works of Kirchhoff, Helmholtz, Hertz, etc."

That is, he skipped classes in order to study electromagnetism on his own.

Minkowski was shocked by Einstein's relativity paper. He'd thought along similar lines, but never quite got over the conceptual difficulties. But his surprise was only partly about the physics. Famously, he wrote to one of his students, Max Born (whom we'll meet later):

> "For me it came as a tremendous surprise, for in his student days Einstein had been a real lazybones. He never bothered about mathematics at all."

And he was right. Einstein had little patience with the over-mathematicazation of physics. Later when he was wrestling with his general theory of relativity, he lamented to Arnold Sommerfeld, a leading senior physicist from Munich,

> "But one thing is certain, never before in my life have I troubled myself over anything so much, and that I have gained great respect for mathematics, whose more subtle parts I considered until now, in my ignorance, as pure luxury! Compared with this problem, the original theory of relativity is childish ."

Minkowski married while in Zurich and he and his wife eventually had a family of two daughters. By 1907-1908 he had come to grips with relativity, having politely written to his former “lazy bones"

"Dear Doctor Einstein,
At our seminar in the W.S. we also wish to discuss your interesting papers on electrodynamics. If you still have available reprints of your article in the Ann. d. Phys. u. Ch., Vol. 17, I would be grateful if you would send us a copy. I was in Zurich recently and was pleased to hear from different quarters about the great interest being shown in your scientific successes.
With best regards, yours sincerely,
H. Minlowski"

From the work that grew of this reacquaintance came one of the more poetic and attention-getting physics talks ever given. On September 21, 1908, in the 80th annual general meeting of the German Society of Scientists and Physicians, in Cologne Minkowski presented a talk entitled “*Raum und Zeit*,” Space and Time. Born wrote later,

“...I went to Cologne, met Minkowski and heard his celebrated lecture ’Space and Time’...He told me later that it came to him as a great shock when Einstein published his paper in which the equivalence of the different local times of observers moving relative to each other was pronounced; for he had reached the same conclusions independently but did not publish them because he wished first to work out the mathematical structure in all its splendor. He never made a priority claim and always gave Einstein his full share in the great discovery.

After having heard Minkowski speak about his ideas, my mind was made up at once, I would go to Göttingen to help him in his work."

Planck had urged Einstein to attend, but he failed to do so. What an amazing event that would have been for the still patent clerk from Bern. The opening paragraph is famous among all physicists today,

> "M. H.! [ladies and gentleman!] The views of space and time, which I would like develop, have sprung from the experimental-physical soil. Therein lies their strength. They tend to be radical. Henceforth space by itself and time by itself, fade away completely into shadow, and only a kind of union of the two will preserve independent permanency."

15.3 The Second Postulate

In the last chapter, we developed Einstein's two postulates of relativity. The first postulate was a throwing down of the philosophical gauntlet: no phenomenon—neither mechanical (like Galileo's) nor electromagnetic—can be used to distinguish motion in a frame of reference moving at a constant velocity (an inertial frame) relative to another inertial frame. But the fun is all in the second postulate: that the speed of light is a constant, an *invariant*. Let's go to the airport.

✎

15.3.1 Time and Space Increments

As a million-miler on a U.S. airline, I spend way too much time in airports. As a physicist in an airport? Well, there's just too much fun to be had. We all enjoy the moving sidewalks as a visual example of relatively moving inertial frames.[1] So let me have my fun and allow me to use the moving sidewalk in our examples. Stay with me.

Figure 15.1(a) and (b) show CouchPeople and WearyTraveler at the airport. CouchPeople have a long lay-over and they're people-watching. WearyTraveler is on the moving sidewalk. Figures 15.1 (a) and (b) show two different times for his journey across the terminal.[2]

If we're in the airport with CouchPeople, we'd label it and them as in the Home Frame (HF) while the WearyTraveler is in the Away Frame (AF). We'll use this for all kinds of examples. Here's the first one.

[1] Okay. You might enjoy them as a way to get from one end of the terminal to another.

[2] Don't you hate the people who just stand on them? Especially those on their phones?

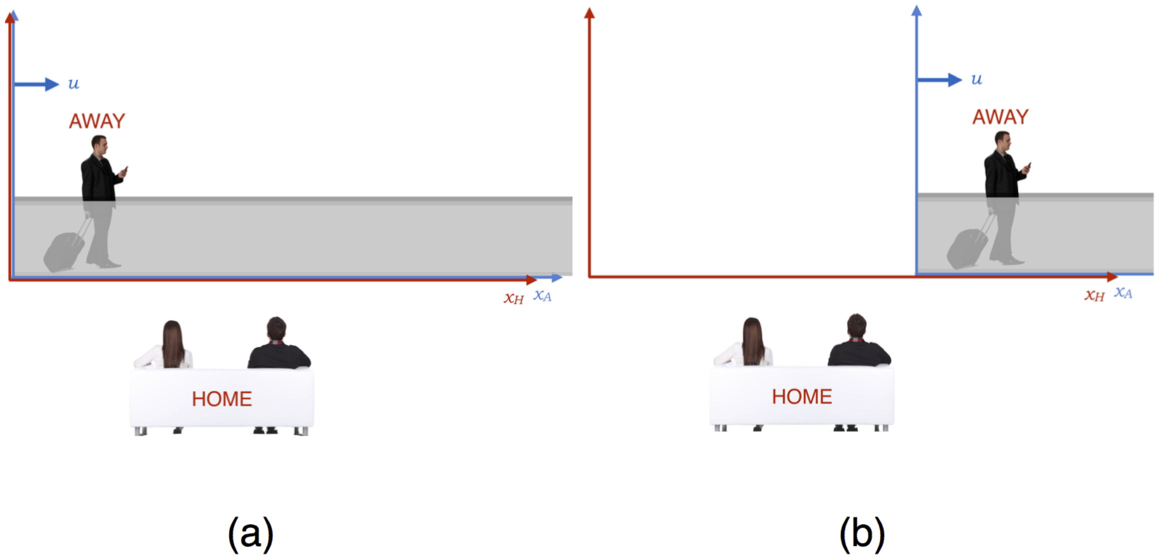


Figure 15.1: sidewalkintro

15.3.2 The Second Postulate's First Surprise

Figure 15.2 is a fake electronic device set up on a test bench that we'll pretend measures the time differences of light passing through the two photodetectors, A and D. What happens is that light goes through A and registers a photo-signal that passed through the cable to the oscilloscope input, B. The time of arrival is then registered on the screen, C. There is an identical unit next to it with another photocell, D with an identically long cable that sends its signal to input E, for time calculation and display at F. What's really going on is that internal to both oscilloscopes is a crystal clock that keeps regular time digitally and the displays at C and E are then presented as clock faces representing the number of pulses that each internal clock registers. In this case, the two devices have been calibrated so that they share a common start time.

Notice too that there are three rulers on the table, each a foot long. So the photo detectors are exactly 3 feet apart. So knowing the distance and then measuring the times, we could determine the speed of light if a beam shines from left to right.

And indeed, that's what we'll do. In Fig. 15.2 (a) the devices are ready for a signal and in (b) a common light beam has been shined on them both from left to right. Notice that they now show different times,

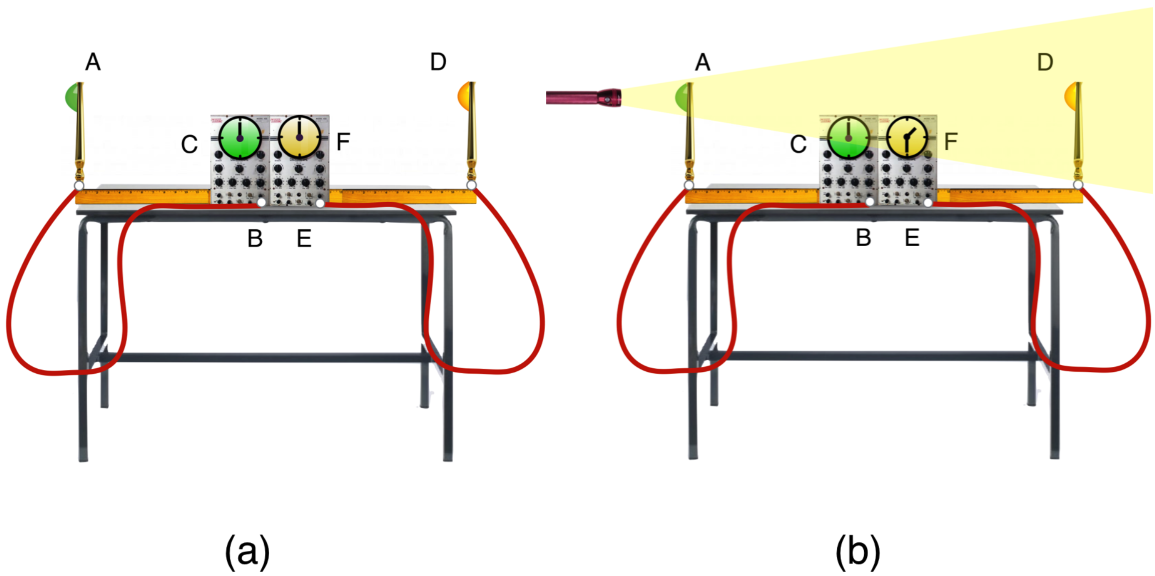


Figure 15.2: timedevice

representing the time that it takes for the beam to go from A to F. Since light travels about a foot in a nanosecond,[3] the time difference that will be registered between our two cartoon devices would be about 3 ns, which is easily discernible with modern electronics.

[3] That's just a rule-of-thumb that one learns in a physics laboratory.

What the Second Postulate says is rather astounding, which we can illustrate with our sidewalk and our fake device. Let's assume that both the CouchPeople and the WearyTraveler have built identical devices. CouchPeople set theirs up in the HF next to the sidewalk, while WearyTraveler sets his up on the sidewalk with him. Then, just like on our test bench above, a single beam of light is directed along the sidewalk, from left to right so that it passes through both sets of apparatus.

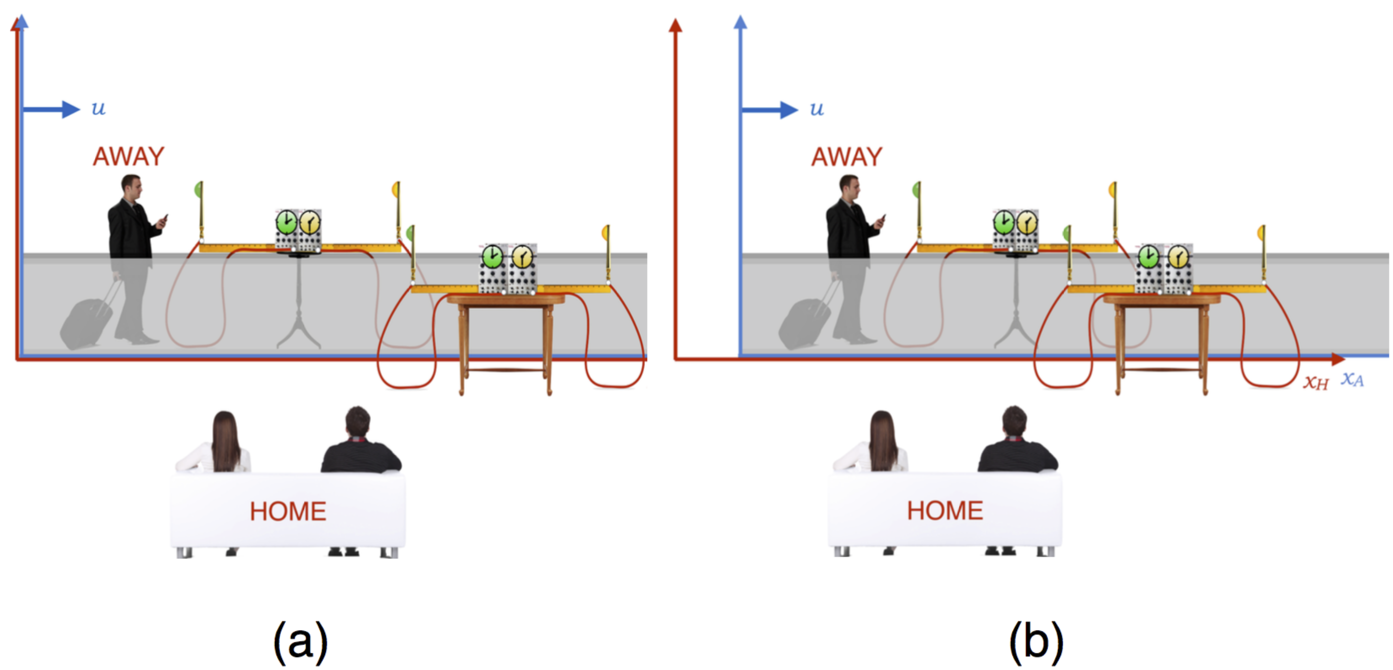
Wait. *Might the beam be slowed down or somehow affected by passing through one device before it gets to the other?*

Glad you asked. *Good question! In principle it might. But our airport people are good scientists and so they first set up their experiment in the airport...one after the other...and shined a light through them both. What they measured was that there is no affect—they measure the same speed for both.*

Figure 15.3 shows our equipment loaded up (a) and with the sidewalk having moved a bit to the right, (b). The experiment comes from shining the light, which we see in Fig. 15.4. Here's where the fun comes in. Let's ask three questions of our travelers:

1. What is the speed of light as measured by the CouchPeople for the HF apparatus?
2. What is the speed of light as measured by the WearyTraveler for his device in the AF?
3. What is the speed of light as measured by the CouchPeople... using the AF device... the one on the sidewalk in the AF?

Figure 15.3: sidewalksetup1



Question 1 is easy. We've already done it in the setup. The HF people measure the speed of light to be 3.0×10^8 m/s. What we all know and love as c.

Question 2 leads to a surprise. Even though WearyTraveler and CouchPeople are each sampling the same beam and even though WearyTraveler is moving away from the source of the light, they he measures

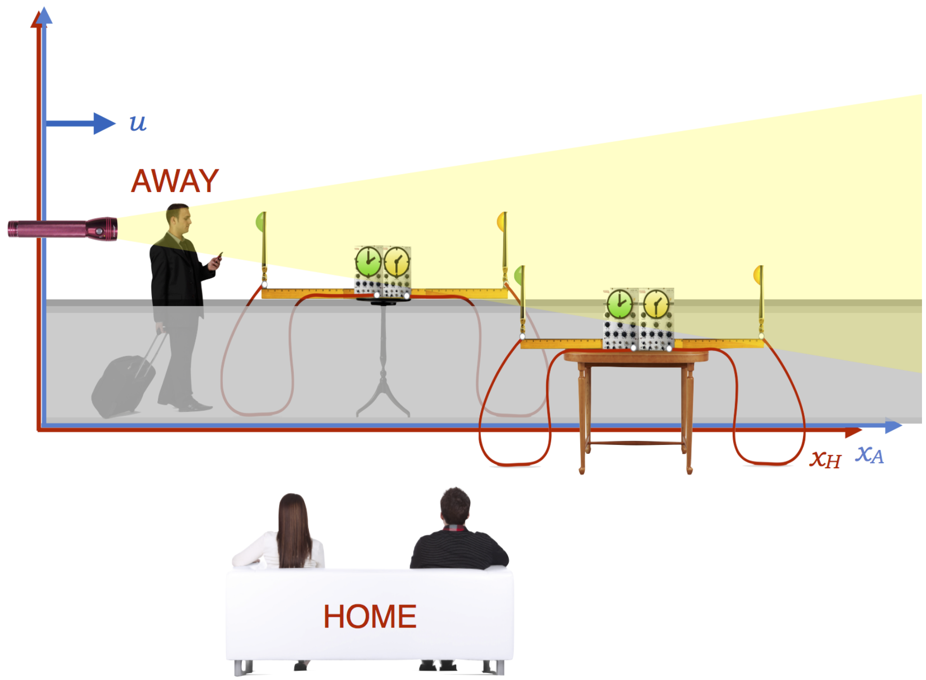


Figure 15.4: sidewalksetup2

the speed of light to also be 3.0×10^8 m/s! You might think that it would somehow have to go faster in order for him to get that same speed. But that's what Einstein's Second Postulate requires. But we're not done with Strange.

What about question 3? The CouchPeople would measure the speed of light for the machine in the sidewalk's frame to be...3.0×10^8 m/s. Now that's really disturbing and our second surprise coming from the Second Postulate.

Wait. *That's crazy! The sidewalk has no affect on the speed of light even though it's moving away from the source of the ight?*

Glad you asked. *Yup. That indeed, is one of the strange things about Special Relativity. Somehow we have to explain this.*

Yes, the Second Postulate suggests strange things about the world.

Wait. *I'm not done with you yet. The Second Postulate just that. It's not a statement of experimental fact. A postulate is only a proposition. A suggestion.*

Glad you asked. *What we'll see is that if we assume the postulate we can derive measurable facts about nature which are a consequence of the postulate and check them. If they work, then we shoudl accept the postulate. If they don't, then it was an interesting try, Albert, but no dice.*

Let's build a clock.

A Light Clock

Back to the sidewalk. Figure 15.5 shows the raw materials for another fake measuring device. A bathroom mirror like on the left side of the figure. Okay, two of them, mounted horizontally as shown on the right side. The mirrors are separated by a distance L and a little hole is drilled, H, to admit a burst of light from the laser pointer. The hole is quickly plugged and the light beam, B, just bounces up and down. We mount it up on the sidewalk as in Fig. 15.6 and WearyTraveler counts the round-trips of the light. Up-down, up-down, Tick-Tock. . . so, yes, it's a light-clock. The round trip time that the light pulse takes go up and down we'll call the increment of time measured *from within the Away Frame (the sidewalk)*, t_A.

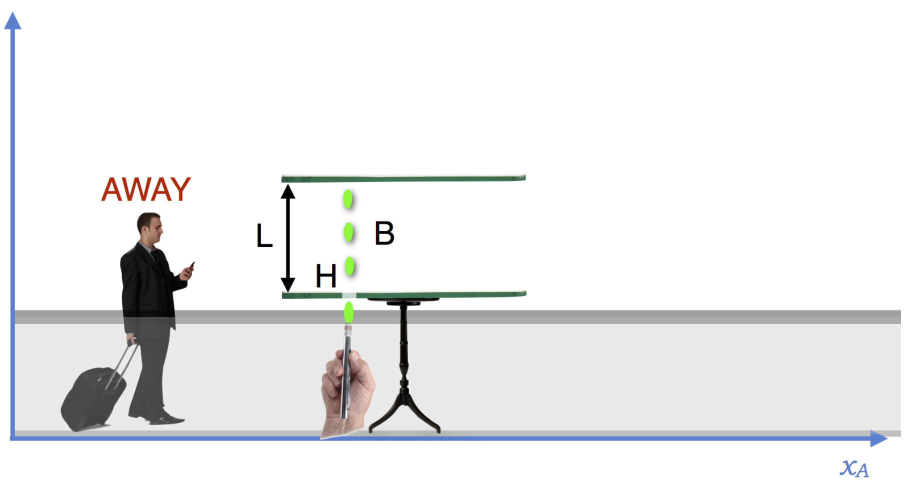


Figure 15.6: mirrorAF

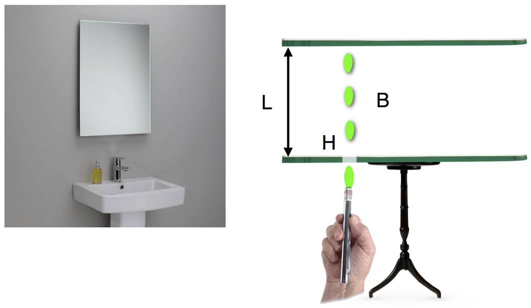


Figure 15.5: mirrors

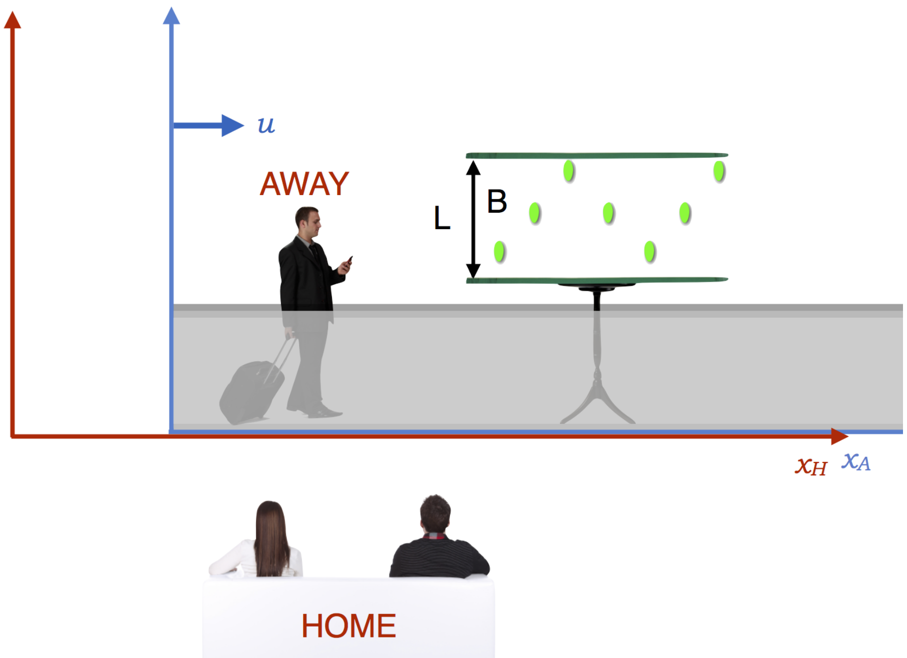


Figure 15.7: mirrorHF

What do CouchPeople see as the contraption moves by them on the sidewalk? The HF view of the ("moving") clock is different. The light pulse certainly has the same vertical up and down motion, but as it goes up and then comes down the sidewalk has moved *horizontally* and so we see a kind of triangular path as shown in Fig. 15.7. So for those of us on the ground, the pulse travels further than for the AF observer and the time that it takes to make a complete trip *as observed from the airport Home Frame*, up and down, we'll call t_H.

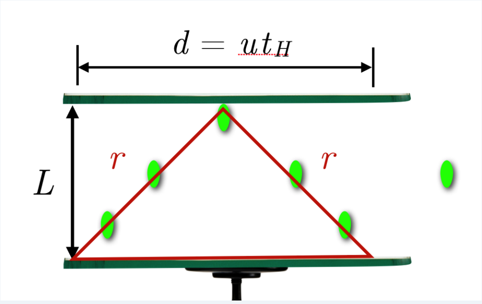
The distances traveled are different, so how do the two times relate to one another if the speed of light along the two different trajectories is actually the same in the two frames—which is what the Second Postulate says? Let's calculate it.

Time Dilation

Example 15.1

the light clock.

Question:



The figure shows the path of the light beam as seen by the HF. Remember, in the AF, the beam is just going up and down vertically, and so travels $2L$ in a full tick-tock cycle of T_A. We need to know the "tick-tock" cycle, t_H, as measured from the HF. The L vertical dimension is unchanged by the motion. The horizontal distance, d is related to the speed of the frame as shown, $d = ut_H$. Finally, the hypotenuse distances that the light actually appears to travel is r.

Solution: I'll sketch this. Let c_A and c_H be the speed of light in the AF and HF. Then:

$$t_A = \frac{2L}{c_A} \text{ and } t_H = \frac{2\sqrt{L^2 + (d/2)^2}}{c_H}$$

Solve each of these equations for L and set them equal to one another.

$$(c_A t_A)^2 = (c_H t_H)^2 - d^2 \text{ ...but } d^2 = u^2 t_H^2$$

$$c_A t_A = t_H \sqrt{c_H^2 - u^2} \text{ ...and if we then use the Second Postulate, } c_H = c_A = c$$

$$t_H = \frac{t_A}{\sqrt{1 - u^2/c^2}}$$

This is an amusing result. It means that a clock in an inertial frame of reference as observed from another inertial frame of reference would appear to keep different time by this factor:

$$t_H = \frac{t_A}{\sqrt{1 - u^2/c^2}} \tag{15.1}$$

$$t_H = \gamma t_A. \tag{15.2}$$

I've introduced a new quantity, γ, which is an important function in relativity called the "relativistic gamma function." (Say "gamma" to a physicist, and she'll know it to be this thing.) This situation in which time intervals would be measured by two observers to be different is called **Time Dilation**. It seems a crazy thing, except that it's truly the way nature works.[4]

[4] ...which we can now declare is the beginning of "crazy" in modern physics.

definition, time dilation

Let's also define a second useful quantity that we'll need it a lot, and that's the ratio of the velocity of a reference frame, u, to that of the speed of light. We call that "beta," β:

$$\beta = u/c \tag{15.3}$$

Since nothing can go faster than the speed of light (we'll see why in a bit), β can only be less than or equal to 1, or $\beta \leq 1$. So that the gamma function can be compactly written:

$$\begin{aligned}\gamma &= \left(\frac{1}{\sqrt{1-\left(\frac{u}{c}\right)^2}}\right) \\ \gamma &= \frac{1}{\sqrt{1-\beta^2}}\end{aligned} \tag{15.4}$$

Now while Eq. 15.4 looks complicated, we don't need to evaluate it for our purposes. Let's graph it and then we can refer to that plot for the whole story. Fig. 15.8 shows γ as a function of β. (Figure 15.9 shows it more precisely in the region $\beta < 0.6$ which might be useful for you.)

Wait. *So, how in my life does this matter?*

Glad you asked. *It matters in big ways and in small ways. Let's get a feel for just how fast, is fast!*

One of the fastest man-made objects might be a rocket with enough speed to reach the escape velocity necessary to break free of Earth's gravitational pull. That speed is 7.9 km/s which is about 20 times the speed of sound. What's the β for such a rocket?

$$\beta = \frac{v_{\text{escape}}}{c} = \frac{7.9 \times 10^3 \text{ m/s}}{3 \times 10^8 \text{ m/s}} = 0.00003$$

So look at Fig. 15.8 or even more usefully, Fig. 15.9. A β of 0.0003 is about the size of a single pixel at the most left-hand portion of that curve, so the γ associated with such a rocket is effectively just 1.0, in fact, it's $\gamma = 1.00000000045$. This means that a clock on the rocket would essentially keep the same time as a clock on Earth which is one of the examples of how relativity is not an everyday concern. Even if you're a rocket scientist.

But what about the electrons in your parent's old TV set? For them, β is closer to 0.5. Those electrons needed to be precisely aimed at the TV screen from the back and precisely scanned across it, But they move so fast from the electron gun at the back of the set that electrical engineers needed to take into relativity into account or Lucy and Desi would have looked funny. So, sure, small things move fast.

What about an object bigger than the whole solar system?

Quasars as an example...

So the relationship between two inertial frames, measuring a time interval is:

$$T_H = \gamma T_A \tag{15.5}$$

As β becomes very close to the speed of light, what happens? Gamma gets bigger and bigger and heads off towards infinity when the denominator of γ becomes very small when $\beta \to 1$, $\gamma \to \frac{1}{\sqrt{1-1}}$.

Wait. *So a clock in such a high-speed frame would run infinitely slowly?*

Glad you asked. *Two responses. First, it's more correct to say that a clock in the AF would appear to run infinitely slowly as viewed from the HF. And second, well, yes, but with a big caveat. We'll see that an object with a mass can never reach the speed of light. But if infinity is a prediction about Nature, absurdity is the cause. That rings true given the paradoxes that we just talked about!*

The velocity of light is obviously special. Remember how our CoachPeople measured the speed of light to be the conventional c for the WearyTraveler's machine? We just saw that time becomes a function of speed—it's warped as viewed across inertial reference frames: clocks appear to move slower. Since c is a speed, so it's a ratio of space and time, if time is warped in order to keep the speed of light constant, something must happen to space also! Hey, you're pretty smart! That's exactly what happens.

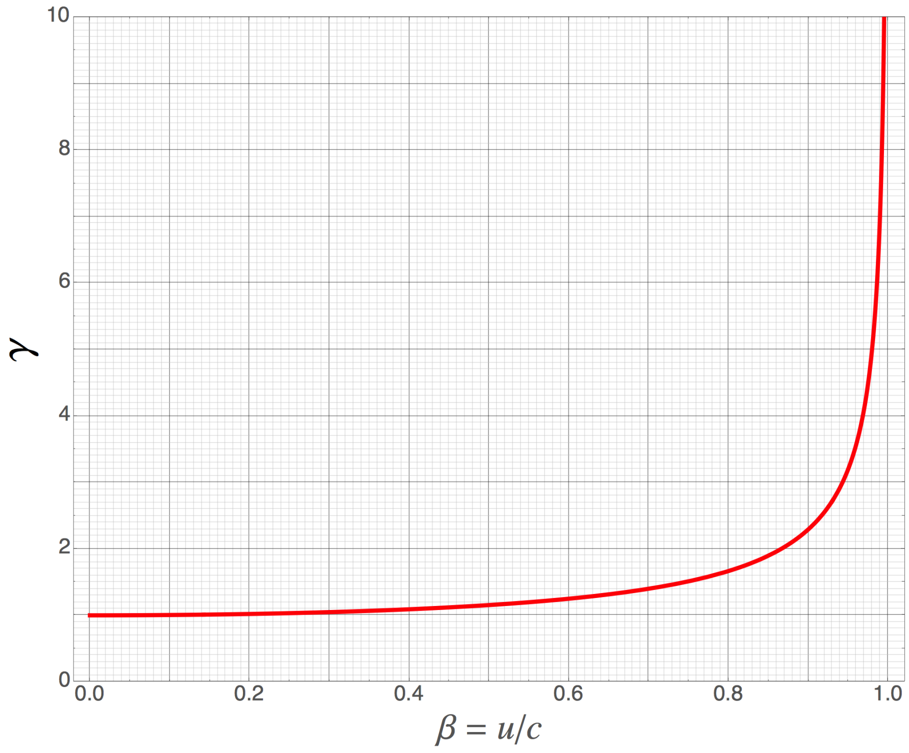


Figure 15.8: The “gamma function,” γ shown over the entire range of β.

Rather than writing out the derivation, which is a little complicated, let me just report the result. For WearyTraveler’s three rulers, which amount to 3 feet in his frame, CouchPeople would not see them to be 3 feet, but something shorter. Space also appears to be warped across inertial frames to the tune of:

$$L_H = \frac{L_A}{\gamma} \tag{15.6}$$

Here, as in Eq. 15.5, the Home and Away frames—this time for lengths—are related through the same γ function, but in the opposite way from Time Dilation. A HF observer would see that a length in an AF would appear to be shorter than that same length would be in the frame in which it’s at rest. Said another way, if we have a meter stick on the moving sidewalk—which would have the length of, well, 1 meter there—as viewed from the ground, that meter stick would appear to be *shorter* than 1 meter. This

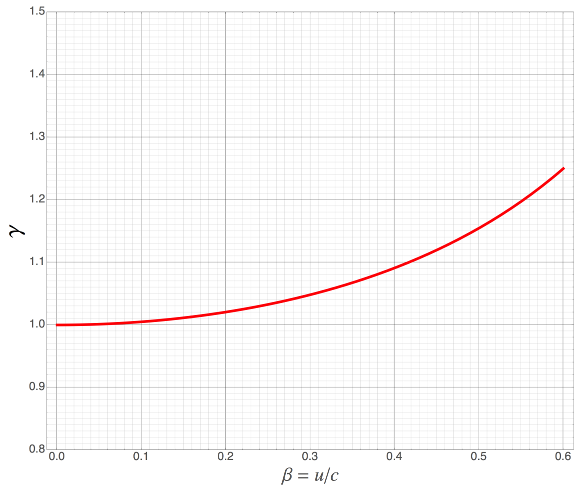


Figure 15.9: The γ function at lower speeds, for $\beta < 0.6$. Notice that the vertical axis starts at 1.0.

phenomenon of observers in co-moving inertial rest frames measuring lengths to be different by virtue of their relative velocities is called **Length Contraction**.

Wait. *Why "dilation" and why "contraction"?*

Glad you asked. *Clocks in the AF appear to run slower as measured from the HF, and so the interval between tick-tock is longer, hence "dilation." You have your eyes "dilated" at the eye-doctor and pupils get large. A length in the AF appears to be shorter when measured from the HF, hence "contraction." You can see this in the two formulae since $\gamma > 1$.*

definition, LC and TD

examples, all using the sidewalk with explicit use of rulers

So there we have it. What does our c-measuring device actually do? First, we use 3 rulers as a length and then we use the internal crystal clocks of the oscilloscopes to measure time. But each of these quantities are messed with from the perspective of the airport frame as compared with the sidewalk frame. So if we fashion a ratio of

$$\frac{\text{(distance in the sidewalk, measured by HF)}}{\text{(time in the sidewalk, measured by HF)}} = \frac{L_H}{T_H}$$
$$= \frac{L_A\gamma}{\gamma T_A}$$
$$= \frac{L_A}{T_A} = c.$$

That last line comes from WearyTraveler's measurement. So the Second Postulate implies a consistency, if not a troubling one.

15.4 Coordinate Transformations, 2

When Einstein forced his two postulates onto Maxwell's Electromagnetism, the outcome was a new set of coordinate transformations, which by now you'd not be surprised to learn treat time in a new way. Let's reprise the Galilean Transformations, Eqns. 14.1 and 14.2:

✎

$$x_H = x_A + ut \tag{15.7}$$
$$t_H = t_A \tag{15.8}$$

Einstein found a set of transformation equations for space and time that had been previously found by Hendrik Lorentz (1853–1928)[5] who had been manipulating Maxwell's Equations also, but with a very different intention and with the firmly held belief that the ether was an absolutely stationary frame of reference. So traditionally these equations are called the Lorentz Transformations. We'll not use them explicitly, but we can learn a lot by just looking at them.

[5] We've already "met" Lorentz when we worked on the forces that **E** and **B** fields apply to electrical charges.

$$x_H = \gamma(x_A + ut_A) \tag{15.9}$$
$$t_H = \gamma\left(t_A + \frac{ux_A}{c^2}\right) \tag{15.10}$$

Look at Eq. 15.9. It looks familiar and indeed, but for the factor of γ, it's identical to Eq. 15.7. Either from Fig. 15.4 or (and?) from the definition in Eq. 15.4, we see that if the relative speed of the AF as compared to the HF is very much slower than c, then γ is for all practical purposes, extremely close to 1:

$$\gamma(u << c) \to 1.$$

If that's the case, then we recover the Galilean Transformation for space coordinates. How about the time transformation?

Equation 15.10 is strange at first since it depends on the space coordinates. It says that the time intervals as measured between two inertial frames would be different, but since we've already gotten our heads around Time Dilation, perhaps this is not too surprising. Further, I'll bet at this point you know what would happen if again, the relative speed between frames is very slow. In that case, $\gamma \to 1$, but also the second term in Eq. 15.10 has the quantity $\frac{u}{c^2}$ in front of the x_A, which when $u << c$ is very close to zero, so we get back that $t_H = t_A$, that unquestioned presumption in the Galilean transformations.

15.4.1 Maxwell's Equations, 20th Century Edition

The transformations of space and time were what Einstein needed in order to make good on his Postulate 2 promise. But now let's think more specifically about electric and magnetic fields in relatively moving frames of reference. Instead of a ruler or a clock on our sidewalk, let's load up a magnet and ask how an AF observer (riding with the magnet) and a HF observer (on the ground) would describe its magnetic field. They would both rely on Maxwell's equations which include x and t variables. But in order to separately apply them, the HF observer would take the field equations and transform the space and time variables according to the Lorentz transformations. Remember this would be required in order to maintain the constant speed of light between the two reference frames. Upon making that transformation, something remarkable happens.

Let me show off for a minute. Please? I want to write one of Maxwell's four equations for just one direction in space. Afficianaos will write this slightly fancier. But I want to make a point. Remember the fact that changing a magnetic field in time creates an electric field? (The magnet moving through a coil of wire, setting up a current?)

Here is a simplified version of one of the equations that describes this phenomenon.

$$\frac{\Delta E}{\Delta y} - \frac{\Delta E}{\Delta z} = -\frac{\Delta B}{\Delta t}$$

There. **E** and **B** are functions of space and time. What you see is on the right hand side how if a magnetic field, B, changes in time (the Δ's mean "change of" remember?), then the result is an electric field that's changing in the by the amount of$y$and$z$. So there are space and time coordinates all over Maxwell's description of light.

If we're observing some electromagnetic phenomenon on the sidewalk from the airport, the constancy of the speed of light forces us to modify those space and time coordinates and this has physical consequences. Figure 15.10 is a simple example. WearyTraveler has a magnet with him in his reference frame. The field lines drawn on the picture are those of a bar magnet and that's what *he sees*! What do CouchPeople see? Remember the x's, y's, z's, and t's in Maxwell's equation above? We have to transform them in order to describe what the CouchPeople see. And it's weird and wonderful.

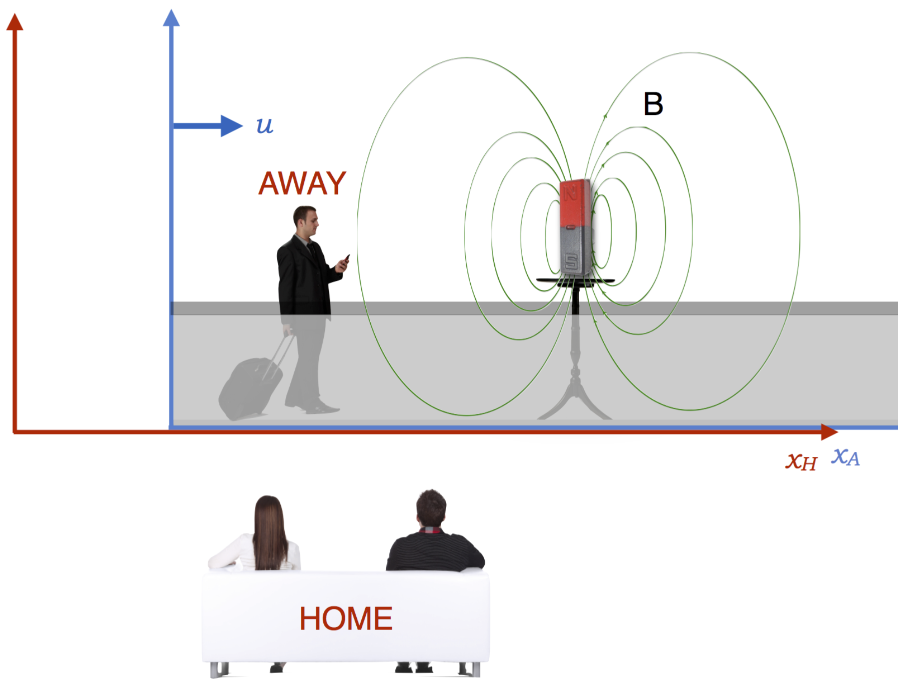


Figure 15.10: sidewalkmagnet

A **B** field, say from our magnet in the AF looks to the HF to be *a mixture of an E and B field*! An electric field in an AF, **E** (like that emeging from a stationary electric charge) when transformed into the HF appears as a *mixture of both an E field and a B field.* So while our AF, WearyTraveler observer sees only a magnetic field from his magnet, our CouchPeople observers would say, "No!" They would see both a magnetic *and* an electric field!

Relativity does it again. Just like it has taught us to merge space and time from separate concepts to a single spacetime. Space and time have no separate meanings any more. It also forces us to conclude that there is really no such thing as an absolute, permanent electric field or an absolute, permanent magnetic field since relatively moving observers will disagree about their natures. What's "relativistically" appropriate is the **Electromagnetic** Field—a single entity—which will manifest itself in different mixtures of **E** and **B** depending on the frame of reference from which it's observed. Spacetime and Electromagnetism as combined things is what has meaningful existence.

What's further a surprise—and indicative of Einstein's Postulate #1—is that Maxwell's Equations themselves turned out to be perfectly invariant with respect to co-moving inertial frames. . . but *Newton's Second Law* is not. Maxwell wins, and Newton loses in Special Relativity! A different, actually pretty complicated relationship needs to be substituted for good old $F = ma$ in order to be relativistically correct. As you might expect, for low speeds of co-moving inertial frames of reference, that more complicated relationship reduces to regular, old $F = ma$ when $\beta << 1$.

This mixture of the individual electric and magnetic field vectors *solves all of the original paradoxes* that we met in the previous chapter. All is well with Maxwell's equations and light, but mechanics turns out to be subtly odd.

15.5 Invariant Intervals

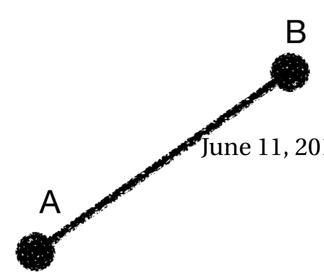
Now you can imagine why this theory is called "relativity." You've heard it said all the time: "Everything is relative." But it's not true! And even Einstein himself disliked the name "special relativity." He wanted to call his theory "Invariant Theory" because for him, what was most important was what stays the same between two relatively moving observers: the laws of physics and the speed of light. But "relativity" stuck. Let's think harder about this.

15.5.1 Space Invariants

Let's do some geometry and take something that's simple, and find out that it's also pretty.

Go to your wall with a pen and draw a straight, diagonal line of length L, from point A to B, as in Fig. 15.11. Mom won't mind, since you're probably not at home. Now take the pink coordinate axes in Fig. 15.12 and place it with the origin at A and the x axis horizontal. From Pythagoras' Theorem, you can calculate the length of your line as

$$L^2 = X^2 + Y^2 \tag{15.11}$$



Now take the yellow set of axes shown in Fig. 15.13 and measure the length of the line again. You'd say that it's

$$L'^2 = X'^2 + Y'^2 \tag{15.12}$$

right? But is the line any different? Of course not, so clearly

$$L = L'.$$

We could add more similarly rotated coordinate systems as in Fig. 15.14 and while the individual x and y coordinates of B would be different, they would all yield the same length, $L = L' = L'' = L''' = \ldots$ etc. We would say that the length of the line is **invariant** with respect to a rotation of the coordinate system. This is an important property of Space: lengths in space are constant, regardless of the reference frame from which they are viewed. Because, all of those different ruler combinations are just different reference frames.

One more step. Let's rotate the line and the rulers all about point A so that the ruler coordinate axes all overlap as in Fig. 15.15and you can see that the B-end of each rotated line traces out a circle, which I'll call the **Invariant Curve**.[6] If we took this to 3 space dimensions, the Invariant Curve would actually be an Invariant Sphere. If we went to 4, or 5 or more space dimensions, the Invariant Curve would be an Invariant *Hypersphere*. The distinguishing feature of all of these curves, even beyond our familiar three dimensional space, is that the invariant quantity always looks like:

$$L^2 = x^2 + y^2 + z^2 + a^2 + b^2 + \ldots$$

The important thing about are those + signs for all of the coordinate combinations. Such a multidimensional space is called a **Euclidean Space** since it obeys all of the rules of geometry going back to Euclid in Hellenistic Greece.

[6] Remember that Eq. **??** is actually the equation of a circle with radius L.

15.5.2 Spacetime Invariants

But we're working in relativity now where we're beginning to mix space and time and so the natural question is whether there's an Invariant Curve...in spacetime? Is there some spacetime "length" that is always constant, regardless of the coordinate system, which is the same as asking whether there's something that *stays the same* even for co-moving, inertial observers?

Maybe since I've asked the question, you know that the answer is "yes" but it's a very subtle point. Let's make a baby. Um, methaphorically.

Causality and Babies

We need to set up a "geometry" of spacetime, which we'll represent as Cartesian coordinates but instead of axes which are both space, we need one of them to represent *time*. Now there are two problems with this:

- Time is not a space-ish coordinate, in the sense that time's unit—*seconds*—is not the same as length's unit, *meters*. So we need to get them both on the same footing and we'll choose space dimensions (meters) as our standard and turn our times into space-lengths by multiplying by c. In that way, ct (length/time times time) is has the units of length, while still *functioning* as a time coordinate. We'll plot ct as the x-axis in our spacetime diagrams just like we have before, except now it will be a length. So for a time interval of 1 second, the time-as-space amount would be $c \times 1$ second or $(3 \times 10^8 \text{ m/s}) \times 1 \text{ s} = 3 \times 10^8$ m. This is also the distance that light would travel in 1 second: 1 *light-second*.[7]
- Second, we can't really draw more than two dimensions on a flat surface, so we'll abstract all of the space coordinates into just one direction and plot that against ct.

Figure 15.16 is a representation of our spacetime axes, just like we've used before.

Now back to our question: remember what an invariant curve is. For different reference frames, that curve is what all observers would agree on. For our drawing of the line on the wall, each pair of rulers-as-axes represents a different observer. They differ in their rotational relationship around their common origin and each "observer" agrees that the line has the length L and that the ends of the different "L's" form the locus of a circle when the observers are put on the same footing. What's an invariant curve in *spacetime*?

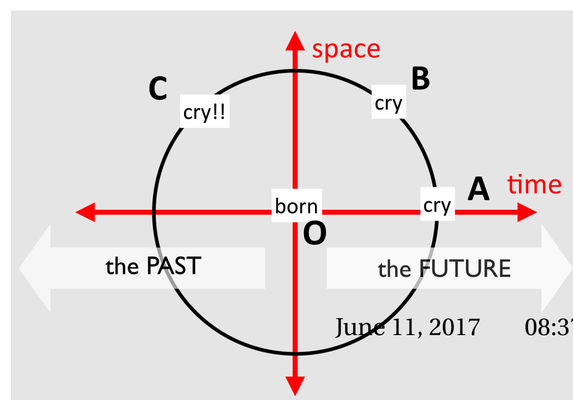
The most natural thing would be to try a circle, just like we had for space, but now in spacetime coordinates, as shown. Since the time axis is horizontal and since we put $ct = 0$ at the center, everything to the right of the center is the *future* and everything to the left, is the *past*.

We'll start our clock (where $t = 0$) at the moment of the birth of a beautiful boy. We'll (Curiously) call him Benjamin (Button). We'll not do this on the airport sidewalk, but rather in a hospital room, Benjamin is born and then without moving in space, he cries. That's represented as point **A** on the plot: his space coordinates stay constant and his young time duration increases slightly.

Let's drive by the hospital in our car and think about what we might observe from within our frame of reference, the HF. The hospital is now the AF and is moving by us in the other direction.[8] We peer into the room and see that Benjamin was indeed born, but when he cries, in the HF coordinate system, he cries at point **B**, where he appears to have moved in space, so vertically on the diagram (since the whole hospital appears to us to have moved in space).

But now let's imagine another car-based observer moving the other direction relative to us. The same blessed event happens, but for that observer the hospital (the AF) is moving in the other direction. We

[7] You've heard of "light-years" which is the same sort of thing, except it's the distance that light would travel in 1 year. It's a handy unit if you're an astrophysicist (or studying QS&BB).



June 11, 2017 08:37

Figure 15.16: spacetimeborn1

can put that on our spacetime diagram. But it's strange: low look at point **C** which obviously doesn't make any sense. That other observer sees that Benjamin first cries... and *then he's born* since the crying happens to the left of $ct = 0$ where Benjamin cries before birth! Since this creates a cause-effect reversal, our "Euclidian" assumption of a circle as the invariant spacetime curve must be wrong.

Space and Time: Doomed to Fade Away

This is where Einstein's critical mathematics teacher comes in because it was he who worked out the technical mathematical basis of Special Relativity and we've been using his terminology of spacetime all along. Famously, Minkowski gave a speech in 1908 80th Assembly of German Natural Scientists and Physicians:

> "The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Einstein later remarked that, "Since the mathematicians have grabbed hold of the theory of relativity, I myself no longer understand it." But he later came to understand the fundamental importance of Minkowski's work and publicly acknowledged that, but unfortunately only after Minkowski unexpectedly died at 45 years old.

15.6 Spacetime

Let's figure out what thinking in terms of spacetime implies. Figure 15.17 lays it out and introduces a new concept.[9] We take it that no signal or material object can move faster than the speed of light. So if we set up a Global coordinate system in spacetime, then this boundary corresponds to a line with a slope of 1.

$$\text{slope} = \frac{x}{ct} = 1$$

which of course implies that

$$\frac{x}{t} = c.$$

This is in one space and one time dimension. If we expand to imagine two space dimensions, then these two boundary lines represent the surface of two cones oriented to the right (the future) and the left (the past) along the ct axis directions.[10]

So anything traveling at the speed of light[11] would travel in spacetime along the top or the bottom diagonal lines.[12]

[9] Notice that I'm now plotting ct rather than t as per the bullet above.

[10] If we expand it all the way to our actual three space dimensions, we have to wrap our heads around the idea of a hyper-cone, which I'll not try to do myself. But feel free if you're so inclined!

[11] . . . .um. That's only light.

[12] The top line would mean going in the $+x$ direction and along the bottom, in the $-x$ direction.

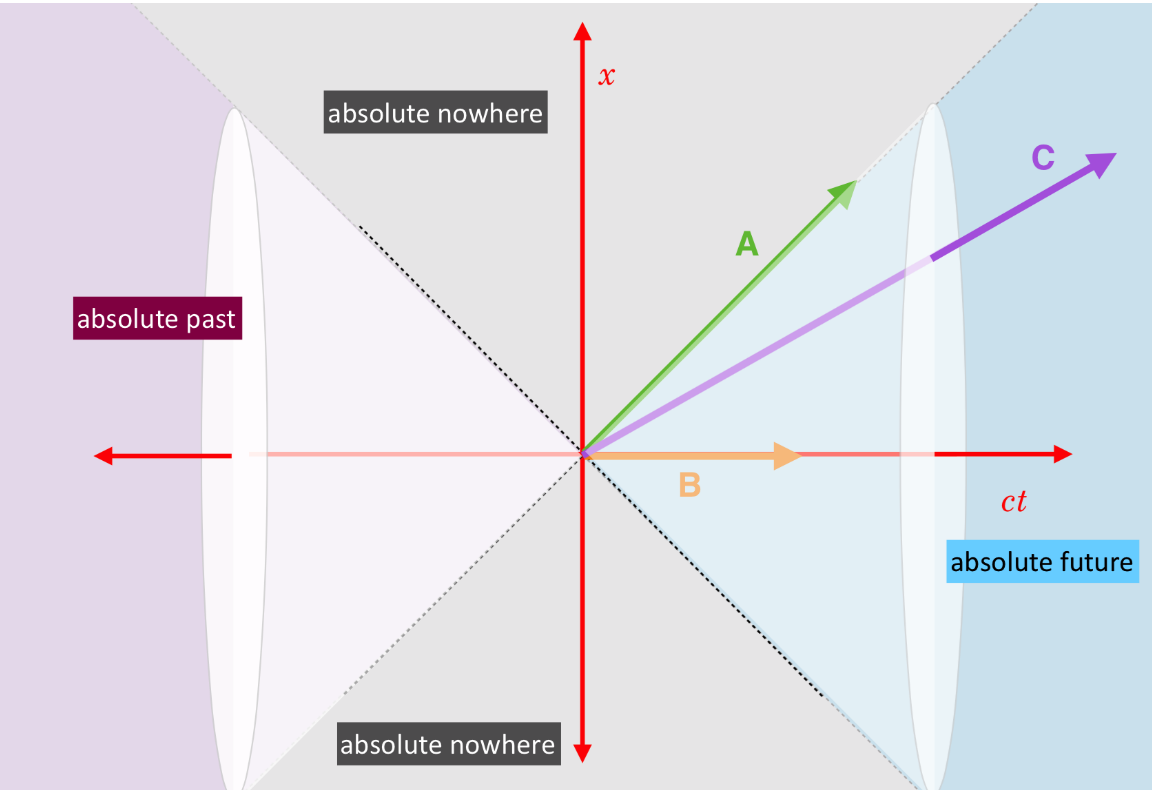


Figure 15.17: worldlines

A is a trajectory of a positively directed light beam.

Anything that's sitting still is still on a special spacetime trajectory, all of it's "motion" is along the ct direction, just in time. B is such a path. Finally, anything moving at speeds less than c would fall within the cone and C is such a trajectory. That's where we live in our everyday, sub-luminal lives.

These cones are special in Relativity and together they are called the **Light Cone**. Since nothing can travel faster than the speed of light, all real, physical trajectories must lie within the Light Cone and such trajectories are called the **Worldlines**. You can think that every object in the universe has a cone in spacetime attached to it that limits what the future might bring and what the past has been. I've drawn a worldline as a straight line, implying moving at a constant velocity, but real-life worldlines accelerate and decelerate and so they would trace out curves—but never with slopes steeper than 1.

What about outside of the Light Cone? Those are regions of space and time that are simply inaccessible to an observer and we call them the **Absolute Nowhere**, in order to make it sound spooky.

What Minkowski discovered was that there is indeed an Invariant Curve for relativity—a "spacetime length" that is the same for all observers in co-moving inertial frames of reference.

> **Wait.** *We've seen that times and lengths and even simultaneity are relative to a frame. How can there be something that's constant?*
>
> **Glad you asked.** *Indeed, separately time intervals and lengths do appear to be different from one frame to another, but separately. A particular combination turns out to be constant, just like x and x' or y and y' are individually different, but Eqs. 15.12 and 15.11 show that a particular combination is constant, namely as L. Watch what's next.*

Remember, the invariant length for just space is

$$L^2 = x^2 + y^2.$$

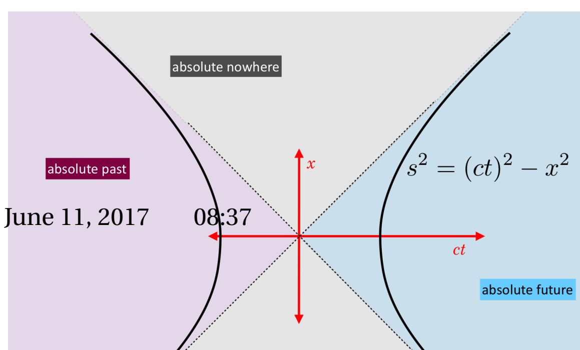
The invariant length—called the Invariant Interval (or just "Interval" for short)—for spacetime in "(1,1) dimensions" (1 time and 1 space dimension) turns out to be:

$$s^2 = c^2 t^2 - x^2. \tag{15.13}$$

That pesky minus sign makes all of the difference and Fig. 15.18 shows how. Equation 15.13 is the formula of a *hyperbola*, not a circle! Relativistic spacetime is hyperbolic and it's called **Minkowski Space**, quite different from Euclidean Space. Figure 15.18 shows this on the spacetime plot where the hyperbolae going left and right are the "real" Invariant Curves for our universe. Any trajectory or set of events will have spacetime points that must lie on a hyperbola, regardless of what reference frame they are in. Just like your wall drawings all have space points that lie on that circle of radius, L, a "length" in hyperbolic space is the same if it goes from 0 to any point on the hyperbolic curve.

This experience with Special Relativity is the first hint that unusual geometries figure into physics. Einstein reluctantly backed into it through his former teacher, but by the time he got to General Relativity, he rushed headlong into even stranger geometries.

We call that distance that always is the same—the one that lands on the hyperbola—the "Interval." It is the length of a line from the origin of the Worldline of an object to the surface of the hyperbola represented by Eq. 15.13. For space we called it L and for spacetime, we'll call the interval, s. Any arrow will do, and each one represents the space and time coordinates of inertial observers moving relative to one another,



June 11, 2017 08:37

each being an AF, AF', AF", and so on. If an arrow lies on the horizontal axis, then that's special and that the object has not moved in position and that it has a value for the Interval of

$$\begin{aligned} s^2 &= c^2 t^2 \\ s &= ct \end{aligned} \tag{15.14}$$

This is represents a unique frame: a worldline totally made up of time and hence from within the rest frame (the HF) of an object. Any motion relative to the rest frame involves a mixture of space and time coordinates that will satisfy Eq. 15.13 and become a family of different arrows. Let's go back to the airport.

It's a sad story. WearyTraveler on the moving sidewalk has spent his entire life there. From a small boy at the beginning of the trip until the current day, he's just moved along. His time is measured by the watch that he was given as a youngster and measures time in his frame. His space hasn't changed on the walkway.

Even more pathetic are the CouchPeople who have been sitting and watching WearyTraveler's life progress as he moves along in front of them. He's in their AF and they are still their own HF. They measure time with their clock. Figure 15.19 shows the sorry tale. Let's take this complicated figure apart:

- This is a picture of WearyTraveler's sorry life taken at two times: early and now.
- We see WearyTraveler as a child, at $T_A = 0$, which coincides with CouchPeople's original time, $T_H = 0$.
- We see that WearyTraveler never changed his position through his life: I've moved the coordinate axis so that WearyTraveler is standing at his origin. This way $x_A = 0$ that he was at as a child is still the same distance, $x_A = 0$ as an adult.
- But we see that in the airport, WearyTraveler appears to have moved from$x_H = 0$ toX_H.
  What do they agree on? The Interval, which as an increment (squared) in spacetime is:

  %

$$\begin{aligned} \Delta s &= (ct_H)^2 - (\Delta x_H)^2 = (ct_A)^2 - (\Delta x_A)^2 \\ \Delta s &= (ct_H)^2 - (\Delta x_H)^2 = (ct_A)^2 \end{aligned}$$

  %

  We've already thought about this situation and we saw different space and time interpretations by each frame. But now we can see that they agree completely on the value of the Interval. Figure 15.20 shows the situation. Here the arrow along the horizontal axis is the AF

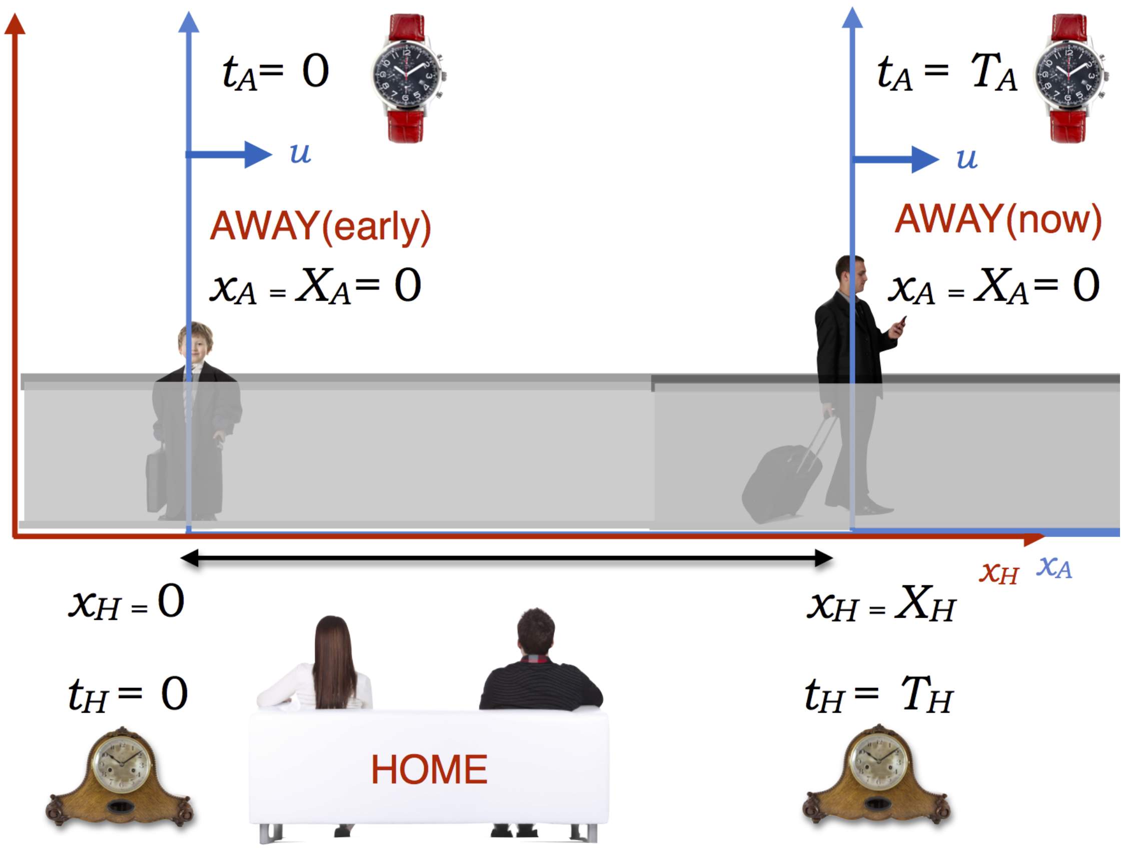


Figure 15.19: life

determination. It's the time ($\times c$) of the WearyTraveler, in his frame. The other arrow measures the Interval for the AF as measured by the HF. These lengths are the same in a hyperbolic space!

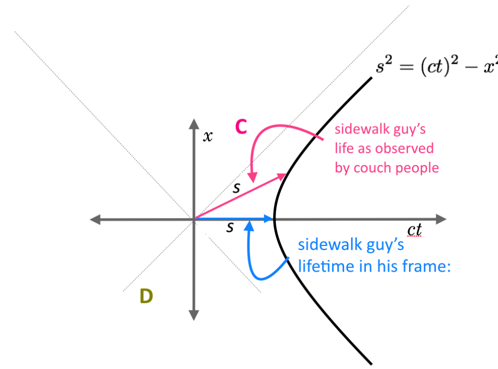


Figure 15.20: spacetimeinvariant

15.7 Why Don't We Live Relativistically?

If relativity is right, why don't we see relativistic effects in everyday life? Let's look at the interval again, for two different frames, a HF and an AF. The interval would be the same for each:

$$s^2 = (ct_A)^2 - x_A^2 = (ct_H)^2 - x_H^2$$

So each represents two different lines on the spacetime plot, pointing at the hyperbola. How about the tiniest time interval that a human might deal with. . . a single second in my HF.

$$ct_H = 3 \times 10^8 \text{m/s} \times 1 \text{ s} = 3 \times 10^8 \text{m} = 300,000 \text{ km}$$

This is almost the distance from the Earth to the Moon, so on human terms the time piece of the interval dwarfs normal human-measure distances that one might encounter. Since normal human speeds are tiny compared with the speed of light, then any time interval in the HF is going to be very close to the value as observed for the AF. So

$$s^2 \approx (ct_A)^2 \approx (ct_H)^2$$

That, in turn, means that the interval arrows for AF and HF in spacetime Fig. 15.21 are very close to one another and so everyday fames would be very similar and the time-space mixing would be negligible.

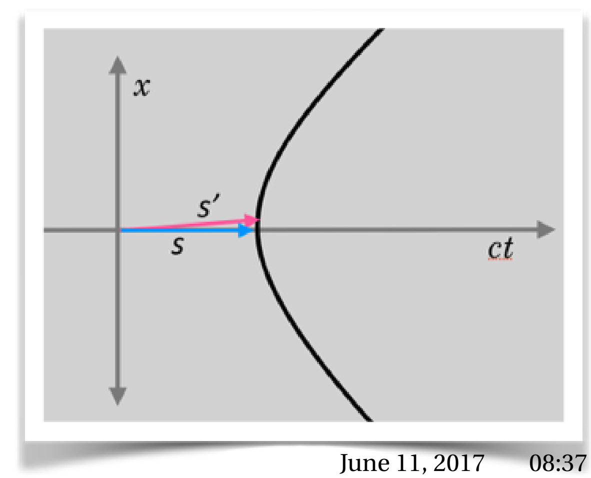


Figure 15.21: longtime

15.8 Simultaneity, Or Something

I've followed Einstein's thinking in reverse from his actual inspiration. Let's fix that now.

15.8.1 A Storm Broke Loose in My Mind

He was fixated on the electromagnetism contradictions and later recalled that after a night of thinking about them—when he was about to give up—that "a storm broke loose in my mind." He suddenly realized that the problem he faced was not about reference frame *speeds*. It was not about *Maxwell's Equations* and \(c\). No. The problem that he was wrestling with was **Time** itself. Here's the thought that lit up his brain:

RIP, Simultaneity

Our blessed, Benjamin-birth tiptoed up to a philosophically dangerous situation, namely that a basic assumption about nature is that causes come before effects, and not the other way around. As silly as that sounds, Einstein had to make a serious philosophical leap in order to preserve it.

Figure **??** shows our airport folks yet again. Not yet tired of their light-measuring machine, they're at it again, this time on the sidewalk alone. They've rigged a light bulb that WearyTraveler has carefully and precisely placed midway between his two light sensors, turning them both to aim at the bulb. Here's the experiment: when the bulb is turned on, what is the difference in times for the light to reach the two sensors?

In the rest frame of the bulb, apparatus, and WearyTraveler... it's obvious. The light should reach both sensors at the same time. In the frame of the airport, it's different: The right hand sensor is running away from the light beam, so it would take the light more time to catch up with it. And the left hand sensor is coming toward the light beam, so it would reach quicker. So WearyTraveler would say, "Simultaneous!" But CouchPeople would say, "No! Left hits before Right.

Can't they calibrate their equipment to take into account the motion? Here's how a classical physicist before Einstein might have thought of this kind of circumstance. The appearance of the lack of simultaneity can be fixed if one just added or subtracted the speed of light and the speed of the sidewalk. Think of it this way. Suppose we've got a tug and a distance L behind it, a barge. The captain in the tug wants to synchronize his watch with the sailor in the barge. When the sailor sees midnight on her watch she yells to the captain, "Midnight!" The captain hears her but does he then set his watch to midnight? No, he knows that by the time he's heard her announcement, some time has passed since the speed of sound is finite and actually humanly slow. So he calculates how much time it would take and subtracts that time from his watch and sets it so that it would match the sailor's watch. That works just fine if they're sitting at port. But what if they're steaming ahead... now he has to adjust the time not just for the distance between tug and barge, but also for the fact that the tug is going away from the barge, so it would take still longer for the sound to reach him. He can do that because!... he knows the speed of sound in air, which is where the sound propagates. It's an absolute reference frame for sound: the speed of sound is constant only there and apparently different for other moving frames. But you can always calculate things back and synchronize watches, etc. In this situation.

Wait. *So why can't WearyTraveler and CouchPeople make that same kind of re-calibration for the relative speeds of light and sidewalk?*

Glad you asked. *Haven't you been listening? Sorry. Too snarky. The difference is that there's no analog of the air, in which the speed of light is c and only there is it c. Since the speed of light is the same value in all inertial frames, there is no way to make that correction.*

The concept of simultaneity is a relative concept: two events that are simultaneous for one inertial frame are not simultaneous in another. Nobody is right and nobody is wrong and this has consequences for what it means to make a measurement, a situation that Einstein called out in the opening paragraphs of his 1905 paper. There, he almost patronizingly notes,[13]

> "We have to take into account that all our judgments in which time plays a part are always judgments of simultaneous events. If, for instance, I say, 'That train arrives here at 7 o'clock,' I mean something like this: 'The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events.'"

This simultaneous event—the clock pointing to 7 when the train arrives—is the case only for observers standing right with you in your frame of reference. Any other relatively moving observer would disagree. Their simulteneity-watch-setting would be different, no worse, no better.

Here's the shocking reasoning: If you cannot rely on things being "simultaneous" then you cannot agree on the Time of an event. If you can't agree on the time of an event, then Time suffers a humiliating demotion from something Absolute to something relative. This is what so excited Einstein. Nobody had before him thought of the possibility that time was not as Newton insisted...that it didn't just flow absolutely from past to future independent of anything external to it.

> Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year. Isaac Newton, Principia

> People like us, who believe in physics, know that the distinction between past, present, and future is only a stubbornly persistent illusion, Einstein in correspondence with the family of his deceased friend, Michele Besso.

> An analysis of the concept of time was my solution. Time cannot be absolutely defined, and there is an inseparable relation between time and signal velocity.

15.8.2 What is Time?

> If you do not ask me what is time, I know it. When you ask me, I cannot tell it. Saint Augustine

This was a profound discovery for Einstein. He realized that as a result there is **no special meaning to the notion of the time of an event**. When he realized this everything about relativity flowed in a couple of weeks from that awakening. He wrote his paper, gave it to his wife to check for mistakes, and then took to his bed out of sheer exhaustion.

[13] This was a very strange paper. One goes many pages into it before running into any mathematics. The first half-dozen pages are stories and careful definitions of what it means to make a measurement. Presumptive for a rookie, unknown scientist.

Later he defined time: It's what a clock reads. Nothing more, nothing less. If what clocks read differ for different observers, then time is no longer an absolute.

Now

What's "now" for you? It's our own personal notion—that we all share—that we're a part of a big universe in which things are happening. . . now. But given two aspects of our discussions, this is a highly complicated idea. First, signals about what's in our now-universe are not instantaneous, but they can only reach us at the finite speed of light, or less. So when I look in the mirror, I'm not seeing my face now. . . I'm seeing my face about 2 nanoseconds ago: 1 ns for the light to bounce off my face to the mirror and another nanosecond for it to bounce from the mirror to my eyes.[14]

In fact, nothing outside of my immediate place shares my now as far as I can determine it. If I look across the room *now* what I see is what the room looked like a few nanoseconds ago. If I gaze at the moon *now* I see what the moon looked like a few minutes ago. If I look at the Andromeda galaxy *now* I see what it looked like 2 million light years ago. So the finite speed of light adds a complication to what we can say about the nature of reality, *now*. That's sort of trivial when you think about it. Troubling, but trivial.

The other aspect of what we've just talked about is that even if we know how far away the other side of the room is—and I can unambiguously measure that—and I know the speed of light, I can arrange the objects that I look at now in what you might think is a proper time-sequence. But the inability to unambiguously identify my *now* as the right one and a *now* from the International Space Station, or an Andromedian observer is even more troubling. They're all proper nows and all of them are legitimate. Each is right and each is different.

In fact, it's even worse. We have this idea that time "flows" and we along with it. Our past is determined—I can't undo that yellow Mustang purchase in 1973—but my future is still to happen (to me) and I can avoid such a purchase next year. But my personal time seems to proceed (at an increasingly urgent clip) from past to future, passing through *now*.

We've seen that space and time are now two sides of the same thing we call spacetime. And we've seen that electric and magnetic phenomena are but two sides of the same thing we call electromagnetism. The consequences of the special theory of relativity seem to be unifying heretofore, different things. But as much as we think of spacetime as a single thing, we can't shake this idea that time is somehow special. After all, I can walk in the positive x direction and I can reverse myself and walk in the negative x direction. As much as I'd like to change my mind about that yellow Mustang, I can't myself walk in the negative time direction.

[14] And, I guess some number of microseconds for the light collection on my retina to be processed by my visual cortex and recognized by my brain. But that's not my concern here.

But! The underlying rules of physics are agnostic about the direction of time. Reverse the sign of time in the equations and the world would be indistinguishable from the other direction. If make a movie of aiming a pool ball so that it collides with another? If I play it back you can't tell whether I'm playing the film forwards or backwards. . . and the rules of physics describe both. So in physics, the direction of time is not so insistently "forward" as it seems to be for us personally. We'll see legitimate instances of this being a legitimate interpretation of real happenings when we get into quantum mechanics.

Suffice it to say, Einstein was the first in history to raise legitimate questions about the nature of time.

15.8.3 What about Causality?

You might be able to think of a circumstance in which maybe an airport officer on her Segway might be going by this little sidewalk parade in the other direction. Might she expect to see the light show in the opposite order? This might conjure up the idea of a problem of causality—that maybe we're back to crying before being born just by virtue of relative motion? Let's think about it slightly differently.

Suppose that we arrange for two light bulbs to fire simultaneously inside of WearyTraveler's frame on the sidewalk. A relatively moving observer (like CouchPeople, or SegwayCop) might say that the left hand bulb fired first, or the right hand bulb fired first. But relativity prevents one from actually observing a reversal of the order of events. . . the time spacing is different, sure. But not which came first and then second. Can't change that. Why? Because in order for that reversal to happen, the bulbs would have to be spaced so far apart that a signal could not reach to the observer or the signal would have to travel faster than the speed of light.[15] It would have a worldline that would be more than that light-speed-limiting 45 degree line on Fig. 15.17. Can't happen. Relativity is safe.

[15] See Diagrammatica: Spacetime Diagrams for examples.

But just as it takes two to tango, we've seen that it still takes both space and time to make a velocity. c is a velocity, and so a distance divided by a time. In our Galilean transformations space quantities change between frames. But as we saw in detail if there's a special velocity that never changes, and if the space coordinates change, then time cannot be above the fray. Time's "marching on" had to be different for different observers and time coordinates are going to adjust in any comparison of events between co-moving frames of reference, just like space coordinates do.

15.9 So, What About That Ether?

When Einstein was doing his Bern-thing, he was in relative isolation. He didn't have access to a university library and so he had to rely on whatever passed through the patent office and whatever he remembered or could get through the mail. For many years it has been a matter of controversy as to just what he knew about research in the areas that touched his eventual Theory of Relativity. Let's go out west.

15.9.1 Albert Michelson

When someone measures something and their name is attached to it, that's a big deal in science and it happens every so often. Usually this indicates a significant discovery. When that experiment is a decade-long failure, well that's even more rare!

Albert Michelson (1852 - 1931) was born in Poland and his family moved to the United States when he was two years old. They were adventurous family—they went to the wild west and became merchants in various mining communities in California and Nevada. Michelson himself went to high school in San Fransisco, living with an aunt where he was a good student. His college education was unusual. He applied in a competition to the relatively new United States Naval Academy at Annapolis, Maryland and was rejected. So he did what anyone would do in such a situation, he got on a train in San Fransisco[16] and went to Washington to see the President. He was nothing, if not persistent, and President Ulysses S. Grant personally admitted him to Annapolis as a midshipman in 1869. He graduated and did two years at sea in the Navy and returned to the Academy as an instructor[17]. He had become a master experimentalist in the measurement of precision optical phenomena and perfected heroically precise techniques to measure the speed of light to very high precision. His expertise led him to study in Europe for a while and to return to making increasingly better measurements of c.[18] He was working on such a measurement using a one mile evacuated tube when he died in 1931.

Figure 15.22: michelson

Michelson was notoriously cranky and difficult. His first wife tried to have him commmitted and a maid sued for abusive treatment. He once had an argument about an experiment with a colleague in a hotel lobby that drew a crowd, maybe because they were loud and maybe because Michelson was still in his pajamas. He won the Nobel Prize in 1907, not for his measurements of c, but rather because of the most famous measurement of "zero" in all of physics and the device he invented in order to do it.

15.9.2 The "Michelson-Morely Experiment"

What Michelson decided to do when he resigned from the Navy and became a member of the faculty at the Case School of Applied Science, now called Case Western Reserve University in Cleveland, Ohio. There he teamed with Edward Morley (1838-1923) to do some of the most audacious experiments of their time: they tried to measure the speed of the Earth relative to the ether.

[16] He was one of the first transcontinental train passengers.

[17] He married the daughter of the head of the physics department!

[18] One of his most precise measurements was between two mountains separated by 22 miles in California, one being the Mount Wilson

Remember, Michelson's time is not so far removed from James Maxwell and his theory of light. He and everyone around him believed that what he'd described were waves that "waved" in the invisible, but persistent ever-present ether. . . the Luminiferous Ether. Everyone presumed that the Earth and the planets were orbiting through this stuff and that we see the Sun, meant that the ether was jiggling as the light (and heat) from the Sun propagated to Earth. The ether was everywhere, but the Earth must be moving through it at some finite speed and Michelson and Morley set out to measure it.

Let's imagine a river that's flowing uniformly from left to right. If you were to take two identical motorized toy boats and set them going in the river, we could measure the river's speed by comparing their motions in two directions. Here's what we would need to know: the distance that each boat travels (W) and the speed that the boats go *relative to the water* (C). Let's say that the width of the river is W feet. One boat is sent racing downstream a distance W and then back upstream to the starting point. . . so it travels $2W$ feet total. Downstream, it would go *with* the current and so faster relative to the shore and in the return, it would fight *against* the current and go slower. If the river's speed is V, then the time downstream would be $t_{\text{down}} = \dfrac{W}{V+C}$ and the time to come back would be $t_{\text{back}} = \dfrac{W}{V-C}$. Since $V+C$ is bigger than $V-C$ and is in the denominator, then the time down is indeed smaller than the time back.

The other trip is across the river and in order to end up exactly opposite where the second boat starts, you'd aim a little upstream so that the river would carry the boat along as it goes across and the trajectory would be diagonal. The time to make this journey can also be calculated in terms of V and $c$and precisely predicted. . . it will be faster than the down and up path. But the point is the following: if the water is not a flowing river, but a swimming pool, it wouldn't matter what direction you went, the times for the trips would be the same. Only if the water is flowing with a finite speed ("C") along with the first paths will the trips' durations be different in time.

Suppose instead of swimmers in a current, we have light in the ether. Since the Earth is moving through the Ether, a beam of light in that direction would be faster (or slower) relative to a beam that moved perpendicular to that Earth-ether "current." The speed of light is of course exceedingly fast and measurement of the absolute wavelength of light would be nearly impossible—after all, a red laser beam consists of light waves is about 650 nanometers, that's 650×10^{-9} m. But Michelson invented a clever device known ever since as the Michelson Interferometer. It's hard to measure the absolute wavelength, but if two beams are brought together so that they interfere and one is slightly out of phase relative to the other they would combine and make a visible interference pattern. The peaks in the interference are a) an indication that the waves are out of phase and b) an indirect measurement of that difference in phase. That quantity in turn could be used to determined the speed of the Earth relative to the Ether. Then Great Acclaim would await Michelson and Morley.

Figure 15.23 is a sketch of the idea. A source of light, A, sends it to a fancy mirror, H, which is designed to reflect half of the light and transmit the other half. So the reflected portion at A goes to B, reflects from another mirror M_1 and then passes back through H onto a telescope, T. Meanwhile the transmitted wave (dashed) goes through H at D, bounces from another mirror, H_2, comes back and follows the first beam from F to G at the telescope, T. Suppose the Earth is moving along the D-E length, then the speed of light in that D-E-F path will be faster than the speed in the A-B-C length and they would interfere at T.

The longer the path-length the more precise the measurement. In practice the experiment was very hard

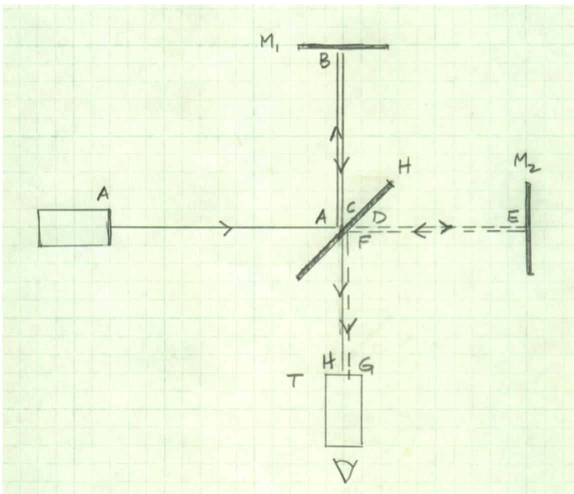


Figure 15.23: MMintme

But Michelson and Morley were even more clever. First, they mounted their device on a very heavy (large inertia, so stable) circular platform that they floated in a big tub of Mercury (a heavy—now known to be dangerous—liquid) and they rotated it around the center so that they would eliminate any bias of any particular direction. Any fringing is a positive measurement, so rotating it is fine. Also they set up a more complicated scheme that I've drawn in Fig. 15.23 and had the paths be many reflections before they were brought to interfere, in essence increasing the length of each path. Figure 15.24 is a drawing from their 1887 published paper and Fig. 15.25 is a photograph of their apparatus from the Case Western archives.

Their result: zero. zilch. nothing. nada. zip. diddly-squat. nil. Over and over, they measured no relative speed to the ether. It was a huge frustration and a crisis in physics and by 1887 everyone was panicking. The ether had to be there! Maxwell's Theory required it. Drastic measures were required: Michelson and Morley suffered over their equipment and the theory community began to make bizarre suggestions. One of the strangest was that the motion through the ether actually caused one of the arms of the apparatus to shrink... literally that the atoms[19] would be closer together and the length would be shorter.

This idea is called Lorentz-Fitzgerald Contraction after the two brave theoreticians who reluctantly proposed it. The amount of the contraction? Well, by now you probably can guess. It was precisely the same amount that comes out of Einstein's Length Contraction formula.

15.9.3 The Superfluous Ether

In Einstein's paper there are no references to any other publication. This is highly unusual and it's even surprising that he got away with it. There's a heartfelt "thanks" to one of the Olympia Academy members, a buddy from the Patent Office,[20] but no reference to the work of Michelson's nor of Lorentz'. The question remains—in no small part because Einstein himself was not consistent in his recollections—did he know of the null Michelson Morley results? Or did he "predict them" after the fact? The general conclusion of

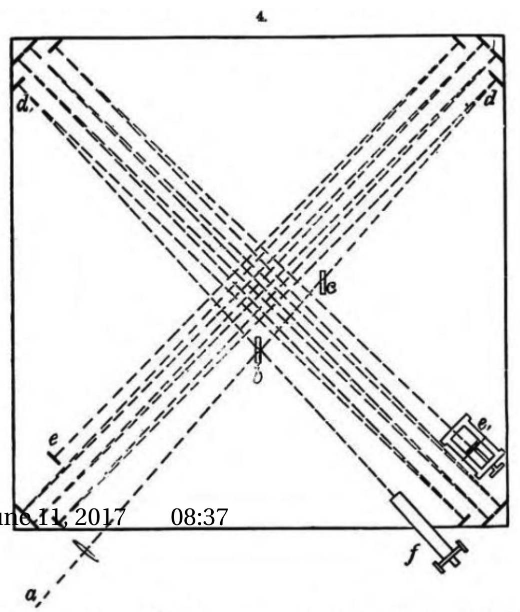


Figure 15.24: MMactual1

most historians is that he was aware of the null ether results but maybe not very familiar with the anxious theoretical work done in the previous 10 years to try to understand those results.

In any case, Einstein's conclusion was clear: he claimed that since there was no way to figure out if an observer is in a privileged reference frame—like Newton's Abosolute Frame, or the frame in which the ether presumably was stationary—then no such frame can exist. All frames, so to speak, are created equal. If you can't detect it, then you can't declare its existence, remember? They are equally likely and none can be picked out as the one that's *really* at rest. And so this 26 year old unknown patent clerk stated quite confidently in his 1905 paper:

> "The introduction of a 'luminiferous ether' will prove to be superfluous inasmuch as the view here to be developed will not require an 'absolutely stationary space' provided with special properties, nor assign a velocity-vector to a point of the empty space in which electromagnetic processes take place."

Quite remarkable.

15.10 The Most Famous "Paradoxes" of Relativity

Michelson never doubted the ether and he wasn't alone. This idea was so deeply ingrained that it was just too much for many to give up on. This put those older scientists in a tough spot and some lived a dual life, accepting the formalism of Relativity, but not holding to the basic consequences. It was not hard to begin to think of all kinds of "impossible" paradoxes that should render the theory crazy, but they have all found explanation, usually with a deep use of the rules of relativity itself.

I'll discuss the tests that confirm Relativity in the next chapter, but we should skim through one of the more colorful challenges that was thrown down to the theory. And then dismiss it. The famous siblings:

Twins

If you want to call Relativity crazy, you don't have to look any further than the famous Twin Paradox. Mr Google will tell you that it provides you with more than a million hits, which isn't bad for an obscure mathematical conundrum of physics! I believe that one of the first challenges on this score came from the French physicist, Paul Langevin in 1911 and it goes like this:

Two twins are born on Earth and one of them gets into a spacecraft and goes to a distant star at a relativistic gamma of 100, $\gamma = 100$. His sister stays on Earth and lives her life. The spaceman twin then turns around and comes back to Earth at the same speed as before and when he arrives he has aged just two years... but he finds that his sister is long gone since 200 years have elapsed on Earth. Now there are two things that are troubling about this, three if you want to be sentimental. First the substitution for a ticking clock by a biological organism often hits people the wrong way. But biology is still chemistry and chemistry's engines are electrical in the end and so the rules of physics, quite independent of thinking, breathing organisms still rule the day. A biological clock is a clock and the rule of physics still apply. No, the paradoxical part comes from asking the question the other way. We assumed in the statement of the problem that the Earth twin was at rest and we reached our conclusion by thinking about it as if the Earth frame was special.

But we know now that that's not fair and so shouldn't we be able to ask the question from the other perspective, as if the space twin was "at rest" and the Earth receded from it and then returned? And then wouldn't it be the space twin who would be 200 years older and the Earth twin only 2 years older? That's the "paradox" part of the Twin Paradox, and the answer is no.

If the twins were separated and then one left in the spaceship and they never re-connected, then indeed, neither frame would be special and the conclusions about the time-dilation would be reciprocal. But that's not what happened in the story. The Earth twin remained in one rest frame for the whole time, while the space twin lived in *two* rest frames, the one going out and the one that returned. That makes all the difference. Now you could ask about the alternative assignment of the space twin being the HF and the Earth twin the AF. The Earth twin still stays in one frame, but in order for them to meet up, even if the space twin is at relative rest on the way out, in order to then catch up, he has to go even faster in a new rest frame than in the first verse of the story. The end result is still the same, he's aged 2 years and the Earth has aged 200 years. What's not relative is this: the Earth twin always stayed in one rest frame while, no matter how you tell the story, the space twin has to participate in two different reference frames. There's no relativity about that. It's not a paradox. Let's put the ladder away.

15.10.1 Fitting in the Garage

One other tricky paradox. We have a garage that's G meters long and our ladder is L meters long and $L > G$. If we run at the garage holding the ladder and run really fast, then from the garage's perspective, the ladder is Lorentz Contracted and it fits. But from the ladder's perspective, the garage is Lorentz contracted, and so the ladder is even less of a fit than before. That's a paradox. Or is it.

15.10.2 Relativity From the Sky

Interstellar space is full of all kinds of particle debris and much of it bombards us constantly—we call them Cosmic Rays and they are a threat, an annoyance, and a useful tool. Many of the interstellar particles are protons which have been accelerated to very high speeds through a mechanism that we don't understand. The process of particles getting to Earth is a complicated one and not unlike the process that we use on

Earth when we intentionally crash protons or electrons into matter in experiments. The protons hit the nitrogen in the atmosphere[21] and create enormous particle showers which can be miles across.[22]

Among the by-products of these collisions is the creation of a particle called the "muon" which is essentially a heavy electron. We'll meet the muon later as it plays an important role in the history of particle physics and its relationship to the electron is one of great interest. Most of the particles that we know about are essentially unstable. That is, they decay into other particles. One moment they're there and the next moment something else is there! We know of only two particles that are stable against decay—that is, our measurements searching for their decay lead to lifetimes that are longer than the lifetime of the universe. Lucky for us, as these two stable particles are the proton and the electron.[23]

The muon is relatively long-lived (on the scale of subnuclear things) and decays on average in about one and a half microsecond, 1.5×10^{-6}~seconds.[24] So how far would a muon go on average if it were traveling at the speed of light?

[21] Did you know that most of the atmosphere is nitrogen and that oxygen is only the second most abundant element?

[22] We'll study these particles later.

[23] The neutron is unstable and decays by itself in about 10 minutes. But when it's bound into a nucleus, that decay is suppressed because? Relativity. Next chapter!

[24] This is the so-called "half life" of the muon. It means that if we start with 100 muons, after 1,5 microseconds there will be 50 left. After another 1.5 microseconds, 25. . . and so on.

QR question muon distance

Figure 15.26 shows the situation. The cascade of particles (I've only drawn in one of the hundreds of thousands of particles in the "shower" induced by the original proton.) leads to a steady rate of about 1 muon through your thumbnail every minute. The issue is the distance that the muons must travel in order to reach the earth. As you saw in your calculation above, on average a muon will decay in less than 500 meters, yet they seem to make it to Earth——more than 50,000 meters! How does that happen? This is many, many "lifetimes" for a muon. If 100 muons are produced in the upper atmosphere, after only about 2000 meters, there would be less than 10 left. There's still more than 45,000 meters to go!

We have the tools to understand this.

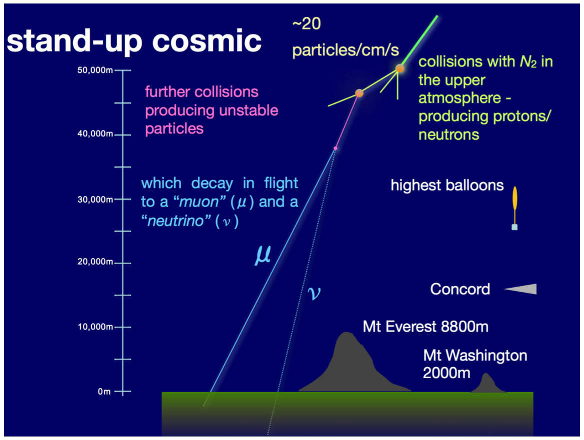


Figure 15.26: cosmic

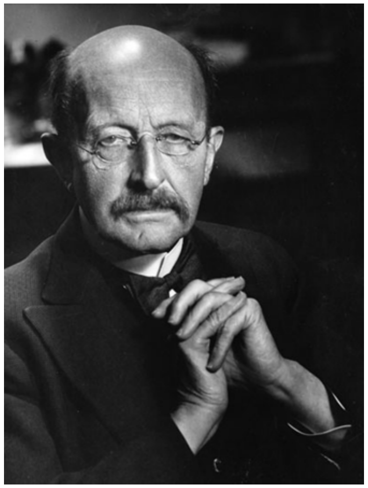
muons

Now discuss the tee shirt equation that I know you've been wanting to understand.

Chapter 16

Quantum Theory

Now For Something Completely Different


Max Planck,

Max Planck, 1945-1946

"When I began my physical studies (in Munich in 1874) and sought advice from my venerable teacher Philipp von Jolly...he portrayed to me physics as a highly developed, almost fully matured science...Possibly in one or another nook there would perhaps be a dust particle or a small bubble to be examined and classified, but the system as a whole stood there fairly secured, and theoretical physics approached visibly that degree of perfection which, for example, geometry has had already for centuries." *Max Planck ( )*

We have seen just how unusual the ideas of Relativity were and, how slowly they were adopted by the physics community. In some ways, the insult to common sense that was Relativity is secondary to the injury that became Quantum Mechanics. By the 1920's, like a slow-motion, one-two punch, physics realized that it had been rocked by the tag-team of those two new subjects which called into question basic facts about nature, but also the status of Reality and Knowledge. Between Relativity and Quantum Mechanics, Ontology and Epistemology—what *is* and what we can *know*—took their biggest hits since Plato.

16.1 Goals

The goals

16.2 A Little Bit of Max Planck

One would be hard-pressed to identify more people in the history of 20th century physics more respected by his or her colleagues than Max Planck. The story of the beginnings of quantum mechanics are essentially solely identified with Planck and Einstein and in this chapter we'll become familiar with their early adventures into the strangeness of quantum mechanics.

Planck was the epitome of German order and precision. He came from a family of academics in law (his father was a prestigious Professor of Constitutional Law in the University of Kiel, and then Munich) and theology (both grandfather and great-grandfather were theology professors at the University of Gottingen) so his path was paved by the expectation that he would be patriotic, honest, moral, fair, and generous... and an academic himself. He was an accomplished musician, and like Einstein, took great comfort in piano and organ... nearly taking music as a career. However, after some physics and mathematics courses at the University of Munich, he became hooked on physics. He described his reasons for choosing physics in his *Scientific Autobiography, and Other Papers* in 1949,

Figure 16.1: planckyoung

> The outside world is something independent from man, something absolute, and the quest for the laws which apply to this absolute appeared to me as the most sublime scientific pursuit in life.

More irony. Quantum mechanics taught us that the outside world is anything but "independent" and "absolute."

Planck excelled and received his PhD at the age of 21 and quickly moved through the hierarchical German academic system, eventually making it to the peak: the University of Berlin in 1888 as the Director of the Insititute for Theoretical Physics—at the age of 30! He held this position until he retired in 1928. His specialty was thermodynamics and he was renown for his clear lectures and brilliance in this, at that time, confusing field. Far from the study of just temperature and the mechanics of heat, thermodynamics had become a theory of statistical treatment of still hypothetical atoms. This wasn't Planck's personal preference—not until he abandoned the idea that the Second law of Thermodynamics was not an absolute rule of nature, but a statistical one. And that switch was in order to solve the problem that is the subject of this chapter.

As we'll see, understanding how objects radiate heat was a theoretical conundrum at the close of the 19th century and it was in solving it that Planck secured his name in the textbooks. But he didn't quit when his formula fit the new data from the laboratories in Berlin. In addition he forced himself into a tortured few months of trying to find a physical interpretation of what his mathematics was telling him. And that was the birth of the "quantum."

Planck was 42 years old when he made his historic leap and defined modern physics. He was awarded the 1918 Nobel Prize for this work and confessed that for a long time he didn't understand his own theory, and then distrusted its conclusions.

He was the permanent Secretary of the Mathematics and Natural Science Section of the Prussian Academy of Sciences from 1912 until 1943. In 1929 the Max Planck Medal was established as the highest award of the German Physical Society, and Planck himself and Albert Einstein were the inaugural recipients.

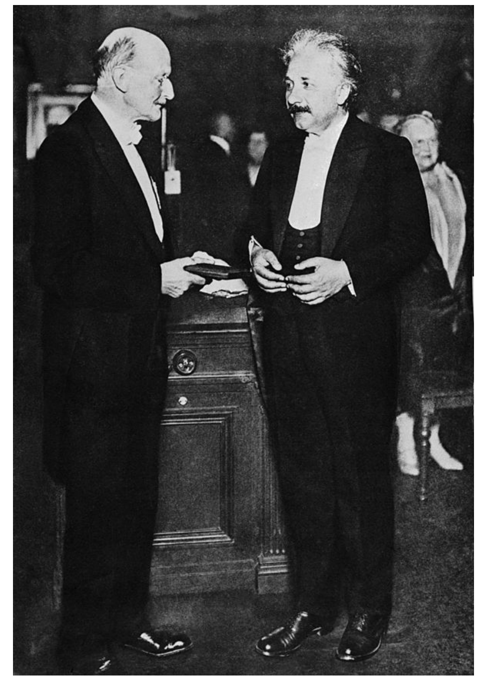


Figure 16.2: planckmedal

Science was extremely important to the German government and society and as the acknowledged leader of all of German science, Planck was a respected advisor to the German government. He as also revered for his fairness by his colleagues and it was no surprise that he personally involved himself in trying to persuade Adolf Hitler away from his racial laws. He failed and when the Academy was reorganized by the Nazis, Planck resigned. He remained in Berlin during World War II, explaining

> I've been here in Berlin at the university since 1889 ... so I'm quite an old-timer. But there really aren't any genuine old Berliners, people who were born here; in the academic word everybody moves around frequently. People go from one university to the next one, but in that sense I'm actually very sedentary. But once I arrived in Berlin, it wasn't easy to move away; for ultimately, this is the centre of all intellectual activity in the whole of Germany.

Remember, by the war he was 80 years old.
His life was full of personal tragedy. He lost his first wife in 1909. They had four children, twin daughters and two sons. Both of his daughters died during childbirth in 1917 and 1919. His youngest son was killed during World War I in 1916. He had one son with his second wife, but remarkably he was tortured and executed by the Gestapo as he had been a part of a plot to assassinate Hitler in 1944.

Planck's home was destroyed in an allied air raid in 1945 and he lost all of his possessions, including his notebooks of a lifetime. When the allies arrived in Berlin he was rescued as an elderly, homeless refugee.

Remarkably he became president of the Kaiser Wilhelm Gesellschaft in 1945 and worked until the end of his life to try to re-establish German science. Remember, it was Planck alone who took the unknown patent clerk's odd scribblings seriously in 1905. Such was his innate fairness and devotion to science.

> In the temple of science are many mansions, and various indeed are they that dwell therein and the motives that have led them thither. Many take to science out of a joyful sense of superior intellectual power; science is their own special sport to which they look for vivid experience and the satisfaction of ambition; many others are to be found in the temple who have offered the products of their brains on this altar for purely utilitarian purposes. Were an angel of the Lord to come and drive all the people belonging to these two categories out of the temple, the assemblage would be seriously depleted, but there would still be some men, of both present and past times, left inside. Our Planck is one of them, and that is why we love him.
> Albert Einstein on Planck's 60th birthday celebration.

16.3 Things Were Heating Up

Complicated ideas are sometimes reduced to catchphrases and quantum theory is no different with its famous motto of "wave-particle duality." Like many bumper-sticker phrases, while a lot is left unsaid, a hint of the truth still shows through. The realization that light waves and light particles somehow share a common reality emerged as an unexpected and unwelcome outcome of ordinary scientific problem-solving.

Waves were well-motivated as we've seen. Thomas Young's demonstrations in the early 1800s seemed to settle the matter of the nature of light: **waves**. Maxwell's theory of electromagnetism was a story about: **waves**. By 1900, light is a: **wave**. Absolutely. No question.

Wait. *So, if light is a wave, then it's not a particle.*

Glad you asked. *Of course, you're right. Showing that light behaves like waves would seem to be simultaneously a disproof that it could also behave like particles. Nobody questioned the logic of this. Until you know who: our patent clerk.*

Few cities in the world were as self-confident as turn-of-the-20th-century Berlin. Increasingly, following Germany's 1870s political unification, world-leading progress in industry, arts, science, and militarism were just a part of life. Theirs was at once an vibrant intellectual environment, within a simultaneously conservative and authoritarian society. It was these latter aspects which troubled the young Einstein and the former which nonetheless pulled him to the city. Unusually connected with progressive industry, science attracted the "best and brightest" to physiology, chemistry, and physics, with successful members of one field often trained in another—Helmholtz, for example, trained as a practicing physician and physiologist and functioned as a physicist.

16.4 Everything Radiates. Everything.

George Carlin says that you can't "preheat" an oven,[1] but this seemingly trivial observation is subtly incorrect. Carlin's oven may not be ready for cooking, but it's still emitting heat.

[1] Just like you can't "preboard" an airplane!

In the mid-19th century, it was realized that heat emission was just a long wavelength (infrared) version of Maxwell's electromagnetic radiation, with a λ just a little longer than that of the color red. Gradually it became apparent that all objects radiate at all wavelengths, not just in the infrared region. "Warm" objects radiate a lot in that region, while "cold" objects radiate less—but not zero. Let's roast marshmellows.

16.4.1 Thermal Radiation

Suppose you're on a camping trip with an open fire pit burning large, hardwood logs very hot. After the fire has been put out, without even touching, you can tell which logs are still hot. They glow—emitting electromagnetic waves in the visible region and just to get near them is to feel a lot of heat. Nonetheless, since it's electromagnetic radiation, Maxwell's equations should describe it, right?

As they cool in the crisp night air, they cease to visibly glow—now it looks like a log—but it would still be warm for some time afterwards. You could detect this reduced warmth by touching it, which might still be unpleasant, but you could sense the temperature without touching—just by putting your palm close to the its surface. The log is *still* emitting electromagnetic radiation, this time at a frequency that your eyes can't detect, but your skin can. Again... *this* should be describable by Maxwell's equations! In essence, the log is both a visible light and an infrared "light" antenna (bright and hot) and when it cools, largely an infrared antenna (mostly just hot).[2]

[2] You can still see the log, but that's because once it's cooled it's reflecting light into your eyes, not because it's emitting its own visible light.

What happened? Of course we say that the log "cooled," but a microscopic observer in the log would say that the agitated molecules of the its atomic lattice slowed down their vibratory motions. Because these molecules involve electric charges and, as we have seen, the acceleration of charges produces electromagnetic radiation, as they vibrate they emit waves that you call predominantly heat and light. The lattice motion slows down, and the dominant frequencies of the radiation change.

But, by the next cold morning had it stopped radiating? No, the molecular motions haven't stopped, but the radiation is much different in character. The log has resumed its dark, now charcoal-like appearance and its temperature has equalized with the air. It's not ice cold. . . it's still radiating in the IR region, just less.

16.4.2 Blackbody Radiation

Suppose, once the log was hot the previous night you had gingerly picked it up and placed it in an enclosed, insulated box, with un-shiny walls. Now, just like overnight, the log continues to emit electromagnetic waves of energy...but they don't disappear into the atmosphere, they encounter the walls, which warm up and *themselves* emit back into the enclosure and the log and so on and so on. It doesn't cool, since the box is insulated from the outside. The log and walls are radiating like mad and it would be classically appealing to imagine a *linguine* mixture of electromagnetic sine waves inside the box at thousands of different wavelengths all to-ing and fro-ing such as in Fig. 16.3.

The box walls and the log will eventually reach a state in which they each emit as much radiation as they absorb. If there is no reflection, the absorption becomes total at all wavelengths and its emission is only due to the thermal motions of the wall's surface: such completely absorbing (and emitting) objects are traditionally called "blackbodies." If you cut a small hole in the box and carefully measured the wavelengths of the radiation that escaped, you'd be taking an unbiased sample of the radiation that's filled the box and "bouncing" all around.

Definition: Blackbody Radiation.
Thermal radiation from a body in equilibrium between emission and absorption.

Now suppose that instead of a log, you heat a chunk of porcelain, or a shaft of steel, or anything else *to the same temperature as your log* and you did the same procedure: put it in the box and make a measurement of the radiation leaking through the hole. You'd find *precisely the same resulting wavelengths leaking through the hole*, regardless of the material in the box or *of* the box. The only thing that matters? The temperature. Such radiation seems to be a universal phenomenon.

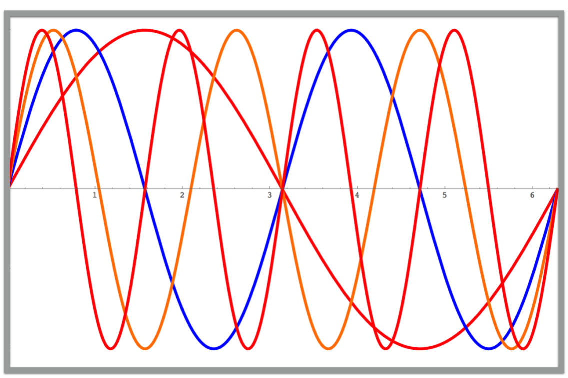


Figure 16.3: pasta2

For centuries, makers of china, blacksmiths, sword-smiths...craftspeople with an oven and a need to bring products to certain temperatures learned that they can accurately guess at the temperature of a glowing object by looking at its color—sword[3], porcelain, or whatever. Like our log and its friends in our box, the trick of relating color to temperature was found to be a universal phenomenon...*it doesn't matter what the material is*, radiation patterns will be the same for objects at the same temperatures.

When some phenomenon appears to be universal, regardless of substance and apparently dependent on only one common variable, Nature's trying to tell us something important. So in the late-1800s, ovens with holes of the sort described above were created in order to precisely measure and characterize this radiation. Why was a problem that needed to be solved. To say that Maxwell's theory was inadequate is a

June 11, 2017 08:37

[3] The art of forging samurai swords is, by legend, 1300 years old. Among the instructions are to heat the sword "until it turns to the color of the moon about to set out on its journey across the heavens on a June or July evening."

huge understatement. It was a disaster. Everyone all failed to explain this most everyday of phenomena: how objects radiate warmth.

Box 16.1 A Word About Temperature(s)

If you watch the weather on TV, you know that we're confused about how to measure temperatures. The rest of the world is not confused, just us since we continue to use an old temperature scale named after the German and Newton contemporary, Daniel Fahrenheit. He was one of the first to use mercury as the medium that would expand and contract with temperature since it would register all the way to the boiling point of water, which alcohol thermometers could not. There's a bit of murkiness about what he did and what was redone after his death. Here's the gist:

What do you do to your sidewalks when it's icy? You put salt (well, these days a different crystal) on them since salt water freezes at a lower temperature than plain water, so your salted ice sidewalk becomes water unless it's *really* cold. How cold does that work? Well just about 0 degrees Fahrenheit and that was the low "set point" that he used to define his scale. Another was when regular water freezes, and the highest was the temperature of a healthy human body. With those three reference temperatures, he then extrapolated to the point at which water boils. The actual numbers that he assigned and those that we use today are different. We use the difference between water's freezing—32°F—and boiling—212°F—to be exactly 180 degrees. This makes the conversion to... yes, that other scale... easier. If I refer to Fahrenheit temperatures at all, I'll write something like "50°F."

The Celsius scale of temperature is named after the Swedish astronomer Anders Celsius, who lived a little after Newton who defined the set points of freezing and boiling plain water and named them 0 and 100.[4] If I refer to a temperature in Celsius, I'll write "-273.15°C." The scientific temperature scale takes into account that there is a temperature limit, a floor that is "absolute zero"—nothing can be colder than absolute zero: the Kelvin scale.

The temperature at which all molecular motions would stop is an unreachable, theoretical limit that's useful for use in physics, chemistry, astronomy, and engineering. It's called the Kelvin scale and it's defined to have the same single-degree unit as that in the Celsius scale, but the zero of the Kelvin scale is that useful absolute zero. That's about -273.15°C The rough numbers to remember are that room temperature is just about 300 K (that's the real way to refer to Kelvin, you don't use the word "degrees," you just say "50 Kelvin" or "50 K." 300 K corresponds to about 27°C and 80°F. For dramatically hot temperatures, I'll sometimes just refer to "5,000 degrees" because, really, the difference between Celsius and Kelvin is small at that level (but I'll really mean Kelvin!). Converting among temperatures is pretty easy, if you're used to it. We'll not do much of this and if you need to, you can always ask Mr. Google.

[4] We now call this scale Celsius, but prior to 1948, it had been called "centigrade" with "100" figuring into the Latin prefix, "centum."

16.4.3 Radiation At Different Temperatures

There were essentially two issues that physicists thought should be understandable about radiating objects using the new electromagnetic theory: the *total amount* of energy radiated at a given temperature and the *frequency spectrum* of the radiation—how much energy is radiated at each frequency (or from $c = \lambda f$, the wavelength distribution).

Wait. *I don't think I understand what the "frequency spectrum" means.*

Glad you asked. *Think of it this way. You can learn the outcome of a baseball game by looking at the final score, which adds up all of the activity for the whole nine innings. That's like the total amount of energy. But there's "structure" in that total score, namely how those totals are distributed within each inning. So the individual inning results are sort of like the frequency spectrum.*

Let's think about our hot log with these two things in mind. Let's pretend that we've got two, 1 cm^2 square devices—like the size of a postage stamp—that each measure the energy of radiation that falls on them, but one, "H," is only sensitive to heat and the other, "V" is only sensitive to visible light.

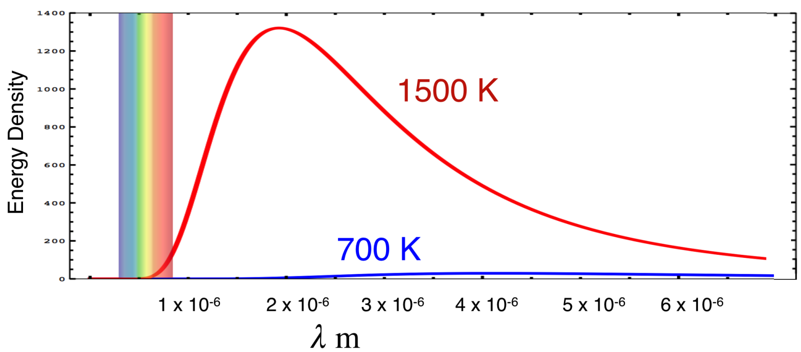


Figure 16.4: blackbody1500700

Wait. *Those must be very special detectors.*

Glad you asked. *Well, as pieces of technology, yes. But you come from the factory with those very detectors as standard equipment. Any 1 cm2 patch of your skin is blind to visible light, but very sensitive to heat radiation, while your eyes are largely blind to heat, and exquisitely tuned to be sensitive to visible light!*

Let's put V and H in each hand and stand a meter away from the hot, glowing log, which we'll assume is at a temperature of about 1500 K. Both devices are recording like mad! But how much radiant energy does V detect as compared with H?

Figure 16.4 shows the modern description of the intensity of the radiation on the vertical axis (don't worry about the actual units) and the radiation wavelength on the horizontal axis in meters (these wavelengths are closer to microns, μ m, 10^{-6} m) for two radiating bodies. The red curve is the radiant intensity as a function of wavelength that would be emitted by our 1500 degree log. The overlaid vertical colored stripe shows the visible wavelength region which you can see just barely clips the total spectrum in its visible red tail.

Not much radiation contributes to our visible, glowing log-light, and what there is is going to be mostly red in color. But boy. There is a lot of the energy output in the 2 micron region—beyond the visible—but smack among the "infrared" wavelengths. This is a special value for us, since for wavelengths longer 2μm radiation falling on human skin isn't reflected but rather begins to penetrate the epidermis stimulating nerves that signal the sensation that we detect as warmth. In this case of our log, at 1500 degrees without touching the surface, you'd detect a lot of warmth.

As the log begins to cool what happens to the radiation? Let's imagine that it has been reduced by almost a factor of two, to 700 degrees and the blue curve in Fig. 16.4 shows the intensity at that temperature. That relatively small temperature reduction results in a huge reduction in the amount of radiation.

16.5 Early Research

You've maybe heard that all hell broke loose in physics in the year 1900? *This* is that hell.

Just before the seminal work that caused the ruckus, scientists had learned how to predict just three things: how much total energy is radiated by a heated object, the peak wavelength radiated by a heated object, and the distribution of radiation for short wavelengths (high frequencies). These are called: Sefan's Law, Wein's Displacement Law, and Wein's Law. Wilhelm Carl Werner Otto Fritz Franz Wien (1864-1928) received the Nobel Prize for his work on radiating bodies in 1911.[5] These three descriptions weren't a fundamental description of what is happening, but essentially "curve-fitting" descriptions of the data. Nobody was satisfied with this.

[5] ...or for having the longest name in recorded scientific history

16.5.1 Stefan's Law

Josef Stefan inferred 1879 that the amount of energy radiated per unit area per unit time (called a "flux") was proportional to a simple, but large function of just the temperature of the radiator. Today, we call this Stefan's law:

$$u(T) = \sigma T^4.$$

This fourth power of the temperature is a lot. If the temperature doubles, then the energy output increases by a factor of 16 (2^4). You instinctively know this already. Think about how different your skin feels if exposed to the air for an hour on a 90 degree day in August as compared with one in October at 45 degrees. The constant of proportionality, now called Stefan's Constant, has the modern value of $\sigma = 5.670400(40) \times 10^{-8}$J·s^{-1}·m^{-2}·K^{-4}. (Notice that the units are energy per unit time per unit area per T^{-4} as you would expect.) This is a general relationship between the energy and the temperature and so σ is independent of the material.

The reason it's an interesting story is that after studying laboratory-based arc lamps and comparing with light from a telescope, he then *took the temperature of the sun* by measuring the energy that was captured in an area on the Earth. He found that the Sun's surface temperature must be approximately $5,430\,^{\circ}$F.[6] Further, in 1885 his empirical relationship was derived theoretically by his student Ludwig Boltzmann who concocted it from thermodynamic and electromagnetic theory.

[6] A more modern value is closer to 5780 C.

So: increasing the temperature of any object (T^4) increases the total energy that object radiates enormously! Why?

16.5.2 Wein's Displacement Law

The second thing that was known through experiment and some mathematical imagination by Wilhelm Wien in 1893 was at what wavelength the peak in the power curve would occur:

$$\lambda_{\text{peak}} = \frac{b}{T}$$

where $b = 2.90 \times 10^{-3}$ m-K and is called Wein's Displacement Constant and is shown in Fig. 16.5.

Figure 16.5: weinsDisplacement

You can see that the peak of the 1500 degree curve in Fig. 16.4 occurs at just about 1.8 microns which matches his prediction in Fig. 16.5. From Stefan's measurement that the Sun has a surface temperature of $5,430\circ$Cwe can calculate what wavelength corresponds to the highest intensity and we'd get:[7]

[7] $5,430$Cis about 5703 K.

$$\lambda_{\text{peak}} = \frac{2.90 \times 10^{-3}}{5703} = 0.509 \mu\text{m}$$

which is right in the middle of our visible sensitivity. It's green-yellow-orange-ish. (The atmosphere changes some of this before we see it.)

Here again, we have an interesting hint and some cute predictions. But they didn't have a model of what's going on.

16.5.3 Wein's Law

Yup. Another Wein's Law, this one is the "distribution" law. This is a little more complicated and is strictly curve-fitting:

$$u(\lambda, T) = \frac{1}{\lambda^5} a e^{-b(\lambda T)}.$$

The constants were gathered by comparing the mathematical shape with experimental results and the consequent curve does a pretty good job of matching the observation. But only at the highest frequencies, or shortest wavelengths... in the ultraviolet region. For heat radiation, the infrared, it was not correct.

16.5.4 What Would Maxwell Say?

In 1900 Lord Raleigh[8] attacked the blackbody radiation problem in a program of research which lasted until after 1905. With James Jeans, they applied the most sophisticated application of classical electromagnetism, mechanics, and thermodynamics ever attempted. They used a model in which atoms vibrated in a radiator's walls and so, radiated setting up electromagnetic standing waves of many frequencies. By using the accepted ideas of how energy is distributed among components of a system that the energy density distribution of these waves came out to be:

$$u_R(f, T) = \frac{8\pi f^2}{c^3} k_B T$$

where k_B is called Boltzmann's Constant[9] and f refers to the frequency of the radiation. This formula fit the data in the low frequency (large wavelength) region—where Wein's formula failed. But, it is obviously nonsense! The energy increases according to the square of the frequency—without bound! There is no upper limit and the energy could approach infinite quantities which was an unwelcome embarrassment. Within the rules of Maxwell's electromagnetism and the energetics of thermodynamics and mechanics, this calculation is unassailable. It has to be right, or the rules are wrong. The Rayleigh-Jeans Formula became the poster child for the failure of classical physics.[10]

[8] The great John William Strutt (the future Lord Rayleigh) had a problem when he was a student—he was brilliant and at the top of his class at Cambridge. But, his problems are not like our problems: as a future Baron Rayleigh, a scientific career was considered a significantly lower-class occupation and was not popular with his family. Nonetheless, he pursued his calling and, as a wealthy man, didn't need an academic career and could devote himself to independent experimental and mathematical research, which he did at his family estate. See, not our kind of problems. He was the first to explain, among many other things, why the sky is blue and he shared the 1904 Nobel Prize for his discovery of the element argon.

[9] Ludwig Boltzmann was a theoretical physicist who also worked during the late 19th century on problems of thermodynamics. He was a strong believer in an atomic picture of matter and was belittled severely for these views. Nonetheless, he developed statistical theories of how atoms would behave in a gas and derived thermodynamic parameters from this model. The toll of abuse that he suffered was too much for him and he committed suicide while on vacation with his family. Tragically, his death was less than a year after Einstein essentially demonstrated the validity of the atomic hypothesis. Had he held on just a little longer, he would have learned of the obscure patent clerk and his model would have been vindicated. A part of his model was a constant called now Boltzmann's Constant that relates energy and temperature in the relation $E = k_B T$. The value is $k_B = 1.3806488 \times 10^{-23}$ m^2 kg s^{-2}K^{-1}.

[10] The story of quantum theory is often told as if the eventual solution was pursued as a reaction to the failure of the Raleigh-Jeans formula. In fact, their work was published *after* Max Planck had solved the problem. Yet, nobody doubted that some missing ingredient to the Rayleigh Jeans model was all that was necessary in order to rid it of its embarrassment and nobody accepted the bizarre, real solution, including its father.

16.6 The Blackbody Spectrum

Let's skip ahead a little and see what the actual energy distribution from radiating substances looks like.

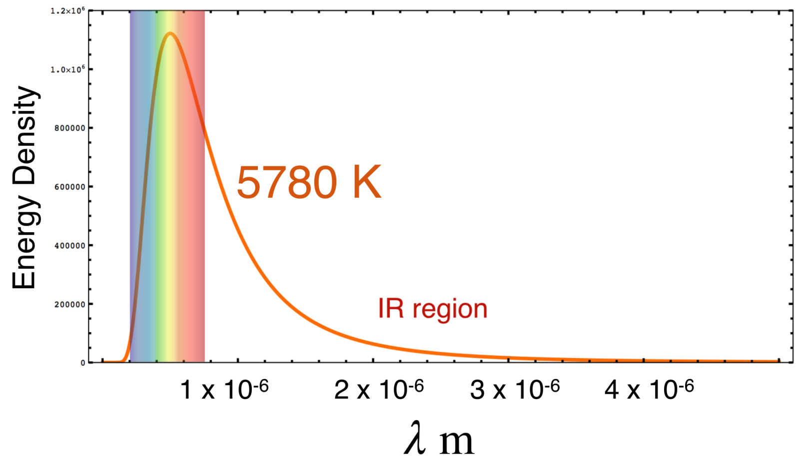


Figure 16.6: blackbodySun

Figure 16.6 is particularly instructive for a number of reasons. It shows the energy density (an energy per unit volume within the blackbody cavity) at different wavelengths for a blackbody at 5780 K, the Sun's temperature. What this means is the following: Look at the curve which peaks at a wavelength just under $\lambda \cong 0.5 \times 10^{-6}$ m. (This is a half of a micron, μm, or it's also 500 nm, or 5000Å. As you can see, this is in the visible wavelength band and the blend of almost all of the colors looks almost white to us.) Follow the orange curve to the right, to that 2 micron point which is where we feel the pleasant warmth of our Sun. Stefan's law tells us that if the Sun were just a little hotter, human life would probably not have developed since it would have been too hot on Earth. That we have evolved a vision sensitivity in the wavelength range dominated by the Sun's emission spectrum seems pretty natural.[11]

[11] Further, some animals such as snakes and bats have sensitivities further to the infrared than we, indicating an evolutionary preference for detection of heat (the infrared region), while others, like penguins, can sense into the ultraviolet region.

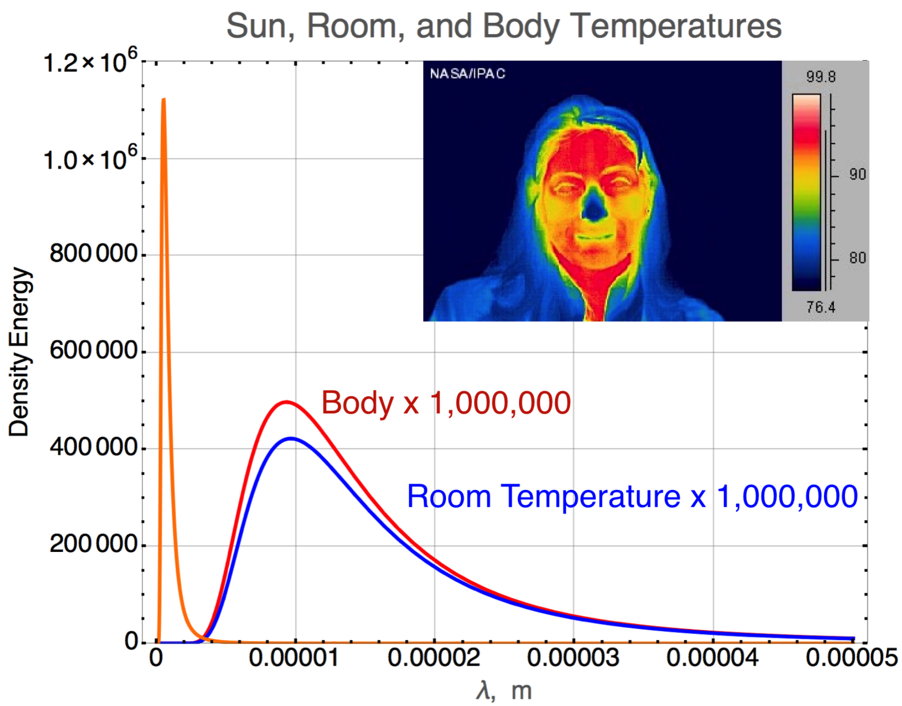


Figure 16.7: bodysunroom

There is another temperature scale of interest to us, namely "body temperature," where we are most comfortable which is about 310 K (about 98.6°F). Figure 16.7 shows the same solar spectrum along with that of a blackbody at around human body temperature...multiplied by a factor of 1,000,000 (!) in order to show its wavelength distribution on the same scale. Notice that it peaks at about $\lambda \cong 10\mu$m, which is just about at the wavelength at which humans radiate most strongly.

You know, you're the little heat-oven. From Stefan's Law, Eq. 16.5.1, your heat output is quite high because of that fourth power of temperature. If your body area is about a meter squared and your body temperature is about 310 K, then your radiation is about 500 Watts. That's a lot of heat! But you're not in isolation, you absorb a lot of heat, so you don't generate all of your output from Big Macs alone (which would require about 10 a day). No, your equilibrium temperature negotiation with your environment given that you absorb, emit, and wear clothes is about 50-100 Watts. That's right. You're effectively a reading-lamp-worth of heat.

Figure 16.7 shows radiation from a body temperature of 310 K compared with that of room temperature, which is customarily chosen to be around 300 K (remember, about 80°F). No wonder a roomful of students can warm up a classroom, each of you radiating like a 50 Watt light bulb.

16.7 The Quantum Is Born

It's not unusual to find that physicists, theoretical and experimental, often continue to work on problems that fit their own specialties. Like most people, they are comfortable in some areas, and less so in others. Max Planck—at the top of his career just before 1900—was an expert in thermodynamics and less enam-

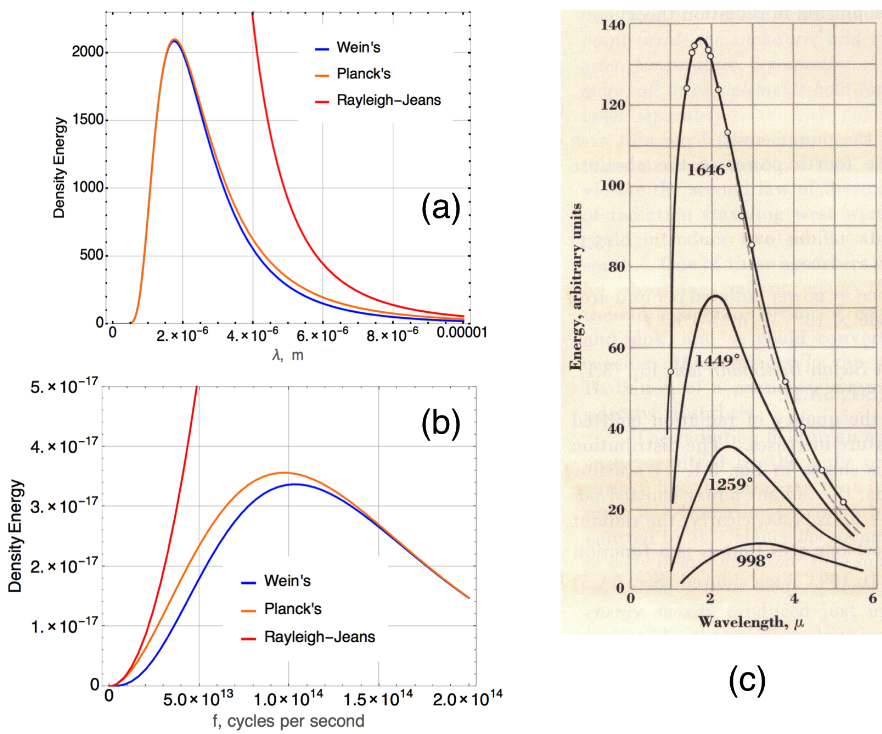


Figure 16.8: everything

ored of the use of statistical methods in physics championed by Boltzmann. As a result, he came to the problem of blackbody radiation with different tastes and approaches than those employed by Raleigh and others. In December of 1900 after a series of fits and starts, he managed to account for the full curve that fit the intensity of radiation as a function of wavelength. He announced it in a meeting in his home institution in Berlin and then one of his experimental colleagues spent the whole night comparing it with the most recent infrared data that were fresh from experiments in the basement: it was bang-on. The curves that I've shown in the previous sections are from Planck's formula and Fig. 16.8 (c) are data and Planck's prediction from that 1900 meeting showing perfect agreement.

So, great. A curve that fit the data. At first, this success was, in a way, mechanical: he introduced an oscillator model like Rayleigh and added some mathematical steps (following Boltzmann) which were very difficult to interpret, but which seemed necessary. He was a fine physicist and was devoted to going beyond just producing a formula that fit data—he needed an actual physical interpretation to go along with it and was determined to find one.

So now we have three different models to describe thermal electromagnetic radiation: Raleigh-Jeans' model, based on Maxwell, Wein's empirical curve, and Planck's model. Wein's fit well, but as the experiments got better, it began to not work.[12]

[12] The problem arose in the IR region, and it's easy to see why. Wein's model was based on measurements, so new data wouldn't have been in his fitting. Second, the IR measurements came later since it was very hard to make precise measurements of objects that were by their very nature near room temperatures. It's why IR telescopes are hard to build and operate. . . if you're trying to detect minute amounts of radiation at nearly the same temperatures as your surroundings, then it's very hard to distinguish the new radiation from the inherent warmth of your detector!

Figure 16.8 shows exactly this situation. The Planck formula works perfectly! Let's look more closely. Figure 16.8 (a) on the upper left compares all of the predictions for 1646 K, which correspond to the top data represented as the circles in Fig. 16.8 (c) on the right. I've also shown side by side the predictions as a function of frequency, Fig. 16.8 (b). What do we see:

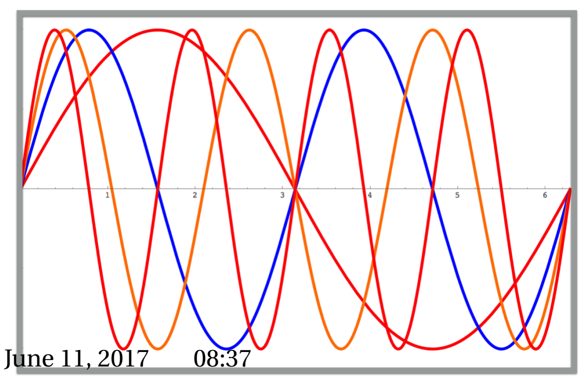
- The Raleigh-Jeans formula fits the Planck model at the longest wavelenths, or the lowest frequencies and fails at short wavelengths and high frequencies.
- The Wein formula fits the Planck model in the opposite regimes: good, for short wavelenths and high frequencies, and bad for the opposite.

How to explain this? Maxwell's theory is surely not wrong! Is it?

16.7.1 Max Planck's Interpretation

So Max Planck had a problem. He'd found a way to solve this thorny, old puzzle but at the price of a strange assumption that he had to make in order to make it work. It's important to realize that he had much going for him: his model fit the data over **all** of the wavelengths that were measurable; it reduced to the classically-acceptable Raleigh-Jeans formula at small frequencies, and the Stefan-Boltzmann result fell out when all of the intensity was added up at each wavelength. Further, he could predict Avogadro's Number, Boltzmann's Constant, and the electric charge. The exchange for all of this success was the introduction of a bizarre idea and the tiniest fundamental constant into physics yet devised.

First, why did the Raleigh-Jeans model break down? Inherent in this "classical" description of how radiation would fit into a finite sized box was that all frequencies are allowable with equal probability. Figure 16.9 is a sketch of what I mean. The rules of Maxwell's theory say that the ends of an E&M wave must be nodes at the walls. So the red curve is the longest wavelength wave that could fit in the box, and from there we could imagine adding more and more waves with shorter and shorter wavelengths—and so higher and higher frequencies!—without end. In the figure we add the yellow curve and then the blue curve and then more and more, each one at a higher frequency than the one before. That's the "linguine" that I referred to above. So since there's no limit to the shortness of wavelength, or more easily seen now,



June 11, 2017 08:37

Figure 16.9: pasta2

no limit to the highest frequency that can fit, the energy in the box would rise to infinity! That's what the red curve in Fig. 16.8 is doing. Planck's idea tamed that high-frequency behavior, which was dubbed the "ultraviolet catastrophe" by the more dramatic of physicists.[13]

His model was that the walls of the blackbody-box indeed contained little oscillating charges—just like Raleigh and Jeans had done—but that they were restricted in the wavelengths at which they could oscillate. He had no physical reason as to why they would be restricted, it was a hypothesis and he built it into his model.

This is sort of hard to visualize, so think about this. Sound is a wave phenomenon and our ears are capable of picking out any frequency, pleasing or not. Frequency analyzers are equally capable of detecting any frequency—sound is not restricted to particular tones. But pianos are. When you strike a key on a piano you produce only particular notes, or frequencies but that in no way says anything about what frequencies sound is capable of, just what pianos are able to make. The walls of a blackbody box are like the piano. Planck postulated that the walls could produce only particular frequencies, but that light and all E&M waves could be any frequency. Just not in a blackbody box.

He called these (piano-like) special radiation bits "bundles" which were later dubbed, "quanta"[14]by Philip Lenard.[15] And he found that he needed to insist that energies of his radiation bundles were directly proportional to their frequency.

This is the beginning of Quantum Theory, which will evolve in the 1920s into a more sophisticated "Quantum Mechanics." The hallmark idea of Quantum Theory is that light—and all electromagnetic radiation is "quantized" into these bundles. Radiation is not continuous. One of the simplest,[16] but most profound equations in the history of physics is this statement ("Planck's law"):

$$E = hf. \tag{16.1}$$

The constant of proportionality is now called Planck's Constant, and he estimated its value in his model. It's tiny and the modern value is: $6.6260755(40) \times 10^{-34}\,\mathrm{J \cdot s}$. Planck's Constant serves as a measure of when quantum weirdness begins to matter and when we can continue to use Maxwell's equations. That's it's so itsy-bitsy explains why we don't see quantum behavior every day. The energies in our lives are much larger. It's the same idea as to why we don't see relativistic effects in our everyday life where speeds are much less than those of light.

So Equation 16.1 says three important things.

- First, it declares that if you give me a frequency, I'll tell you the energy precisely. For example an IR frequency of 1 micron corresponds to $1.98782207 \times 10^{-19}$ Joules. Not $1.98782206 \times 10^{-19}$ or $1.98782208 \times 10^{-19}$ or anything else.

[13] Why "ultraviolet"? The UV spectrum is that on the other side of the visible, the shorter wavelengths than blue or violet... a sort of short-hand for generically, "short wavelengths" or "high frequency."

[14] Here is the origin of the word "quanta" or "quantum." It is Latin in origin for "*quantus*," meaning "how much."

[15] Lenard became a Nazi and doggedly tried to prosecute Einstein for his "Jewish Physics."

[16] Did I say that some of the most profound solutions would come in very simple mathematical equations? Did I?

- Likewise, if you want radiation of $1.98782207 \times 10^{-25}$ Joules, then you must be content with radiation that has a wavelength of 1 meter. Not half a meter or 1.2 m, but 1 meter.
- Finally, zero energy corresponds to a frequency of $f = 0$ identically. We'll see later that this is not possible.

Further, one can build up different amounts of energy in only two ways: change the frequency, f, or have more bundles. In fact, the more general way to write Planck's law is

$$E = nhf \tag{16.2}$$

where n is an integer, ranging from 0 to any finite value. For any finite frequency, the radiated energy is finite and equal to hf, $2hf$, $3hf$, ...and so on.

This is strange. If we extrapolate the quantum idea to life, it makes playgrounds strange.

The idea is as if the frequency of a child's swing is only adjustable in discrete amounts. Picture it going back and forth, each time it returns I give it a little push to restore the amplitude lost to friction. Now to make the ride more exciting, I push the swing just a little harder. But nothing changes. A little harder still, nope. Again and again, more and more and then... then all of the sudden the swing suddenly goes further. And it stays there. More pushing, no change... and so on. The pushes are discrete until the right "quantum" of push is applied corresponding to the energy going from nhf to $(n+1)hf$. This weird hf is a tiny, tiny energy and could indeed be mistaken for zero in anything but an atomic environment. So we'd not see quantum effects in our swing-pushing, since they'd be tiny.

This is why the Raleigh-Jeans formula matches the Planck formula at the smallest frequencies. That's where the energies in the presumed blackbody wall-oscillators are the tiniest so the differences between classical and quantum behavior is too small to detect. Why? Because Planck's constant is so small.

Planck fought against his own interpretation.[17] In that fight he was in good company, since nobody liked his idea. Nobody, except Einstein. It's important to realize in hindsight that Planck's approach is a little contrived. Without a physical basis, he tied the hands of the blackbody wall oscillators and chose not to mess with the Maxwell interpretation of E&M radiation.

This is not the mark of someone behaving unscientifically...it's the mark of someone facing nearly insurmountable conceptual difficulties while inventing a new subject. His scientific stature surely even raised the stakes for him personally, as when one is as distinguished as he was, a mistake could carry a significant embarrassment. But, in spite of that risk, Planck followed the physics as far as he could, publicly, and forthrightly.

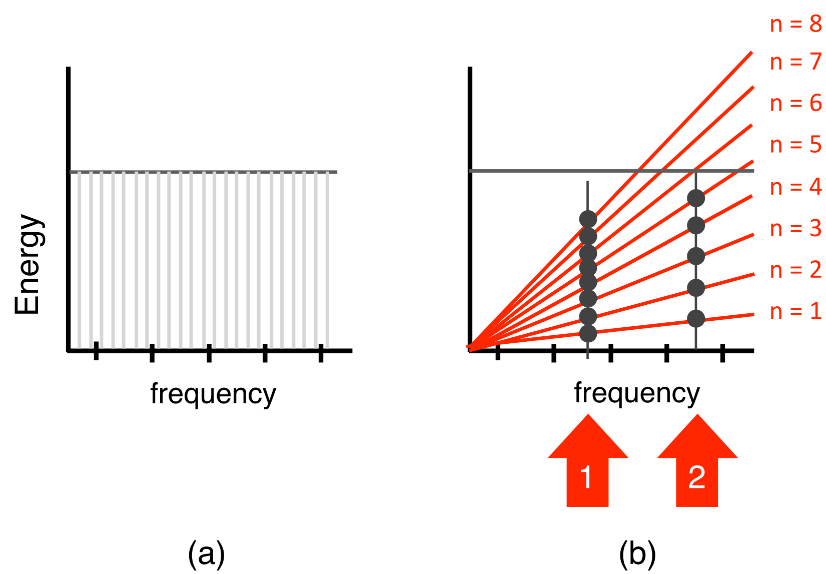
[17] Although he suspected that he had stumbled onto something significant and on a stroll with one of his young sons indicated that he had had "a conception today as revolutionary and as great as the kind of thought that Newton had."

16.7.2 How Does It Turn Over?

I've not touched yet why Planck's curve in Fig. 16.8 turns over at the higher frequencies, where the Maxwell-description blows up. This too comes from Equation 16.2.

The requirement of Maxwell's theory that all frequencies of radiation are equally likely suggests that for a given energy, we could populate it with all frequencies. Figure 16.10(a) is a cartoon of that idea. Every frequency works to create energy E.

Figure 16.10: planckexplain



Notice that Planck's law is just the equation of a straight line, passing through the origin. This straight line would have a tiny slope, Planck's constant, but it's straight nonetheless. In Fig. 16.10(b) are plotted many such straight lines from Planck's law, each corresponding to a different value of n. At the smaller frequencies, like the one labeled 1, many different values of n contribute. But at a higher frequency, like the one labeled 2 notice that many fewer values of n are necessary in order to get to energy, E. It takes fewer quanta of higher energy than at lower energy, so the curve in Fig. 16.8 turns over and falls at the lower wavelengths and higher frequencies since fewer values of n are involved.

16.8 The Quantum Grows Up

We've already met Einstein's singularly unusual way of looking at the world through relativity. In that same year that he was not working very hard at the Patent Office, he was literally changing the world in one time-gulp: remember, in 1905, now called his "miracle year," he published his relativity theory, proved the existence of atoms by explaining Brownian Motion, and correctly reinterpreted Planck's formula as a description of light. Einstein was the first Quantum Mechanic.

16.8.1 Photoelectricity, Revisited

Remember that one of the results of Hertz's experiments was the inadvertent discovery that illuminating some metals with ultraviolet (UV) light prompted a current to flow from its surface. This current could be accelerated in an electric field and was found to be negatively charged and so likely to be made up of J. J. Thomson's electrons. But, the characteristics of this "photocurrent" were inconsistent with what one would expect employing a Maxwellian electromagnetism description of the illuminating UV light. For example, the electric field, $\mathbf{E}$, of the UV light should apply a force to the electrons (remember, $\mathbf{F} = e\mathbf{E}$) and shake them out of their orbits. The *stronger* the field—the more intense the light—the higher the force and the *faster* the electrons should be ejected. But that's not what happened. Instead:

- If one changed the intensity of the illuminating light, the kinetic energies of the emitted electrons *did not change*
- If one changed the intensity of the illuminating light, the *current did increase*, as if there were more electrons.
- If one increased the frequency, f, of the light, *the kinetic energies increased*.
- If one switched the light source on and off, the photocurrent *started/stopped immediately*.

All four of these results were counter to what would be expected if UV light acted as waves.

While the early observations were qualitative, by the early 1900s physicists were beginning to make very precise measurements of the characteristics of photoelectricity and the precision continued as surface preparation and vacuums improved. By the late 1800s it was clear that the following was the sequence of events in establishing a photocurrent: one shines UV light on a highly polished surface of (predominantly, an alkaline) metal in a good vacuum. Beginning with a very low frequency, nothing is observed until at some particular frequency, f_0, a small photocurrent begins to flow from the surface. As the *frequency* is increased beyond that minimum, the kinetic energies of the ejected electrons grows: E(photoelectrons) ~ some function(frequency.

Not only did the observations appear to contradict a wave-like behavior for UV light, it also ran counter to the notions of what held electrons in the metal. It was believed that either the electrons were freely moving inside or that they were bound in oscillators. In either case, the application of a wave of light energy should cause them to absorb that energy and eventually break free, either of the surface or of the oscillator. Calculations suggested that the delays that would be expected would be of the order of seconds or even a minute or so. And yet, the near-instantaneous release of photoelectrons was measured to be on the order of nanoseconds.

So, the problems with photoelectricity caused problems for both Maxwell and Lorentz' electromagnetism as well as the most promising notions of the electronic structure of matter. It should be noted that there was another way to get electrons to be ejected from metalic surfaces...by heating them. This "thermionic" emission mechanism had a similar tendency for the electrons to not be emitted until a minimum amount of energy in the form of heat was applied. This minimum energy is called the *work function* and was the same for similar materials as the low-frequency threshold for photoemission. This led to the conclusion that the photoelectrons and thermionic electrons originate from the same material structure. All of this was going on while Planck was solving thermal radiation. Hmm.

16.8.2 Einstein's Take on the Quantum

It was into this fray that Einstein leaped with his first crucial paper of June 9, 1905, "*On a Heuristic Point of View Concerning the Generation and Transformation of Light*." (His first relativity paper was second that year, the Brownian motion paper, third.) Remember, he had been a "closet" devotee of Maxwell's theory because of its success in explaining the *propagation* of free electromagnetic waves in optical and other frequencies, Einstein was unstinting in his praise. It works. By itself.

But when light *interacted with matter* strange things happen: namely, blackbody radiation and the photoelectric effect. Einstein picked on the weak spot in Planck's argument that had somehow linked together quantized radiators with continuous radiation and drilled right into it proposing that not only would the radiators oscillate according to quantized frequencies, *but that that the quantization would remain*...that light is a kind of lumpy, resultant wavefront—not the more familiar continuous wave.

That is, Einstein insisted that electromagnetic radiation is not continuous, but is itself also quantized—particles. The particles of radiation eventually got the name, **photons**, although not from Einstein and not immediately—it was the 1926 invention of the American chemist, G. N. Lewis.

Wait. *How can a wave be a particle?*

Glad you asked. *This quantum picture of light doesn't spring easily to mind! Previously "obvious" wave-like observations would have been done with bright light, which in Einstein's picture, is light containing an enormous number of photons. We'll understand how this happens in a bit once Quantum Mechanics is invented. Even though a thorougly confirmed description of nature, it's still conceptually problematic for all of us. It's one of the best examples of having to trust what the mathematics tells us, even when inconceivable by humans.*

Einstein argued, the wavelike properties are more apparent as a kind of cooperative relationship among the photons. Bright (meaning, Intense) light hides the granularity. But, when one is dealing with the emission of single or few oscillators, then the lumpiness becomes apparent. For the moment, let's follow Einstein's argument about photoelectricity.

16.8.3 Photoelectricty Explained

If we take f to be the frequency of the light, Einstein was proposing that the energy of the particles of light was identical to Planck's formula for the his wall-oscillators. The energy of a single quantum of light—a photon—comes from Planck's law:

$$E = hf.$$

The energy of an intense beam of light of frequency f would come from the collection of individual photons

$$E = nhf.$$

And the intensity of a beam light is in turn a measure of the number, n of individual photons in that beam.

Classical Picture	Einstein's Picture
Intensity of light $\propto |\mathbf{E}^2|$	Intensity of light $\propto nhf$

He explained photoelectricity by also blending the pictures of waves and particles. The UV light that falls on a photoelectric surface? Those hit the electrons in the metal and behave as if they are *particles*. Plus there is a direct correspondence between the classical, Maxwell nature of light and the quantum nature: intensity and photon number.

The interaction kinematics is precisely "fixed target scattering" that we considered in Chapter 6.5 with one difference. The in order to knock an electron free, the photon needs a minimum energy since the metal binds the electron. If a photon does not have enough energy to liberate an electron from its atom, it can't. But according to Planck, the energy is proportional to the frequency. So if some light isn't sufficient, just raise the frequency. Then, energy conservation would demand that

$$\text{energy of photon} = \text{energy required in order to free an electron} + \text{KE}_{\text{left}} \tag{16.3}$$

$$E = hf = \phi + \frac{1}{2} m_e (v_{\text{left}})^2 \tag{16.4}$$

where ϕ is the so-called "work function" of the material, the amount of binding. KE_{left} is the kinetic energy that the ejected electron has after it's been freed. When turned around to be the maximum kinetic energy of an ejected photoelectron that could be ejected,

$$\text{KE}_{\text{left}} = hf - \phi.$$

This linear dependence on frequency in Eq. 16.8.3 was Einstein's prediction.

It was not until about 1914 that the linear relationship predicted by Eq. 16.8.3 gained ground and, in a series of famously precise and careful experiments in 1916, it was (reluctantly) determined to be correct. I say reluctantly, as it was Robert Millikan at the University of Chicago who performed them, first calling Einstein's idea "bold, and not to say, reckless" and then later lamenting in his publication, "Einstein's photoelectric equation... cannot in my judgment be looked upon at present as resting upon any sort of a satisfactory theoretical foundation. . . "

He learned to console himself with his Nobel Prize for measuring the charge of the electron.

16.8.4 The Compton Effect

Nobody would ever accuse Einstein to be a *reluctant* revolutionary, as certainly Max Planck was. Rather, he was well aware of the revolutionary aspects of all of his 1905 work, and especially the quantum idea. In 1916 he kicked it up a notch to conclude that the light quantum would not only kick out electrons, but if it hit a molecule, then that molecule would recoil mechanically and that the photon that recoiled would itself have less energy, like pool balls, and so its frequency would be reduced.

So we can just use plain-old energy and momentum conservation for the scattering process, just like we had in Chapter 8:

$$\begin{aligned} A + B &\rightarrow A + B \\ \text{X-rays} + \text{atom} &\rightarrow \text{scattered X-rays} + \text{recoiling atom} \end{aligned} \tag{16.5}$$

Energy conservation would be the simple:

$$E_0(\text{photon}) + E_0(\text{atom}) = E(\text{photon}) + E(\text{atom})$$

Here's what we would know:

- the photon's initial energy and frequency: $E_0(\text{photon}) = hf_0$
- the atom's initial rest energy: $E_0(\text{atom}) = m_0(\text{atom})c^2$
- the photon's final energy and frequency: $E(\text{photon}) = hf$

If one could reliably prepare x-rays of a known frequency and measure frequencies of an x-ray beam, the particle hypothesis for photons could be tested. The E energy could be determined by measuring the frequency of the scattered x-ray beam.

This process is the first time that special relativity and quantum theory were combined together! That's because in order to calculated the momentum of the incoming photons, Einstein had to take his relativistic energy equation to places...that it had never gone before.

Remember the relativistic formula for the total energy of an object having momentum, p

$$E = \sqrt{p^2c^2 + m_o^2c^4}.$$

Although unimaginable in 1905, if a body has no rest mass, then m_o=0 and

$$E = pc.$$

A photon has energy and it has a momentum and that mixes up particle and wavelike characteristics:

$$\begin{aligned} E &= pc = hf = hc/\lambda, \text{ so}: \\ p &= \frac{hf}{c} = \frac{h}{\lambda} \end{aligned} \tag{16.6}$$

Equation 16.6 is the expression for the momentum of a single photon. It's related to its wavelength or its frequency! Talk about mixing metaphors!

So we can add to what we know:

- p_0(initial photon) $= hf_0/c$
- p(final photon) $= hf/c$

and a momentum conservation equation could then be written.

If this pool ball picture is right, then since the outgoing atom would carry away energy, then the outgoing photon would have less. So $E < E_0$, which means that $f < f_0$ or $\lambda > \lambda_0$.

Arthur Holly Compton

By 1923—eight years after Einstein's idea of the photon momentum, and seventeen years after his original prediction of photons, Arthur Holly Compton, an American, succeeded in slamming the door tightly against any doubt of the particle nature of light.[18] He studied the elastic scattering reaction

$$\text{X rays} + \text{Carbon} \to \text{X rays} + \text{Carbon} \tag{16.7}$$

$$\gamma + e \to \gamma' + e'. \tag{16.8}$$

[18] When I was seven or eight years old, I would occasionally play with a neighbor's visitor in the suburbs of Chicago. I can remember my parents marveling at the fact that his grandfather had a Nobel Prize. My sometimes-friend's name was "Compy."

This is a standard notation, where the Greek letter gamma (γ) always represents a photon and the primes indicate here that the scattered particles have different characteristics—different states—from the initial ones. The use of X rays was in part to facilitate the measurements of the final state, as we'll see. But, they are also sufficiently high in energy that the electron in the carbon target is essentially at rest relative to the incoming photon. This greatly simplifies the mathematics of what to expect. It is strictly "billiard ball" kinematics, albeit with tiny, bizarre billiard balls.

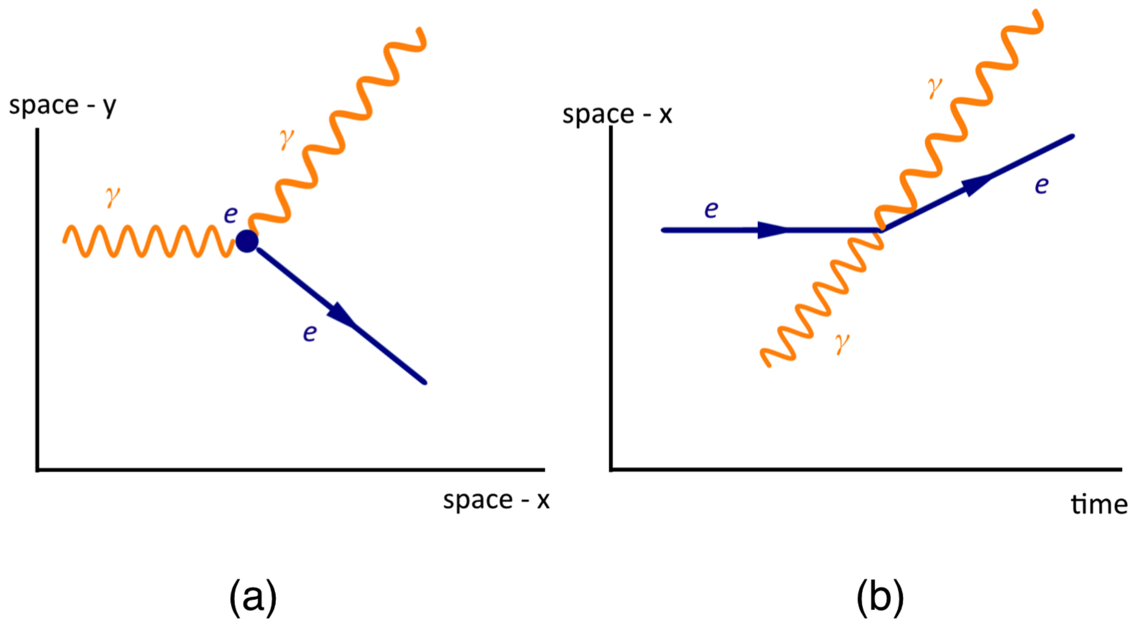
Particle 2

photon

symbol:	γ				
charge:	$0e$	mass:	$m_\gamma = 0\ \text{MeV/c}^2$	spin:	1
category	boson, aka intermediate vector boson			category	elementary

Figure 16.11 shows our three diagrams for this process—now called "Compton Scattering." And the results are shown in Fig. 16.11.

Figure 16.11: comptondiagrams



The results were conclusive. In Fig. 16.11 the solid curve is the prediction of Einstein's model and the data fall right on top at a wavelength that's a little longer than the initial beam's. It took this long for Einstein's vindication about the quantum nature of light. Acceptance came slowly: in 1916, Planck nominated him for membership in the Prussian Academy of Sciences, writing in part,

> That he may sometimes have missed the mark in his speculations, as for example in his hypothesis of light quanta, cannot really be held too much against him. For it is not possible to introduce fundamentally new ideas, even in the most exact sciences, without occasionally taking a risk.

Even his Nobel Prize in 1921 was somewhat subdued in nature: "The Nobel Prize in Physics 1921 was awarded to Albert Einstein 'for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect'." In fact, this was a scandal for the Nobel Prize organization. There had been a dispute in the committee borne of antisemitism and animosity towards Einstein's pacifism. As a result there was *no prize awarded in physics in 1921*, even though the world knew that Einstein was due at least one, if not more than one Nobel. Instead, the 1921 prize to him was awarded in 1922. He rarely mentioned it himself, and did not attend the award ceremony. His divorce decree from Mileva in 1919 stipulated that when he won the Nobel Prize, that the award money would go to her and their children.

16.9 What's the Meaning of This?

I know you're wondering how the photon can be representative of the waves of electromagnetic radiation and at the same time (!?) be a particle. We'll see some of the explanation—that we physicists, cling to as an excuse for our inability to picture this apparent contradiction—in the next chapter. But we can see the waviness and particulate features in experiment with the aid of modern photodetection techniques.

16.9.1 Modern Examples of Photoelectricity

You know, you should always wash your hands. And since the first installation in the O'Hare Airport in 1980's, in most U.S. airport restrooms it's pretty easy. You just put your hands underneath the spigot and water comes out. Magic? No, it's photoelectricity. There are three kinds of photoelectricity, all of which impact our lives in a regular way now. They are

- Photoemissive. Light falls on a surface and electrons are ejected. These are typically metals.
- Photoconductive. Light falls on a material and liberates electrons, but they don't fly off, rather they become an electrical current within the material. These are semiconductor materials.
- Photovoltaic. Light falls on the top of a two-layer material and causes a voltage to appear between the layers. These too are semiconductor materials.

These are all devices that convert light into electricity in one way or another. The first one in the list is the historically oldest—it's the story that we just went through: shine light of the right frequency on some surfaces and electrons will be emitted. It's not as widely used as the other two since photoemissive detectors are actually pretty large and cumbersome. We use them in laboratories to detect the passage of particles as we saw in Chapter **??**. What's neat about them is that by setting up voltages between "stages" in the vacuum of a phototube the little bit of electron current that originally comes from collected light can actually be amplified many times (the "gain" is a measure of that multiplication and a factor of 10^5 is not unusual) and a measurable current can result from even a small amount of light and can happen reliably within 10s of picoseconds. So we can glue or press phototubes on pieces of special plastics that emit light when a charged particle goes through them and then collect that light in the phototube. Infrared vision goggles also are photoemissive devices. Some materials (semiconductors) are even sensitive to infrared light and so one can "see" heat in the dark with such devices.

Photoconductive devices can be infrared detectors. These are the devices in the faucets in airport restrooms. If you look under the faucet you'll see a little redish (maybe) window. It contains an infrared LED which is constantly beaming low-power, invisible infrared light under the faucet. When you put your hands in the beam, it reflects back into the window where there's also an infrared photo-receiver. This in turn operates an electric switch and eventually, a valve and the water turns on. Likewise your TV remote control is often infrared with the emitter in your hand and the receiver in the set or cable box. But photoconductive materials can also be in your phone. CCDs are a particular kind of engineered materials which are sensitive to light and create current in your camera.

Photovoltaic devices create a voltage when light falls on them and they are the workhorse of photo panels on dwellings. Solar cells are increasingly made of photovoltaic materials: sunlight becomes power for your home all because of Planck and Einstein.

While the photoelectric effect described by Einstein is only strictly the first mechanism, all of these require quantum physics in order to work and so 1905 was the birth of many modern devices, so ubiquitous that today you hardly realize their sophistication.

16.9.2 What's Coming

The "quantum revolution" had a handful of intellectual and experimental boosts. Insight by one or two people caused the scientific world to shudder as revolution seemed to happen over and over again. I noted that Einstein received the Nobel Prize only for photoelectricity. While relativity was indeed revolutionary, his ideas of the quantum were held against him for many years. They were just too unbelievable to be... believed. So when the Nobel Committee finally got around its racism, only apparently the most audacious of his many worthy ideas qualified for acknowledgement. He should have received additional[19] Nobel Prize for Special Relativity, General Relativity, Brownian Motion, and his idea of quantum statistics which we'll touch on later. He had an unprecedented five Nobel-worthy successes. Photoelectricity needs to stand for all of them.

[19] A number of scientists have received two Nobel Prizes, the first being Madam Curie who received one in Chemistry and one in Physics.

Part III Physics and Cosmology of My Generation

Chapter 17
Antimatter

Paul Dirac's Second Big Score


Paul Dirac and Werner Heisenberg

Paul Dirac, 1902-1984

"Heisenberg, why do you dance?" A question of Dirac's to Heisenberg when they were ship-bound for Japan together. Heisenberg liked to dance before the dinners and replied that when there were nice girls he felt like dancing with them. Dirac thought about that and eventually asked Heisenberg, "Heisenberg, how do you know beforehand that the girls are nice?"

When Paul Dirac went to Stockholm to accept his Nobel Prize in 1933 he did all of the standard things—the parties, the banquet, and of course his address to the Royal Court and guests and family. While families of Nobel Laureates always attend the ceremony to see their spouse, parent, or child inducted into history, Paul's father didn't attend. He wasn't invited.

Quantum Mechanics was difficult enough. Relativity was tough too, but somehow a little more accessible, right? Putting them together—which everyone knew had to be done, but nobody could figure out—proved to be the opening of a floodgate that let in all manner of odd realizations of just how Nature works at the deepest level.

17.1 Goals of this chapter:

- Understand:
  - How to calculate distance, time, and speed for uniform and constantly accelerated, linear motion
  - That falling objects all have the same acceleration near the Earth.
  - How to graph simple motion parameters
  - How to read graphs of realistic motion parameters
- Appreciate:
  - The algebraic narratives in the development of the formulas
  - The shape of the trajectory of a projectile
- Be familiar with:
  - Ideas of motion before Galileo
  - Galileo's life
  - Galileo's experiments with motion

key concepts

17.2 A Little Bit of Dirac

Antimatter is the stuff of science fiction —an almost a silly-sounding thing. From blockbuster Hollywood to pulp science fiction, the idea of spooky "stuff" that somehow cancels "normal stuff" has been a part of our cultural imagination for a long time. When you first hear of it, you're skeptical but that's nothing compared with the struggles of the man who invented it—it took him almost three years of frustration before he found a way to interpret what his mathematics was telling him. After very public arguments among many people, he quietly settled in on the most ludicrous interpretation of all, which of course turned out to be the only way Nature would have it.

We're now on the path to Oz that is modern particle physics and the first bricks on that road were laid by Paul Dirac. We're going to live with antimatter throughout the rest of our story. And we'll be puzzled about antimatter right to the present day. And the adventurer who took us there was a very quiet man.

The story of Dirac's young life is legendary, not for its inspiring boy-makes-good theme, but because of the nastiness that he suffered at the hands of his father. Charles Dirac was an expatriate Swiss teacher of French in Bristol at a secondary school affiliated with the University of Bristol.[1]

"Strict" doesn't do justice to the way in which he treated Paul, and in a different way, his older brother and Paul's mother. There was a darkness in the Dirac household which arguably led to tragic suicide of his

[1] Cary Grant grew up blocks from Paul Dirac and attended the same primary school, but a year younger!

brother and a discipline that so affected Paul that even in his eighties he could still register deep emotion. He almost never spoke of his childhood, but when he did it was with bitterness.

Paul was commanded from an early age to speak only French with Charles. His mother, brother, and sister ate in the kitchen, while Paul had his wordless dinner at the dining table with his father. Silence was a matter of practicality since any mis-step in his French would result in severe punishment. He also suffered from a delicate digestion throughout his life and there were times when even at the dining table he could not keep his food down. Yet even after such embarrassments, he was still forced to resume eating until his plate was empty. So speaking only when it was absolutely required and eating sparingly were his early choices and habits throughout his life.

Paul was a gifted mathematics student. His brother was as well and wanted to study medicine, but Charles would have nothing but Bristol-educated (free), employable sons so both graduated as engineers. Paul received a bachelor's degree in Electrical Engineering at the age of 19 and after a disastrous industrial internship, went back for a second degree in mathematics. By the time he managed to get admitted to graduate school at Cambridge, he was only 21 years old with two university degrees. At first his classmates were mystified by strange student in the front row correcting the mistakes of lecturers...and then recognized him as a genius.

Years later, Paul's daughter speculated that this was a tragic blow to the uncle she never met as he was miserable in his short life and took his own life when Paul was in graduate school.

The Cambridge Physics Department was associated with The Cavendish Laboratory and Trinity College, which boasted Isaac Newton as an alumnus. By the time Dirac arrived, Rutherford had settled in as Cavendish Director where experimental physics was in the now familiar full-speed-ahead-Rutherford mode. But theoretical physics lagged. Arthur Eddington of the Einstein-eclipse fame, had been named the Director of the Cambridge Observatory a few years before Rutherford arrived and so Relativity was well-represented (but not very popular outside of Eddington's circle). But Quantum Mechanics was less well practiced than General Relativity.

Paul wanted to continue to study Relativity, but didn't get his first choice of faculty advisor and was instead admitted to St John's College, where mathematics and mathematical physics was studied. The man who took him under his wing was Ralph Fowler, exactly the right mentor. Fowler was one of the few experts on Quantum Mechanics at Cambridge and so Paul quickly became expert in that field—so much so that he'd transform it before he graduated: Yes, before Dirac even gained his Ph.D., he was well on his way to revolutionary discoveries greatly impressing the mighty Eddington, as well as everyone on the campus and the continent.

The British would say, "where maths was studied."

Cambridge University is organized in a set of colleges, of which there are 31 now. This is a part of its 500 year history.

Paul was an odd companion—silent at the common St Johns College meals to the point of exasperation for those around him—he did eventually establish friendships with students and faculty and participated fully in the seminars and journal clubs. Yet, he seemed to relate to people in a highly mechanical

His working habits were to completely isolate himself in a spare room for calculation six days a week. Then he always reserved Sunday for walking, and would take off in his only suit and tie for nearly all-day treks in the countryside, a habit he maintained his whole life.

Figure 17.1: Dirac and his family.

Unlike his Cambridge regimen, at FSU, he was much more sociable and found the American campus and department life agreeable and genial. After a lifetime of working one way, he completely switched!

way, seemingly incapable of being insulted or embarrassed and accordingly was often unconsciously the source of discomfort for others. He wasn't intentionally unkind, he was simply unemotional— matter-of-fact to a maddening degree. Not unfriendly, he just wasn't...anything. Eventually people learned to accept him, even affectionately, and he, them. But Dirac Stories abounded.

Paul's letters home were regular but were typically only a few lines, almost never referring to his work. Even when his fame was growing, the Bristol Dirac household had to read of his successes in the newspaper or learn from a neighbor. It just didn't occur to Paul to tell them. Or he didn't want to.

This story does end happily. Paul eventually married the sister of a colleague whom he met at Princeton on one of his scientific stays in the U.S., officially adopting her two children and then together having two of their own. He was appointed the Lucasian Chair (remember, that's Newton's old position) which he maintained following his regular 6+1 weekly schedule (see the side-note above) for almost 30 years, retiring from it at 65 year's of age, as required.

But he wasn't ready to quit and accepted a position at Florida State University where he enjoyed the more personable nature of the American faculty life. So out of a difficult beginning, a happy life evolved. Sufficient reward perhaps for his intellectually lonely work. Because he was often way ahead of everyone.

It is difficult to imagine modern physics that doesn't hold a debt of gratitude to Paul Dirac. To many, he was the second most important theoretical physicist of the 20th century, but because of his low-key manner and the deeply complicated topics that he mastered, he's not as well known today as perhaps he should be. (Indeed, it's hard to imagine his picture on the cover of *Time* magazine!) His contributions were in Quantum Mechanics, General Relativity, Cosmology, materials science, and even some dabbling in experimental physics with a patient colleague.

In 1925 while Dirac was a senior graduate student, Heisenberg visited Cambridge and talked on his still-forming ideas of Quantum Mechanics. They struck up an odd friendship (the athletic and highly social Heisenberg and the quiet, anemic-looking Dirac) and soon after Heisenberg sent Dirac a draft of his famous paper. On one of his Sunday walks he was stunned to realize that a particular mathematical tool that Heisenberg used was identical in form to an old, formal description of "regular" classical mechanics. From this inspiration, Dirac was able to show a connection between the "regular" mechanics of Newton and its subsequent 300 years of development and the new, seemingly foreign Quantum Mechanics. The two descriptions were connected by the Planck Constant, h. Even though it's a tiny, tiny number, were it to become zero on in Dirac's mathematical repackaging, the quantum description would pass over into that old classical description. This was the first time that anyone had succeeded in making that connection. So all he did in his first professional publication was fix the apparent disconnect between the Quantum World and the previous centuries of physics. This shocked the whole of the Revolutionary European Physics Crew of those struggling with the new subject.

In fact, by the time Dirac received his Ph.D. in 1926 (at the age of 24) he published a remarkable 11 papers in which he also showed that the Schroedinger and Heisenberg descriptions of Quantum Mechanics—so different on their face—were mathematically identical. As a student, he tied everything up in to a single picture!

From his degree he took what was becoming the standard trip around Europe: he spent time working in Neils Bohr's institute in Copenhagen, establishing a life-long relationship with the revered Quantum Patriarch; a stint at Gottingen Germany where he worked with Max Born, and met the unusual Robert Oppenheimer;[2] and then to Leiden where he worked with the tragic Paul Ehrenfest.[3]

Cambridge University worked hard to bring him back to the fold and in 1927 he returned as a Fellow and began his career as a teacher and researcher. It was in his capacity as an instructor of Quantum Mechanics in 1930 that he wrote his famous *The Principles of Quantum Mechanics* which is on every physicist's shelf and is as readable today as then. It's impossible to understate the importance of that book. It set the intellectual stage for all of us as the first clear textbook on Quantum Mechanics on which most subsequent texts were based.[4] Every word in that book seems absolutely necessary and no extra words are used—it's quite a pretty piece of scientific literature. Rather an unusual legacy for a 28 year old. (The photograph at the beginning of this chapter is of Dirac taken at just about this time.)

But his most legendary contributions came before 1930: in particular the problems he tackled between 1926–28. He fixed an apparent incompatibility between Quantum Mechanics and Special Relativity, but right before that, he put photons and electrons on the same mathematical footing. Not bad for someone still in his 20's.

Paul Dirac passed away in Tallahassee and is buried there. He was unabashedly content with his new U.S. life, and even blossomed socially in Florida with many friends and the more relaxed atmosphere of the American university campus.

17.3 The Dirac Equation

You know that when someone has an equation named after them...well, there's a story behind that. We saw in the last chapter that in order to make sense of the spectra of even the simplest atoms, something else was required of Schroedinger's original wavefunction—in Wolfgang Pauli's hands it seemed to be made of two pieces according to the ad-hoc idea of spin—the regular wavefunction with the uncalled-for spin addition sort of duct-taped onto it. In this way, the wavefunction for an electron would be represented as

[2] Robert Oppenheimer was a theoretical physicist who was educated in Cambridge and worked for Max Born where he was on the front lines in applying Quantum Mechanics to problems in atomic, nuclear, and even astrophysical problems. He was eccentric and a ferocious worker, often going without food when engaged on a problem and sometimes enthusiastically taking over seminars to the point of irritation by other attendees. He came back to the U.S. and joined the faculty at Berkeley where he essentially founded modern theoretical physics in the U.S. A passionate advocate for social change, he had collaborators and friends that got him into terrible trouble in the 1950's when he lost his security clearance in a very public and humiliating set of Congressional Hearings. This was a big deal since it was Oppenhemier who was tapped to lead the Manhattan Project that organized an enormous collection of physicists, chemists, and engineers to build the first nuclear bombs that ended World War II in the Pacific. The FBI never trusted him and by the time of his triumph as the leader of the project that ended the war, they had much ammunition to use against him—and they did. A colleague noted sarcastically that he he been an Englishman, he would have been knighted for his contributions. But in the McCarthy era in the U.S. he was branded a traitor. He died at the age of 62 from lung cancer.

[3] Ehrenfest was a Dutch theoretical physicist who tragically killed himself when in a depression over his self-perceived inability to keep up with the pace of Quantum Mechanics. Paul was one of the last to see him alive and berated himself for not recognizing that Ehrenfest was troubled to that degree.

[4] In that sense, we all learned Quantum Mechanics from Paul.

Two Component Wavefunction

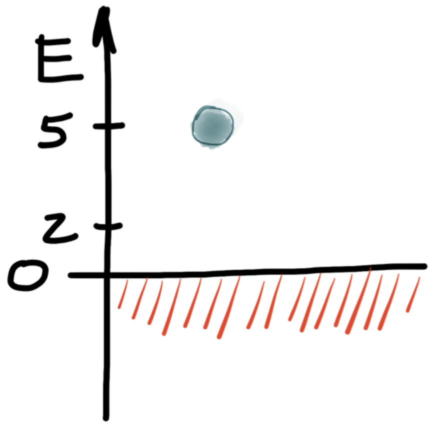


Figure 17.2: The energy of a classical particle can be anything down to zero.

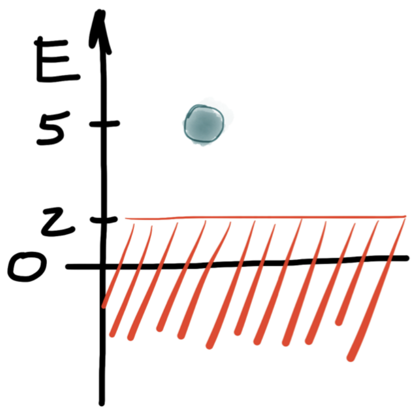


Figure 17.3: The relativistic energy of a particle is made up of the kinetic part (here $E = 5$ and the energy due to mass, here $E = mc^2 = 2$. Energies cannot be less than that due to mass.

either $\psi_\uparrow$ or $\psi_\downarrow$ which represent the spin up (+1/2) or spin down (–1/2) parts. We say that the state of an electron is then represented by a "two component" wavefunction, a double- valued quantity.

Along with the idea of spin came the the Pauli Exclusion Principle which was just indefensibly asserted: No two electrons can occupy the same "state." It was bold and it worked. Everyone was puzzled.

Completely independent of the lack of formal reason for the messy spin solution, Quantum Mechanics also suffered from a more serious embarrassment. The Special Theory of Relativity was by the 1920s essentially undisputed. Yet when one mixed Relativity with Quantum Mechanics of Schroedinger, absurdity was the result. Electromagnetism had been pliable enough to accommodate Relativity but there was no such nicety when Relativity was mixed in with Quantum Mechanics.

There were basically two reasons for this. First, in Schroedinger's picture a wavefunction seemed to be dependent on one's rest frame and that can't work. The chemistry of an atom can't be different for one observer over another. This was a kind of practical problem.

But the second reason is more serious and went to the heart of the quantum idea. Here's how to see it. When we compare the energies of a non-relativistic particle with those of a relativistic one we can see why. The "regular" Kinetic Energy is

$$E = 1/2mv^2,$$

which we can slightly rewrite by using the definition of momentum $p = mv$ to eliminate the v in favor of p. We get:

$$E = \frac{p^2}{2m}.$$

The energy is proportional to the square of speed or momentum. As the momentum goes to zero, the energy does also.

Now take a walk on the wild side: It's obvious that I can walk slower and slower and slower...all the way to a energy of zero in classical physics. Once I reach zero speed, I stop. My kinetic energy is spent and there's certainly no kind of walking that I can do to nudge me into a negative kinetic energy! So far, so good. We can see this in Fig. 17.2 where the energy of a particle in arbitrary units is shown and the disallowed energies are hatched in red. For a classical particle, the energy can go all the way to 0, but not below. Here, "energy" means kinetic energy.

Relativity mixed with Quantum Mechanics messed with this obvious sounding idea of zero as the "lowest you can go."

$$E^2 = p^2c^2 + m^2c^4 \tag{17.1}$$

This is a lot different from the classical kinetic energy in two subtle ways. First, as the momentum goes to zero, the energy settles in to just the rest mass energy, a finite value. As long as a classical object has a mass it will always have a finite amount of energy. We can see this in the cartoon of Fig. 17.3 where the energy now is both kinetic and due to mass and while the particle can slow to zero kinetic energy, there's still always the mass energy remaining. But in Schroedinger's Quantum Mechanics it's a different story.

Remember quantum behavior is unusual since it's *not* continuous. It's jerky. An electron can go slower and slower, but there is nothing to prevent it from quantum-jumping anywhere. And when Relativity was mixed in and negative energy states seemed a consequence, then that was exactly what Schroedinger's Equation predicted: an electron could change its energy from something finite and positive, to quantum-jumping right past zero, all the way to a negative value! Figure 17.4 shows that: a particle with energy of 5 units jumping past 0 into negative oblivion!

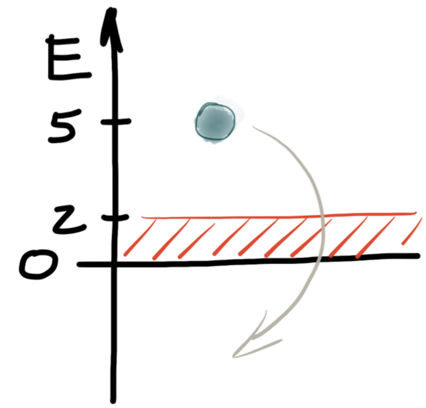


Figure 17.4: In Quantum Mechanics there is nothing to prevent a positive energy electron jumping into a negative energy state, especially one predicted by Relativity!

We can see this from what we learned for the relativistic form of energy:

That's the second way that Relativity messes with Quantum Mechanics. In Schroedinger's equation, what matters is the energy itself, but the relativistic form is a *square*, E^2. So because of that obnoxious "2," the energy by itself has to be:

$$E = \pm\sqrt{p^2c^2 + m^2c^4}.$$

Since both the positive and the negative roots when squared give the primary equation above, both are solutions. So as in Fig. 17.4 a particle with positive energy can quantum-jump to any negative energy!

Forgetting units for a minute: If the relativistic energy squared of an electron were 25, then both +5 and –5 would be solutions. If the rest mass energy of the electron were 2 in these units, then the kinetic energy of the positive electron would be 3 and the kinetic energy of the negative electron would be –3. Should an electron quantum-jump from total energy +5 to –5, then presumably it would be accompanied by a photon of energy 6. Nothing like that had been observed.

It's these two-values of the square root that permit a quantum particle to jump to a negative value for the − sign as well as the more sane positive value of $+E$. If such a thing happened in a classical theory, we'd just throw out the negative solution since it's not reachable. But we have to keep them both in a Relativistic description since quantum-jumping particle could get there!

But it gets worse. When Relativity is added, you also end up with the possibility of a negative probability! *Now that's just insulting.* Negative probabilities don't make any sense at all! Indeed Schroedinger's first attempt at his Quantum Mechanics was just such a relativistic description, but when he ran into the negative probabilities he gave up on Relativity and ran for the hills (the Alps with his girlfriend) settling for his non-relativistic version.

You Do It 17.1. Dirac's Energy

or copy the solution

From the relativistic energy equation, show that if the total energy in Fig. 17.3 is 5 arbitrary energy-units and that if the mass energy in those units is 3, then the energy of motion of that particle is 4 units. Likewise, what happens when the p goes to zero, what total energy results?

There was something incomplete about either Quantum Mechanics or Special Relativity. It was too good to toss out entirely, but at the same time it had to be a part of the truth. Everyone knew this but nobody had any idea how to fix it.

17.3.1 Relativity and Quantum Mechanics

Paul Dirac always had his own approach to theoretical physics. He said that he "played" with equations without regard to their apparent connection to what was in vogue. To him the beauty of the mathematics was most important and we would have to say that this approach worked out pretty well for him. But normal people can't work this way. Somehow hidden in his personal scientific method was a canny physical insight—something that can't be taught.

Wait. *I've never heard of Paul Dirac.*

Glad you asked. *The most famous physicist whom you've never heard of (unless you read the last chapter...then, you're an expert). Ask any pro, and she'll tell you that next to Einstein, that Paul Dirac was probably the most imaginative and influential physicist of the 20th century. You're about to see why with antimatter in a second, and then in the next chapter, our whole notion of particles. All seeds sown by this unusual man.*

Dirac broke the Schroedinger-Relativity problem down and started over, while keeping the squared energy problem squarely in mind. After trial and error, he constructed an equation that was linear in energy (no square root), but he paid a price: instead of the single wavefunction of Schroedinger or the dual wavefunction of Pauli's, what Dirac found that he had to contend with were *four* different wavefunctions. This was good news and bad news.

Definition: Dirac Equation.
The model that Paul Dirac created that merged Quantum Mechanics and Special Relativity.

The good news was that his formulation was invariant with respect to any rest frame. Check. He no longer had the negative probability problem. Check. And he found a surprise: Two of his four wavefunctions had all of the features of Pauli's two spin wavefunction![5] So spin just popped out of the equations for free, without being introduced by hand as Pauli had done! This was quite remarkable. Quantum Mechanical Spin went from being a wild guess to an actual requirement of Relativity. Spin is a purely relativistic quantity.

[5] He found this by subjecting them to a hypothetical magnetic field interaction.

But he still had the negative energy problem—it didn't go away with his new equation. And he had to figure out what the other two wavefunctions meant. He thought pretty hard about about the consequences of these extra solutions over a couple of years, frustrating everyone who had to listen to him as he worked this out. Eventually he came up with a scenario that was controversial, to say the least.

17.3.2 The Vacuum, Spruced Up

We're about to enter a subject that's going to also plague us to the present day: Nothing. Can Emptiness happen? As I hinted when we talked about Newton's Cosmology this subject is an ancient one and most associated with Aristotle who claimed that the very idea of space was defined by objects themselves. Take away objects and there's no way to speak of the space between things, if there are no things to be between. So emptiness didn't exist for him. But we saw that Newton was one of the first to flip that idea in favor of Emptiness as a vessel into which "stuff" in the universe is added. Then, Einstein seemed to bring Emptiness back by ridding science of Newton's Absolute Space. Clearly, much ado about Nothing!

But the quantum realm complicates the subject. Before Dirac, Empty Space was either declared impossible on logical grounds (Aristotle and Einstein), or deemed necessary by definition (Newton). But

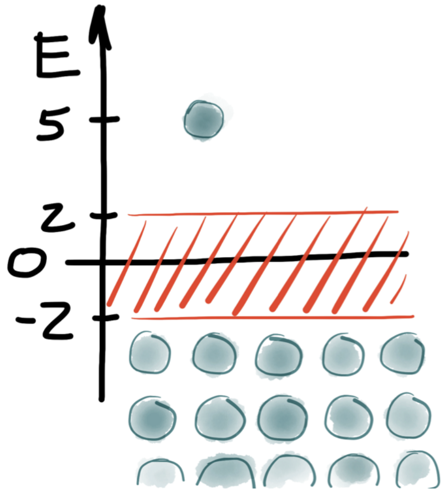


Figure 17.5: In Dirac's picture, positive energy electrons co-occupy the world with negative energy electrons, where the latter fill the vacuum, each occupying each available energy and not overlapping according to the Pauli Exclusion Principle.

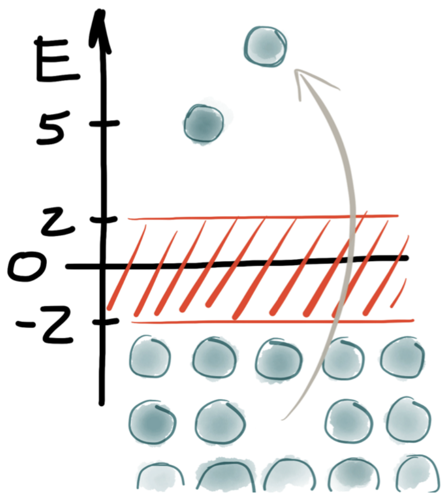


Figure 17.6: In some way, one of Dirac's negative energy electrons has been kicked free, to live a long, happy life among the positive energy world.

nobody questioned that Nothing meant anything but *Absolutely Empty*. Nobody suggested that Nothing could actually be full. That's that path that Dirac tentatively went down in his somewhat desperate attempt to understand the second pair of wavefunctions that arrived, unbidden out of his equation. The Vacuum is a strange place in the quantum world and Paul Dirac first taught us that.

How this gets mixed up in the Quantum Story is the following. The idea that the negative energies couldn't be ignored didn't seem to correspond to the fact that we're here. Everything seeks the lowest energy state, and quantum entities are no different. So if negative energies are available, why haven't all objects quantum mechanically plunged into the lowest, negative energy state...of negative infinity? Here is where his imagery went to work.

Dirac always claimed that the geometry from his mathematical training and isometric drafting from his engineering training guided his physics. He needed a mental picture and what he visualized in this most difficult of all problems was a whole new world that we're not familiar with. In this new world he needed to allow for negative energies, but somehow stop particles from jumping into them...so he took Pauli's Exclusion Principle at its word and simply asserted that all possible negative energies must already be filled with electrons. Since no two electrons could occupy any state, then if the negative energy slots were all unavailable—no positive energy electron could quantum-jump to them. So Dirac's Vacuum is empty on the one side—no particles with positive energies (empty, like a traditional Vacuum) and full on the other side—all negative energy "slots" are filled with negative energy electrons as suggested in Fig. 17.5.

But the real world is active. And he knew that something else must happen besides the empty positive energies and full negative ones. Suppose that some energy is deposited into his dual world on one of those negative energy electrons. If the photon has enough energy, it could liberate it from the confined negative energy sea to the free world of positive energy matter— where we're aware. The picture in your head should be something like that in Fig. 17.6. Think about what has to happen in order for this newly juiced-up negative electron to become "real" as a positive energy, regular electron. The energy has to overcome the negative value to get it from -4 units of energy just to get it to zero. But it's not real yet, since it doesn't have enough mass energy to qualify as a real electron, so that deposit has to have at least 2 more units. So 6 units of energy would get it to just enough to make the positive energy electron...at rest. Any more than 6? Well that just means that the electron has kinetic energy once it's liberated, here more than the 5 units of the spectator electron in Fig. 17.6.

But that's easy to picture. What's left behind where the negative energy electron was before its gift of energy? Dirac called this defect in the sea...a "hole." The characteristics of that hole are very interesting.

17.3.3 The Hole

Not quite Alice's Rabbit Hole, but Dirac's hole idea had bizarre consequences just the same. What would the characteristics be of a hole? This negative energy world, in Dirac's imagination, is all around us but just like a fish is unaware of the water its immersed in, we don't notice it. So let's take each of the primary qualities that define "electron" one at a time.

How about the about the electric charge of a hole: in its full state, the sea has an infinite number of electrons and so it's electric charge is...infinite. We'll call it Q_{sea}.

Infinite. That's just terrific.

Likewise the energy of all of the individual electrons similarly adds to infinity. Let's call it E_{sea}. What Dirac showed was that relative to the sea of the negative charge, negative energy electrons, that a hole—thought of as an absence of a negative charge, negative energy electron—had defined properties which are opposites of regular electrons.

✎

For example, when an electron is removed from the negative sea, what's left over? Well, it would be $[Q_{\text{sea}} - (-e)]$. But *relative to the "sea"* you'd remove the big background charge of that sea and the charge left over of the hole would be:

$$Q_{\text{hole}} = [Q_{\text{sea}} - (-e)] - Q_{\text{sea}} = +e$$

a positive electric charge. Likewise the energy of the hole, also *relative to the sea* would be:

$$E_{\text{hole}} = [E_{\text{sea}} - (-E_e)] - E_{\text{sea}} = +E_e$$

a positive energy.

Didn't like that? Let's count the charges in a different way. Let's suppose that instead of infinity, the total negative electric charge of our story-vacuum is -10. Let's write it as

$$-1-1-1-1-1-1-1-1-1-1 = -10.$$

Now let's take out one of our electrons:

$$-1-1-1-1-1-0-1-1-1-1 = -10+1.$$

Notice that the left side includes a "hole" where there was a negative electron and that the right side describes that as the addition of $+1$, something with a positive charge. So in the language of the previous paragraph,

$$(-10-(-1))-(-10)=+1$$

since here $Q_{sea} = -10$ rather than negative infinity. Got it? Relative to the sea of negative charge, the removal of one of them acts like a positive charge.

So this hole that's left behind would *behave* as if it were a relatively positively charged object with a positive energy.

Now this confused him, but he pressed on with an interpretation. Remember, Dirac's Equation—which is what we call it now—did an amazing thing: it accounted for spin, it reduced to the Schroedinger Equation for electron velocities small relative to c, it also accounted for a slightly different but experimentally correct atomic spectrum. It had a lot going for it. Except: what are those negative energy solutions about? They seemed to point to a positively charged particle when a photon comes along and promotes one into the positive state.

At first Dirac thought that maybe his equation had revealed a reason for the proton to exist—that it was maybe the positively charged hole.[6] But it was pointed out by Herman Weyl and Robert Oppenheimer that whatever the hole is, it had to have the same mass as an electron, so it couldn't be an explanation for the proton. Back to the drawing board.

[6] This is a new way of thinking in and of itself: that an equation might be an actual reason for a particle. It's modern reasoning, which we employ all the time and Dirac was the first to do so.

People became tired of Dirac's trying to find an interpretation for his equation: "The saddest chapter of modern physics is and remains the Dirac theory." A letter from Heisenberg to Pauli. "I find the present situation quite absurd and on that account, almost out of despair, I have taken up another field..." A letter from Heisenberg to Bohr.

Definition: Antiparticle.
An antiparticle is a particle of the same mass, but opposite electrical charge. All particles have antiparticle counterparts, but some of them are their own antiparticle (like the photon).

Definition: Positron.
The antiparticle of the electron got its own special name.

After a couple of years in 1930 Dirac took the huge leap of concluding that the hole behaved like a separate, distinct particle. A new particle. A partner of the electron, but the *anti-partner* of the electron—the "anti-electron," (his words) or as a later journal editor would later suggest, the "positron."

The Dirac Equation would then account for four wavefuctions which uncovered the full electron-family: two spins of each the electron and its anti-particle cousin, the positron. It was an audacious move. He even went so far as to predict that the proton also should have a negatively charged anti-particle cousin, an anti-proton. (That prediction came was confirmed in 1955.) Plus he imagined how it might be unlocked from the negative sea—by the addition of some sort of energy, as we pictured in Fig. 17.6.

Where would energy of that sort come from? It has to be neutral and luckily, at this time the neutron hadn't yet been discovered and so the only candidate was a photon and that's what Dirac suggested: a photon with enough energy could strike a negative energy electron, promote it to a positive energy electron and leave behind a positive energy, antiproton.

We need a nomenclature for antimatter and for reactions. You can probably sense Feynman diagrams are coming! We'll write a process with an arrow connecting what happens at the beginning (the ingredients) with what happens when the dust settles (the cake). So like a chemical reaction. We'll call all particles by their names, which will be a Latin or Greek letter and usually write an antiparticle with the particle's name and a bar over the top of it.[7] We'd write this particular transition of a photon becoming an electron and an anti electron as:

$$\gamma \to e\bar{e}. \tag{17.2}$$

Where did the photon go? For now, we'll say it "converted" into the electron- positron pair and tell you how we deal with its mysterious disappearance in the next chapter. So with lots of photons flying around the universe and presumably lots and lots of negative energy electrons in the sea, you'd maybe expect that electron-positron pairs would be popping up all over. And you'd be right.

As if on schedule, soon after it was proposed by the young Paul, the positron showed up in dramatic fashion in the hands of a young Carl.

[7] Sometimes, I'll be explicit and perhaps write e^+ to indicate the positive election. Context will rule.

As we will see, the hole idea ceased to be an acceptable explanation for antimatter in favor of a more general description. But the hole idea was found to be fruitful in another area of physics, now called Condensed Matter Physics, or Solid State Physics. These are the fields that study materials and their behavior. One of the more interesting materials is responsible for the electronic device on which you might be reading this right now: the semiconductor, which is most precisely described as a material that under the right conditions of voltage behaves at an interface as if conductors (electrons) and holes set up currents which can be switched on and off. The holes have all of the features of electrons, but move in the opposite direction in the presence of an electric field. In an important way, Dirac was the instigator of the most wide-sweeping technological advance in the history of the 20th century!

17.4 Following the Mathematics

We've just crossed a line. To this point we've been messing with common sense in Relativity and Quantum Mechanics, but leaving things like what it means to travel near the speed of light, or how to manipulate a wavefunction as technical recipes. This antimatter business, and how it came about is a different story. Now we're talking about reality. What must be the case. And what Dirac suggests about what must be the case should be testable, and it appears to make little sense. But the mathematics made us do it!

Wait. *You mean that we're going to start to believe in mathematics as if we're forced into it? Don't we have a choice to just say "no"?*

Glad you asked. *Sorry, but being pulled by the mathematics into uncomfortable interpretations of just how Nature seems to function marks the beginning of the modern approach to physics. It wasn't everyone's cup of tea. This is different from everything that came before: the ground starts to shift continually from this point. What made sense yesterday, is cast aside today. What is actually real yesterday is now uncertain today. The Quantum Heroes had to metaphorically close their eyes and walk forward, guided only by their penciled scriblings. They had to cast aside common sense and begin to imagine that our brains were not evolved to actually understand Nature in her most fundamental ways. Sure, we can make predictions and test those predictions, but we can't sensibly describe what's actually happening!*

This was hard for the pioneers. They were inventing modern physics and you can identify how different people had different reactions:

1. Some ignored any interpretation of Quantum Mechanics altogether, letting the mathematics speak and not worrying beyond that. Many (or most?) physicists work in that mode today.
2. Some worked incredibly hard to find a way to describe in words (human language!) what it all means.
3. Some concluded that the whole enterprise was at best incomplete, if not wrong! That was Einstein's position as he became older.

As troublesome as the first painful lessons were, it was only going to get worse!

To remove the suspense, let me say that most of us have learned to live with this state of affairs. We're prepared to "follow the mathematics" and live with the consequences, content with the notion that our brains and our communication skills need not to have evolved beyond simple appreciation for macroscopic objects and everyday velocities. There's no reason why what goes on in the realm of atoms should conform to our notion of common sense. It's strange and beautiful to be able to probe the inner structure of the Universe. It's fascinating that humans are able to find ways to understand things at this level. At least that's my story and I'm sticking to it! So prepare yourself for an unusual ride from this point.

17.5 Just In Time

The drama of discovery that sometimes goes like this: scientist Moe *over there* predicts something that nobody ever dreamed of before while scientist Larry, *over here* has been puzzling over just that phenomenon without knowing of Moe's idea. As if by chance, one hears of the other and both get trips to Stockholm.

That's what happened between Dirac (Moe) and a young Caltech researcher by the name of Carl Anderson (Larry)...but it's even more intriguing since right under Dirac's nose, a group within Rutherford's Cambridge (Curly?) laboratory was puzzling over the same thing without knowledge of even their neighbor's prediction! In fact many people had a surprise in their data that they missed or rather, *dis*missed.

Actually, we're protected here on Earth by the capturing of many of these particles by the magnetic field that surrounds the Earth. One of the unsolved problems for interplanetary space travel is the medical danger from these particles for astronauts who would be outside of our protective magnetic belt for months.

17.5.1 Cosmic Rays

As you read this, you're under attack. Bombarding you from above are hundreds of elementary particles going through your body in a few minutes. Every once in a while, thousands blast through you like a torrential rain of electricity. Feel it? No, you don't although the errant particle can interact with your DNA and perhaps induce some mutation. But as the bag of water that you mostly are, you're mostly unaffected by their intrusion.

"Cosmic Rays" have been a puzzle since about 1910 when they were first taken seriously. Remember how Rutherford used ionization to detect and measure the amount of charged radiations in decays?

Charged alpha, beta, and gamma rays would discharge, electrically charged-up electroscopes. It was puzzling to find that if one was patient enough that even when an electroscope was not near any obvious decaying nuclei, that they would still discharge and it was correctly determined that there must be radiation from the soil on Earth that contributed equaling things out.

The clever Jesuit Theodor Wulf tested this idea. In 1910 he climbed to the top of the Eiffel Tower with an electroscope of his own design and showed that even when you got away from the earthy stuff, the discharge continued. That something *above* the ground was responsible for much of that discharge. By 1914 crazy people like Rudolf Hess were risking their lives in balloons, going nearly 30,000 feet into the atmosphere and finding that their electrscopes' loss of charge not only continued, but was more quicker the higher they went!

It became clear that Cosmic Rays were, well, Cosmic! They come from outer space.

Since those early days studies of this large bombardment by charged particles have revealed remarkable data about them. First, by noticing that they spiraled around the Earth's puny magnetic field in a particular direction—East to West— their origins must be positive particles (we think now, mostly protons and not higher mass nuclei). Next by building detectors over large distances, the size of whole counties, it's clear that a few times an hour that the surface is blasted with huge "air showers" covering large land area with millions of particles. This led to the correct idea that whatever was hitting the upper atmosphere was causing a gigantic cascade of particle production before the Earth (and scientists' instruments) gets in the way. This was the state of the art around 1932. By now we know that the energies of these strangely energetic protons are enormous, as much as 8 orders of magnitude higher than the highest energy accelerators — many orders of magnitude higher than any conceivable man- made accelerator could ever achieve on Earth.

Their origins might be from a variety of sources, most likely shock waves from exploding stars if they're local (from within the Milky Way) or the spittle from terribly angry whole galaxies called Active Galactic Nuclei that shoot out streams of particles from their cores. In any case, much of the birth pangs of Particle Physics came from studies of Cosmic Rays in the 1930's and 1940's as clever experimenters improved on old instrumentation techniques and invented whole new ways of detecting particles. That's where the Dirac story gets interesting. In California and unbeknownst to him, across town at Rutherford's lab.

17.5.2 Anderson's First Experiment

Carl Anderson wasn't in the habit of disregarding his boss. Robert Millikan was legendary, authoritative, and pretty sure of himself. For good reason, since experiments that he'd done had contributed mightily to the burgeoning story of quantum physics—it was he who demonstrated that the electron had a fixed

Figure 17.7: Next time you're flying across country, imagine looking out your window at cruising altitude and seeing a tiny balloon with a man in full suit and tie manipulating a crude scientific instrument. The Captain had just announced that we're at 30,000 when I took this picture. This is the height that Rudolf Hess climbed in his balloon to measure Cosmic Rays.

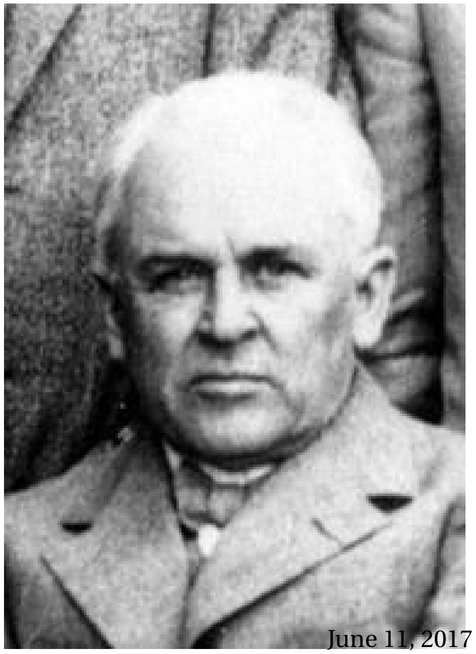


Figure 17.8: Robert Millikan in 1935.

charge and that it couldn't apparently be divided any more than that fixed quantity which we now call *e*. Never shy, he loudly proclaimed that Einstein's photon idea was wrong and then proceeded to demonstrate the opposite in pioneering measurements of the Photoelectric Effect (which hardly tempered his blistering criticism of the photon idea).

Millikan invented the term "Cosmic Rays."

He also had taken up study of Cosmic Rays at the California Institute of Technology—aka Cal Tech—which was becoming the most sophisticated center of experimental physics and observational astronomy in the U.S. He had a theory that Cosmic Rays were high energy photons—the left-over "birth-pangs" of creation— and was in a feud with his former University of Chicago colleague Arthur Compton who suggested that they were protons. Eventually, again, Millikan was found to be wrong when it was shown that the cosmic rays were bent in the Earth's feeble magnetic field. No matter. Millikan was quick to come to a conclusion and would defend it to the end, verbally steamrolling anyone in the way. He certainly wasn't much interested in backing down in measurements of his own devising, in a field he largely pioneered, especially from his former Ph.D. student, Carl.

Anderson had been a student of Millikan's at Cal Tech and then stayed as a young researcher after his doctorate and eventually spent his whole career there as a professor. He recalled that there was a three year stretch when he never saw his advisor one time, perhaps confirming the horror stories of graduate student life in some pockets of academic science (not at our institution, of course). So he became pretty self-reliant and an expert at constructing large Wilson Cloud Chambers and building scary high-current magnets in which to put them.[8] This was the preferred mode of investigating Cosmic Rays. Taking their portraits. Randomly.

[8] The whole laboratory's lights would dim when Carl would turn on his magnets.

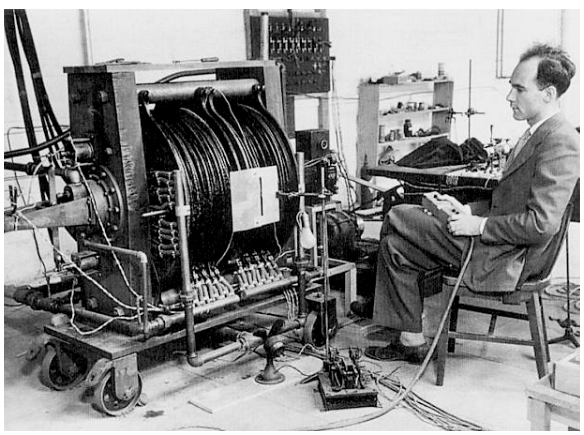


Figure 17.9: Carl Anderson working in front of his magnetized coud chamber. We all dress in wool suits in our laboratories.

Cloud Chambers

By the time that Carl Anderson was studying Cosmic Rays, the Cloud Chamber method was well-established, and very human-intensive. A person would just blindly take hundreds of photographs of the chamber, sufficiently illuminated in order to show the tracks as white dots. By analyzing the density of the dots on a track, researchers had determined that the heaviest particles—protons and nuclear fragments—left dense, heavily ionizing tracks. Electrons left much less dense tracks and were uniform, time after time. Remember, people thought that electrons and protons were all that existed, as the neutron had not yet been discovered (showing up in a cloud chamber in 1932) and so they became adept at picking out their familiar proton and electron patterns.

What Anderson did was build that very large magnet around his cloud chamber, allowing him to bend the protons one way and electrons in the opposite way. By knowing the curvature and the strength of the magnetic field, he could measure the energies of the particles and guess at the Cosmic Ray energies themselves. At least, that was the plan. He fired up his magnet with a uniform 24 kGauss (a small bar magnet has a field strength of about 100 Gauss) solenoidal field then started taking pictures. Lots of them. 1500 of them of which almost all were blank (it was a chancy thing, this picture-taking) but out that bunch a dozen or so were problematic.

You Do It 17.2. title

Draw a plate and a track of cosmic ray electron coming from the top. Assume that the magnetic field is oriented so that positively charged particles bend to the right.

or copy the solution

Right away he began to see ambiguous results. Remember, he could distinguish between a positive proton and a negative electron. What he saw appeared to be positive tracks (because of how they were bent) but with the track densities of electrons, which made no sense. But how could he tell which direction they were going? If a track was a negative electron, but came from above, it would bend one way and appear to be normal, but it could logically have been a positive electron coming from below and be abnormal.

“In the spirit of scientific conservatism we tended at first toward the former interpretation, i.e., that these particles were upward-moving, negative electrons. This led to frequent and at times somewhat heated discussions between Professor Millikan and myself in which he repeatedly pointed out that everyone knows that cosmic-ray particles travel downward, and not upward, except in extremely rare instances, and that therefore, these particles must be downward-moving protons. This point of view was very difficult to accept, however, since in nearly all cases the [thickness of the track] was too low for particles of proton mass.” Anderson, C. D., American Journal of Physics 29, no. 12 (December 1961): 825.

That the track densities of these anomalies were not proton-like and since he was not so Millikan-stubborn, he pursued it and did something clever: he inserted a plate of lead inside his chamber so that when a particle would go through it, energy would be lost and the bending when it emerged would be tighter, a smaller radius. That way he could tell which direction a track was going: if the a track came from the top, it would bend little before the plate, and then more after passing through it. If from the bottom, the opposite. Therein lay the surprise.

On August 2nd, 1932 (18 years to the day before I was born), he captured an iconic event. This was the clearest example of a particle that was obviously an electron (since the track density was electron-like-lite), obviously coming from below (since the curvature greatly became tighter above the plate of Lead), obviously not a proton (since they had learned that a low energy proton would barely have struggled to get out of the Lead at all), and obviously positive (because of the direction that it bent). Staring at him directly was the first definitive evidence of an anti-electron, which he called the "positron" in the paper that he quickly wrote and sent to press, against the wishes of his boss. By September, the rest of the world would know that antimatter existed.

Well, not quite.

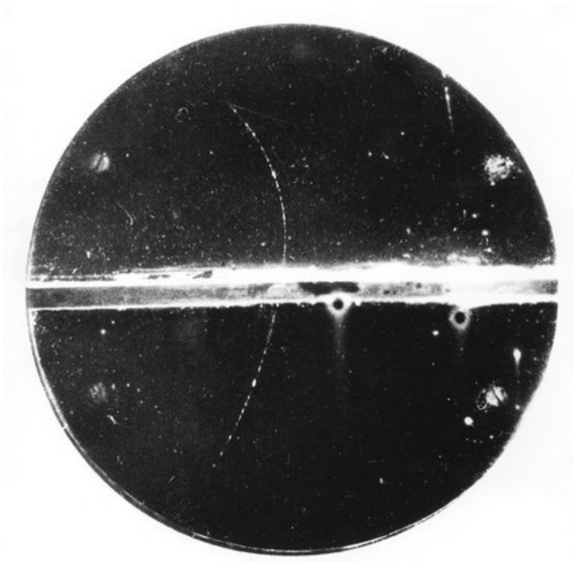


Figure 17.10: The famous photograph capturing the first acknowledged antimatter particle ever discovered. The bend of the particle is according to a positive electrical charge and the loss of energy (tighter spiral) shows that it comes up from the bottom.

Chapter 18

Quantum Mechanics, Grown Up

Paul Dirac's First Big Score


Paul Dirac as a young man.

Paul Dirac, 1902-1984

"In science one tries to tell people, in such a way as to be understood by everyone, something that no one ever knew before. But in the case of poetry, it's the exact opposite!" Paul Dirac as quoted in *Brighter Than a Thousand Suns : A Personal History of the Atomic Scientists (1958)* Robert Jungk

When Paul Dirac went to Stockholm to accept his Nobel Prize in 1933 he did all of the standard things—the parties, the banquet, and of course his address to the Royal Court and guests and family. While families of Nobel Laureates always attend the ceremony to see their spouse, parent, or child inducted into history, Paul's father didn't attend. He wasn't invited.

Quantum Mechanics was difficult enough. Relativity was tough too, but somehow a little more accessible, right? Putting them together—which everyone knew had to be done, but nobody could figure out—proved to be the opening of a floodgate that let in all manner of odd realizations of just how Nature works at the deepest level.

18.1 Goals of this chapter:

- Understand:
  - How to calculate distance, time, and speed for uniform and constantly accelerated, linear motion
  - That falling objects all have the same acceleration near the Earth.
  - How to graph simple motion parameters
  - How to read graphs of realistic motion parameters
- Appreciate:
  - The algebraic narratives in the development of the formulas
  - The shape of the trajectory of a projectile
- Be familiar with:
  - Ideas of motion before Galileo
  - Galileo's life
  - Galileo's experiments with motion

key concepts

18.2 A Little Bit of...haven't decided

In order to dig into modern theories of EPP we want to be able to describe how elementary particles interact with one another and how different theories specify those interactions. In practice we'll be able to concoct a straightforward method[1] to do this but under the hood, these tricks are based on one of the more complicated bits of mathematical physics that's ever been devised. This *tour de force* has a fancy name: "Relativistic Quantum Field Theory" (RQFT) and while a mouthful, it's a catchy phrase and sure to impress at parties. Since you're about to understand what it means, feel free to use it. If you value your privacy.

[1] Easy to say, right?

Definition: RQFT.
Relativistic Quantum Field Theory is the model of how elementary particles behave both quantum mechanically and relativistically.

Wait. *Antimatter was pretty hard to grasp. Can't we come to earth?*

Glad you asked. *Well, you'd better fasten your seatbelt. I've worried out loud a little about "what an equation means" but here it's going to be a critical consideration. You'll see that while we're really talented at turning the crank, we're a little hamstrung when it comes to thinking really deeply about what's actually attached to the crank. Like any group of people, there are physicists who just want the answers and don't worry about the conceptual stuff. And there are others who stay awake at night worrying about exactly what the symbols in the equations all mean!*

What I'm going to describe is likely to frustrate you at first. That's partly because it's a tough subject taught to second year graduate physics students and the details are pretty technical and we just can't go there. It's also a subject which we understand almost solely through mathematics and not words or even reliable mental images. But of course I'm going to use words and pictures, so you'll have to accept metaphor and analogy as substitute for six years of university mathematics.

That's okay with you, right?

Perhaps surprisingly while RQFT is a sophisticated model and well-defined set of rules, it's not entirely understood in each and every solemn detail and some people devote themselves to tying up a few formal, mathematical lose ends.[2]So do we wait around until it's perfect? No: welcome to practical science at its best. The RQFT recipes make predictions which are incredibly precise and experiments confirm those predictions with exquisite accuracy. And that's good enough for most of us and so we don't worry too much about the niceties. All of the progress in most areas of Quantum Mechanics in the last 50 years is due to this theory.[3] So stick with me and I'll highlight the concepts and few formulae that we'll need to make progress. I promise that by using a bit of it you'll grow to tolerate, if not enjoy RQFT. Like us!

[2] This means they try to be sure that every single step be logically consistent with every other and that there are no unwarranted mathematical assumptions.

[3] When it's all tied up in a neat bow, we'll take a couple of hours off and have a party.

We're going to walk a mental tightrope together without a net here. You and I both have this natural desire to put a mental (here, read "mathematical") concept into words, and when we can't do it, it's frustrating. The wavefunction falls into that category and now we're going to go even further into the realm of abstraction. But Quantum Mechanics is different and I feel your pain: living with what a successful equation "means" is a burden that takes some getting used to.

Even in Relativity when we were faced with some odd mechanical circumstances like length contraction or time dilation we didn't have to wonder what a meter stick was or struggle with the "concept" of a clock. So the meaning of the objects affected by the theory was not at issue and while the behavior of these objects was unusual, we had a feel for it.

Sometimes I'll have to say something like, "If we take literally what the equation says, then we have to assume that X is the case." Then I'll describe X which will cause you to shake your head and say, "No, the World can't be like that!" Trust me: that's the best we can do. We have to take the position that if RQFT works so well and yet is a little unnerving when its mathematics is deconstructed into words, well, who's problem is that? Nature continues to function just fine and I suspect she's not staying up nights worrying about our inability to fathom to our satisfaction what She's doing. Again, RQFT works to many decimal places of prediction. Obviously there's something very right about it.

As Feynman put it: "It is not a question of whether a theory is philosophically delightful, or easy to understand, or perfectly reasonable from the point of view of common sense. The theory of quantum electrodynamics describes nature as absurd from the point of view of commonsense. And it agrees fully with experiment. So I hope you can accept nature as she is—absurd." Oerter, Robert (2006-09-26). The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics (p. 131). Penguin Group. Kindle Edition.

With that apologetic introduction, let's take stock of what we know up to this point in our story, which corresponds to the situation in about 1926 when—guess who—Dirac wrote a seminal paper that changed everything. What we knew by then was that electrons were originally found to behave like particles and then grudgingly, were described by deBroglie and then Schroedinger as if they were also waves.

- So electrons: → first interpreted as particles & then also as waves.
- So photons, → first interpreted as waves & then also as particles.

Nature complied:

- particle like behavior of light was confirmed in 1923 in the Compton Experiment: ✓.
- wave-like behavior of electrons was confirmed in 1927 in the Davisson-Germer Experiment: ✓.

So a shaky stand-off was under way when Dirac was in graduate school: particles and waves seem to be opposite concepts, but yet shared by all forms of energy and matter.

A small difficulty: the theories that described both electrons and photons were different! Schroedinger's theory was of non-relativistic electron wavefunctions (and Dirac fixed that in 1928 as we saw). Electromagnetism is clearly a relativistic theory—after all photons move at the speed of light. But the propagation of EM waves and even their interaction with quantum mechanical electrons was still described by Maxwell's theory—it wasn't really quantum mechanical at all! So something was wrong and Paul Dirac quietly set about to fix that in 1926, a year before he wrote down his relativistic electron theory.

I'll try to transcribe his symbolic mathematical narrative that describes this into a story of words and pictures. Again, please don't blame the messenger!

18.3 Uncertainty Principle, Unplugged

You might not have thought of this, but "identity" is a philosophical problem that Dirac inadvertently solved. Of course, he wasn't trying to solve it, he was working on something prosaic, like how atoms absorb and emit light. But let me lead with this because it's so simple a problem that I'll bet you never thought about it before and you'd be in good company. Since Plato worried about it, not much thought had gone into it.

Look around you: what among the things that you experience in everyday life are identical? Sofas are all similar in many respects, but they're really all different in particular features. The big pieces in our living rooms share "sofa-ness" with other such furniture but even if they are the same brand, small differences remain among them rendering them imperfect, a mere shadow of the True Sofa. From those billiard ball examples I'm always using, even the "8 balls" from each set are very similar, but in tiny ways each would be different from one another—slightly different thickness of paint at the micron level, for example. Even mass produced identical objects are not *really* identical.

As I hinted, this is an old subject. Plato surmised that the sofas that we perceive in our lowly human lives aren't actually real, but that only the pure, Idea or Form of a Sofa is the actual, true one and that the apparent sofas that we know then "participate" in the Form of the Sofa. You see, reality for Plato consisted only things that are absolutely true and that was his realm of the Forms, those perfect ideas which are the source of the things we actually perceive. That's why he despised art . It's an imitation of something we see, which itself is an imitation of the Real Form of that thing.

But an electron? That's a different kind of thing. An electron in an iron atom in an arbitrary hemoglobin molecule in my blood stream is absolutely identical to an electron in the atomic hydrogen in the upper atmosphere of a planet in the galaxy Andromeda. Or exactly identical to an electron created in the Big Bang. *Absolutely identical* is a feature that only elementary particles can possess and the ideas seeded by Dirac led to an understanding of how this might be the case and how to write a theory of all of particles

In Plato's language? An electron in our lives is actually its own Platonic Form!

based on that idea. He called the theory that he first published in 1927 **Quantum Electrodynamics**, a quantum theory of Maxwell's and Einstein's electromagnetism. It required some unusual ideas.[4]

[4] These will be some of those "X" things I mentioned before.

Let's think again about the sort of atomic excitation problem from your high school chemistry class, or our description of it in Bohr's and then Schroedinger's Quantum Mechanics. An electron in some state in an atom is exposed to an electromagnetic wave, absorbs a single photon from that wave, and is excited into a higher orbital in the atom. If you wait a bit, that electron will jump down from that excited state and in the process release energy in the form of another photon whose energy is exactly the difference between those two orbitals.

Fine. What actually happened to the photon that was absorbed? Does it wrap itself around the electron? Do they join hands and swing together in that higher state? Does the electron actually get fat by somehow eating the photon? All silly sounding ideas, but something has to happen to it!

The fate of the photon fell out of Dirac's careful quantum mechanical analysis of the problem. Prior to his idea, the quantum mechanical description of atomic transitions treated the photon as if it were a classical, Maxwell wave while treating the atom as a quantum mechanical thing. Shouldn't there be a single Quantum Mechanics that describes the whole system, photon and atom?

Dirac divided the system into three pieces which he wrote explicitly as an equation of the form:

$$\text{Whole System} = \text{Atom alone} + \text{EM field alone} + \text{coupling between the Atom \& EM field.}$$

That last piece is a delicate combination of the relevant pieces of the two systems merged together — the "interaction piece" we call it —with Quantum Mechanics. Following his uncanny instincts and some inspired mathematics he found a single set of equations which made the atom and the field resonate together only if the photons are *individually counted*. He needed a tallying symbolism that started from zero and counted up or down in single photon units. Literally.

"The light-quantum has the **peculiarity that it apparently ceases to exist** when it is in one of its stationary states, namely, the zero state, in which its momentum, and therefore also its energy, are zero. When a light-quantum is absorbed it can be considered to jump into this zero state, and when one is emitted it can be considered to jump from the zero state to one in which it is physically in evidence, **so that it appears to have been created**. Since there is no limit to the number of light-quanta that may be created in this way, we must suppose that there are an infinite number of light-quanta in the zero state..." P. A. M. Dirac, Proc. R. Soc. Lond. A 1927 114, 243-265 His remarkable Photon-quantum paper. The emphases are mine.

Supposed he wanted to describe an electromagnetic wave of a particular frequency with 1000 photons. He wrote $|1,000>$ If one of those photons was absorbed by a nearby atom, he'd then write the wave as $|999>$. The mathematical operations that took one to the other are related to the actual physics of a photon-electron collision, but it would only work if the photons could be individually counted and in this case, decreased by exactly one. That is, the interaction piece above would do something like this:

$$|999> = \mathscr{H}_{int}|1,000> .$$

Here the term $\mathscr{H}_{int}$ is where the mathematical action is. It's a fancy mathematical "operator" that contains the atomic physics.

With these new tools he could explain all interactions of an atom de-exciting from an excited state, an atom exciting to a high state in the presence of a radiation field, to even a new description of a free electron interacting with an electromagnetic field. Notice that this is an inherently particle-like description since we can count photons here, one by one.

Here is the idea and the language: When a photon is emitted from an excited atom, the EM field count goes up by 1 and we say that a photon is **created** in the EM field. If a photon is absorbed by an atom the field count goes down to 1 less and we say that a photon was **annihilated** from the EM field, like in our 1,000 photon example above. In these counting symbols, for the absorption mechanism we have $|1>\rightarrow|0>$.

Wait. *What in the world is: $|0>$???*

Glad you asked. *Oh, it's nothing. In symbols.*

This "zero" part is interesting. Dirac's "picture" was that a perfectly acceptable state for a photon to jump into to is...a "zero state"—a state of there being no photons, the quantum jump to oblivion. Here's another one of those "X" times. The mathematics requires that there be a state of photons in which there are no photons. Say that again:

I was just at a dinner with friends in France outside of the CERN laboratory. Of the eight of us, only two were American. If there's one thing that seems to separate Americans from Europeans, it's ice. Our way of speaking was pointed out to me by a particularly perceptive physicists: she noted that when you order a soda in the US you ask for "ice" or "no ice." Like "no ice" is an actual thing. "I'll have some 'no ice,' please." They all thought that it should be one word: noice. A state of ice, in which there is...no ice.

No particles is a perfectly acceptable state of a particle. *Key Observation 19*

So the photon in an atomic excitation is not absorption into the atom or the transitioning atomic electron—in this formalism, it disappears. Into the Photon-Vacuum state. Let's go to the carnival.

18.3.1 Whac-A-Mole Quantum Mechanics

You've all done it. So have I. Every carnival has the stress-relieving device in which you beat a toy mole on the head with a mallet. Every time you whack it, it disappears...and somewhere else up pops an identical mole which stays there until you whack it. The process goes on and on and there seems to be this unseen, boundless source of identical moles. Before you paid your token there weren't any visible moles—a Mole-Vacuum state. Once you begin, the Mole-Vacuum spontaneously creates a mole until you annihilate it and this goes on over and over.

Back to Dirac's Photons...When one is absorbed, where does it go to? And where does it come from when one is emitted? RQFT demands that there is this invisible, boundless source of photons which we say resides in a field...I'll call it briefly the "Ur-Electromagnetic Field." [5]

We'll use the word "field" (lower case) as that of Maxwell, but this Field is an entirely different...shall I say, animal—a primordial *Field* with a capital "F." The Maxwell field was a continuous disturbance in Electricity and Magnetism which has a particular value and direction at every point in space. The thing

[5] Ur was the ancient city in Sumaria that was presumed to be the source of all of Abrahamic tradition. The Source! Also an Ur-text is an original text or language from which other languages originate. "Ur" is an Old German word meaning "primitive" which is a good analog here to what we'll want to think of as the primitive source of all matter.

that oscillates and makes the electrons in the radio antenna in your car oscillate along. Our more modern Ur-EM Field is the source of all photons which has the ability to oscillate at every point in space and each represents the potential to become a quantum of the EM Field—we say an "excitation" of the Field. I'll drop the silly "Ur" but I want you to think of this stuff as a primordial substance (sort of!) of fundamental importance in our world: the EM Field.

We have to think of it as *everywhere* and *everytime*. When a photon is created, it's squirted out of the EM Field and becomes real. When a photon is annihilated, it's sucked into the EM Field and disappears. This goes on all over the Universe, all the time, for all time. It'll happen tomorrow, a lot.

I hope you can see that this idea of photons appearing and disappearing is a little like how Dirac's positive electrons disappear and appear along with his holes—the theory that he came up with a year after his photon Field work. But while the photon story is the modern interpretation of Dirac's atom-idea, it's also the modern version of his hole idea! His holes were just his first try to attach some meaning to his mathematics.

The modern way to interpret Dirac's hole idea is that each electron is the excitation of an Ur-Electron Field. Now we've got two such Fields, simultaneously comprising our Vacuum.[6]

Figure 18.1: mole

[6] If a copy-cat vendor created a Whac-A-Racoon game, and merged the two, sometimes you'd see a Raccoon and sometimes a Mole...there would be a Raccoon-Vacuum as well as a Mole-Vacuum from which only Moles would be produced from the Mole Vacuum, and likewise for the Raccoon-Vacuum. They're separate and when they spit out their particular varmint, you see them and can count them...and can interact with them.

18.3.2 What Particles Are

In fact our modern interpretation is that all elementary particles are excitations of their Fields: an Ur-Field for every one. Were there a state in which there are no excitations among the collection of all of these Fields, well, that would be our Vacuum. It would be *empty* in the traditional sense that if there are no excitations of any of the Fields, then there are no particles of any kind. But it would also be *full* in the sense that these Ur-Fields are there all the time and they contain energy and at every point in space, the **ability** to create a quantum appropriate to their particular field. This is going to be huge at the beginning of the 21st century in both particle physics and in cosmology. You wait.

This description of photons (excitations of the Ur-EM Field) and electrons (excitations of the Ur-Electron Field) puts all particles on the same footing. The mathematical tools we use to manipulate them are the same. After Dirac, the task from the 1930's and into the early 1950's was to formalize this idea and rid it of some mathematical embarrassments into the a very sophisticated, internally coherent description of RQFT. The rules of how Nature Works. Why Relativity figures so prominently in the title we'll see in a bit.

"Wheeler said, "Feynman, I know why all the electrons have the same charge and the same mass." "Why?" Feynman asked. "Because they are all the same electron!" replied Wheeler." Oerter, Robert (2006-09-26). The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics (p. 103). Penguin Group. Kindle Edition.

18.3.3 Back to Plato

But circle back to how I started this discussion: the identical nature of all electrons falls right out of this idea! Because there is only one Ur-Electron Field, every electron is identical to every other since *each is an excitation of the same primordial Field*. You have to fit into your head—where believe me it's crowded now—the idea that these particle fields exist everywhere and everytime and when we collide particles together or when Nature causes natural reactions and decays, it's the tickling of the Fields by the particles and their propensity to do so that account for all of physics.

RQFT is now entrenched: Elementary particles are individually created or annihilated when their Fields are disturbed. Where they go to and where they come from is this increasingly strange Vacuum.

Before we start drawing the pictures that describe all of this formalism, let's see how Quantum Mechanics and Relativity force our Fields into a strange, jittery dance and then see how some clever people devised an experiment to confirm this whole thing.

18.3.4 Pop

32 degrees Fahrenheit is a special temperature. Above that value Nature behaves one way, and below, another. There are many parameters like that in our world that delineate some qualitative boundary — some are temperatures, some are weights or forces or lengths. In Field Theory there is a special length that separates Dr Jekyll and Mr Hyde-like personalities of Nature. This cross-over point can be stated as a particular value of length (or time, Dr Jekyll) in which the Heisenberg Uncertainty Principle (the quantum side, Mr Hyde) and $E = mc^2$ (the relativity side) reveal a strange truth about our Vacuum. Truly, Nothing turns out to be a wild place and Dirac let it loose.

Let's use the Uncertainty Principle as our guide and see where it leads us. Remember it's

$$\Delta x \Delta p \sim h$$

where the ~ symbol means "almost."[7] Whatever Δx and Δp are, their product has to equal h. If our distance uncertainty is small—we're peering at space very finely and the momentum uncertainty must be large in order that their product is always h. In that situation our ability to know about speed or momentum is greatly decreased. Likewise, if we're casual about our investigation of space so that Δx is relatively large, then our ability to know the speed or momentum of something has sharpened up.

Remember Compton Scattering? This was the reaction of photons with free electrons in which the momentum and energy calculation was done by presuming that the photon behaved like a solid, tiny ball. When I went through that calculation in an off-hand way, I highlighted a particular wavelength,

Symbolically it's

$$e\gamma \to e'\gamma'$$

where the primes on the outgoing particles serve to remind us that the outgoing energies and momenta (wavelength, for the photon) are changed in the scattering.

[7] There's actually a factor of 4π involved, but that won't change our argument.

By the way Compton Scattering was re-calculated using the Dirac Equation with Dirac's 1927 counting-photons idea and it was found to only work *only* if both the negative energy states and the positive energy states were included in the calculation. Otherwise, the wrong answer resulted—another indication that the whole thing was right on the money.

called now the Compton Wavelength, which had the value

$$\lambda_C = \frac{h}{mc}.$$

Any object of mass m would have its own, unique Compton Wavelength. I'm going to show you that λ_C is a special length: for a bit of space greater than λ_C, conditions are one way, and for a bit of space that's less than λ_C something different happens. It's the dividing line in Quantum Mechanics which opens a door to havoc in the Vacuum.

For the electron, $\lambda_C(e) = 2.4263102175 \pm 33 \times 10^{-12}$ m. For the proton, $\lambda_C(p) = 1.3214098446 \pm 19 \times 10^{-15}$ m

Now let's imagine that we trap a single electron in a little box of side Δx (in one dimension). Where is the electron? Well, you don't know for sure, but you do know that it's within the box, and so that Δx is a representation of the electron's uncertainty in position. (This of course implies an uncertainty in momentum or velocity.)

But now let's make the box smaller and smaller to the point where it's *so tiny* that its sides are equal to $\lambda_C(e)$. So let's do that. I'll take the Uncertainty relation, replace the uncertainty in position Δx with the Compton Wavelength for an electron and then multiply the whole thing by 1...written in an equivalent but useful way, $1 = \frac{c}{c}$ in the middle...and turn the crank:

$$\begin{aligned}
\Delta x \Delta p &\sim h \\
\frac{h}{mc}\Delta p &\sim h \\
\frac{h}{mc^2}\Delta pc &\sim h \\
\Delta pc &\sim mc^2 \qquad (18.1)
\end{aligned}$$

Notice in the third line that when the two "c's" are distributed the mc becomes mc^2.. Now that's interesting. Let's look at what this last line says which comes from just manipulating the equation to solve for Δpc.

From the relativistic energy relation $E^2 = (pc)^2 + (mc^2)^2$, the pc piece is related to the motion energy of the electron. We've got this electron boxed in to a size that's so tiny...that its kinetic energy *uncertainty*—not its actual kinetic energy, but just the uncertainty of it— is as large as the rest energy of an electron.

What's not forbidden by any rule of physics, must happen.[8] What this suggests is that if we squeeze an electron into a space smaller than its Compton wavelength that shiny, new particles could be produced *purely from the Uncertainty Principle.* No direct, or actual input of energy...just the inability to be precise about energy is sufficient to have to admit to reality: the mass equivalent of that imprecision in energy. That's wild. *It's like the possibility is the reality.*

[8] Have you ever thought about physics like that before?

Not only can we create matter this way, we lose yet another formerly concrete idea. We put into the box, a single elecron, but by changing *the box* we lose the ability to be sure that there's only one electron. Into the clutches of Uncertainty goes the certainty of counting electrons. In some volumes it might be 1, but in other volumes or other circumstances it may be 3 ...or more. Try to make it 1 by putting a particle in a box that's really tiny? You might create more particles.

In fact, you don't need a particle in that box to create matter this way. Just build a little box with nothing in it that's a factor of 2 smaller than the Compton wavelength of, say electrons—and then an electron and positron could pop out of the vacuum. Unbidden, just by the tininess of space itself. It's as if you look at the horizon of the ocean, it looks flat from horizon to horizon. But if you get closer and closer to the water, it looks more and more agitated and frothy. The Vacuum is indeed frothy, except not with sea water. But with elementary particles.

Again, the Uncertainty Principle is the reason. If we accept it—and you'd better!!—then we have to accept the consequences. We used to think of the Vacuum as a state of Nature in which nothing exists and in which there is no energy. That is, we used to believe that $E_{Vacuum} = 0$, exactly. But the Uncertainty Principle says that energy cannot be precisely anything, unless viewed over an infinite time. So the energy of the Vacuum is actually fluctuating around the value of $0 \pm \Delta E$, and that rippling energy fluctuation can produce particle-antiparticle pairs when the $\Delta E \sim 2mc^2$. And does.

Wait. *Why the "2"?*

Glad you asked. *As reckless as Nature seems to be, there are some things that He's pretty particular about and one of them is that electric charges are always balanced. So any frothy particle production from the Vacuum had better add up to zero electric charge, since the Vacuum itself has zero electric charge. So what is produced in the froth? Equal numbers of particles and antiparticles, two by two.*

Gone in Quantum Mechanics is the luxury of imagining that there is some state of Nature with absolutely zero energy! So Nothing...seems to be constantly idling at a very low value perhaps, but it's not shut off. There are observable consequences to this which we'll talk about in a bit.

So how do we force the Vacuum into play? By forcing particles to come together in a volume so small that particle number becomes a variable and we find that particles then appear to change into other particles—if there's enough energy to make them real. What's actually happening—if we strictly read the mathematics and turn it into words—is that in all reactions particles go into the Vacuum where their particular Ur-Field reservoir lives and other particles come out of of their Ur-Fields. Which ones? How many? Well, that's what the laws of Nature dictate. That's where the forces of Nature come into play.

So now to summarize. What's actually in the Vacuum? The mathematics suggests that for every kind of particle there is a "fundamental" field. The Vacuum is the state of those Fields when they've not actually produced any particles. When is that? Well, never, since the Uncertainty Principle forces the spontaneous production of particle-antiparticle pairs all the time.

This is really quite profound and will become more so when we think about the implications for cosmology. But try to imagine just how beautifully active spacetime now is! Close your eyes and pretend to remove all of the furniture around you. Eliminate the room, the building...the Earth, Sun, stars...even any stray radiation from temperature. Create a space in which there are. no. things. There are, however, the Fields and Heisenberg quietly insisting that Zero is not an option, and so your empty mind-universe is suddenly populated by millions and millions of particles. It won't sit still. There is no such thing as nothing. At least in our neck of the woods.

Dirac started this way of thinking, and then couldn't control the direction it took when other brilliant people followed the mathematics and the physics to the highly confirmed theory of Relativistic Quantum Field Theory that we all know and love today. He grew to dislike some of the consequences.

Now let's learn how we'll create particle reactions for any theory in a strictly rules-based way and develop the set of tools that we'll use in the rest of our story. The development of these techniques was not a pretty sight but was a couple of decades of anguish and confusion, oddly blended with impressive prediction-confirmation successes. Such...is Science.

18.4 Feynman Diagrams for Real

Let's cut to the chase and work out the tools we'll need in order to unravel the forces and particles that feel them. The calculations in RQFT are very involved. For example when I teach the trade to second year graduate students the first full-blown calculation I do is our old friend, Compton Scattering. It has features which are illustrative of other processes and since it involves two particles "in" and two particles "out" I can develop some tools that will become useful later in their bewildering journey. It also can be quickly shown to reduce to a non-relativistic, non quantum mechanical result when proper approximations are used.[9]

In order to go from the beginning to end of the calculation of the probability of a photon scattering from an electron requires about 3 hours of hand-scribbling on the chalkboard and about 25 pages of my own handwritten notes. Every line is a mathematical step and the opportunities for mistakes are frustratingly large for the professor and amusing for the students.

[9] In fact, that non-relativistic, non-quantum mechanical result is a calculation that they all learn in undergraduate school that explains why the sky is blue. Go ask Mr. Google about Thomson Scattering and Raleigh Scattering.

Definition: mathematical.
it's silly

That's why we're all indebted to Richard Feynman. His reformulation of Quantum Mechanics had a particularly pretty abstraction into little cartoons that can each stand for the whole of the "histories" of a particle as I explained before. Of course the pictures are Feynman Diagrams which I've used in a non-standard, classical sense just to get you ready to think in terms of Spacetime. RQFT in Feynman's hands is visually instructive—you can "see" the processes as they unfold— yet they really are a quite sophisticated mathematical tool. We're not going to do the mathematics, but we'll make use of the "Feynman Rules" almost in the same way that he intended them to be used. They take into account all of the surprises that Dirac invented and reduce them to little stories.

18.4.1 The Dance

Back to the atom. Our picture of the excitation of an electron will become successively more sophisticated with the Vacuum actually upsetting the simplest explanation of atomic spectra. But let's stay with the simple idea and keep track of what happens in the excitation and emission cycle: EM wave comes in and excites and electron from its ground state, it spends a bit of time in a larger orbit, and then it de-excites back to the ground state by emitting a photon. Let's do it in a really, really detailed way using Dirac's creation and annihilation explanation in the process. Stay with me.

The dance of the electron and the photon takes us through six steps. The partners approach one another, come together in a resonance, and then separate. But the Vacuum is at work between each step swapping particles in and out, like a persistent former boyfriend who keeps trying to cut in. The steps are these:

1. The electron and the photon have their separate existences: the electron in its orbit and the photon as a part of some external radiation field, like a sharply tuned laser beam.
2. They come together and are each annihilated into their respective Ur-Fields.
3. A new, excited electron is created from the Electron-Field.
4. That electron's now in the high state and executes its own solo moves, fully showing off on the high stage.
5. But the high life doesn't last long and the electron is annihilated. Aw.
6. But not to worry, because no sooner has the glittery, solo electron faded from view, but a new electron is created back in the ground state and a new photon is also created and our pair go on to dance again some day.

Each of these moves are executed according to rules that depend on the forces involved, here those of Quantum Electrodynamics.

We need to begin the use of important language, that of the Initial State and the Final State of interactions. It's straightforward, but important since what we can actually measure with our real-world apparatuses.

Let's turn our dance into a liaison. From across the room two partners spy one another and move gracefully together, twirl around executing whatever romantic Latin ballroom technique you like to imagine, and then wistfully move apart at the end. Three stages of this little romance: the Initial State, the Intermediate State, and the Final State. The Intermediate State is where the action is. Before they come together, you're unaware of whether they're going to Fox Trot, Waltz, Tango, or Boogaloo. There's lots of movement and close analysis would reveal their choice. After they're done, their Final State is one of separating and going on to whatever their next project is. We can diagram this three step[10] performance in Spacetime as shown in Fig. 18.2. You see them come in, go out, and do something in the middle. This Dancefloor Spacetime Diagram is generic. Until dancing with three partners becomes standard, this diagram will accommodate all ballroom dancing.

So the Initial State and the Final State we will think of as free dancers (particles). They're far from one another, are not yet interacting, and frankly in the model, untouched by the Vacuum or Quantum Mechanical complications[11]

So with that caveat the lines represent particles, in which the wavelike nature of the Quantum Amplitudes have been buried inside of the nice sketch: Feynman's vision is decidedly a Particle Picture. There's no denying the conceptual import of that and I'm a victim of thinking that way, and so will you too. But deeply imbedded are all of the rules of Quantum Mechanics with the dual nature of particles and waves, but we don't have to worry about that complicating, mind-hurting idea when we use Feynman's Rules. So back to our atom.

The initial state consists of an electron and a photon and the final state's a different electron and photon and what happens in between is governed by the rules of QED (and any other Laws of Nature that allow electrons and photons to interact).

The rules are the following. Figure **??** shows two lines which will be used to represent electrons (and other particles we'll meet). One thing is always true of an electron line and that is that an arrow always points in the direction in time that the electron is headed. For an electron that's somehow managed to violate the laws of RQFT and is just merrily moving along in spacetime without any interactions, we'd just drawn the line and the arrow. It's a "free" electron.

But for an electron that undergoes an interaction with another particle, we draw a dot at the point where they come together as the top line in Fig.~ 18.3. The dot signifies two things. First, it denotes the force strength of whatever model of particles is underway. Second it signals the departure (or arrival) of

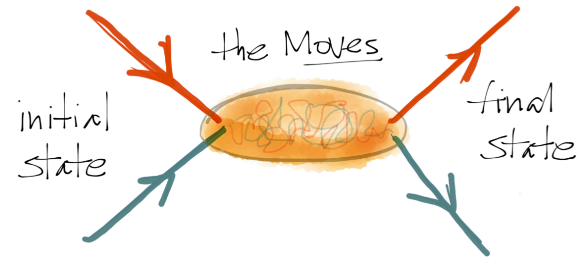


Figure 18.2: The Dancefloor Spacetime Diagram: two dancers come together—bust some moves—and separate.

[10] No pun intended!

[11] Not quite. In fact, according to Feynman's completely self-sufficient description of Quantum Mechanics—you know, he reinvented Quantum Mechanics for his Ph.D. thesis—these initial states as drawn by a single trajectory are not possible. In fact to him a particle takes *all* possible paths to go from one place to another. The dancers approach one another simultaneously by approaching normally, but also one of them goes out the door, down the hall, and back into the room at the other end, the other also goes to the moon and back...and so on. Each possible path has a probability associated with it. The normal path for people-sized objects? Of course it's the one with the overwhelmingly highest probability. But the other paths have to be included or the answer is wrong. We've already encountered this fact in the description of the two-slit experiment. Our quantum softball's amplitudes go through *both* slits and interfere on the screen. What the lines in the dancer diagram represent is the sum of all possible paths and is therefore a really complicated object.

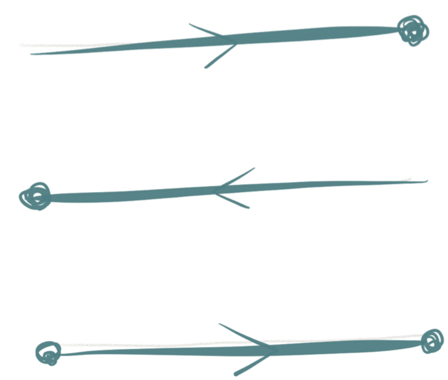


Figure 18.3: The top line represents an electron coming in from the left and interacting at the dot. The second line is an electron coming in from the right (!) and interacting at the dot and the third line represents an electorn that is "born" at the dot (some previous interaction) and lives for a while and then interacts again at the right-hand don.

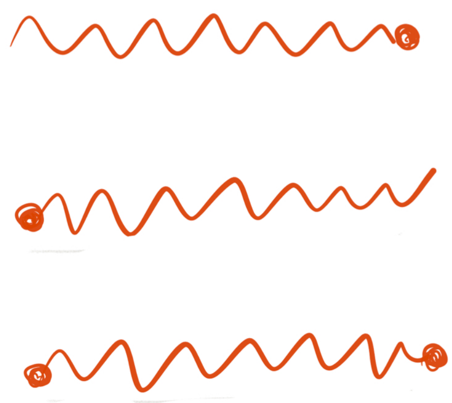


Figure 18.4: Like the electron picture in Fig. 18.3, here the same thing can be represented for photons.

that electron from the stage—it's gone to the Electron-Field and been annihilated if the dot comes after the arrow. If the dot comes before the arrow? Then, happily, a shiny new electron has been created.

Figure 18.4 similarly shows three photon graphs. The top one is just a free photon, a part of a happy EM field, maybe the sunlight, maybe an angry gamma ray from a supernova. (Notice, no arrow.) For a photon that undergoes an interaction with another particle, we do the same thing as with the electrons, namely put a dot at one end or the other (or both) to signify the force "coupling" strength and the annihilation or creation at that point of photons.

Not to be too pedantic, but what does the dance card look like for the catchy Compton Scattering Boogie? The initial state of Compton scattering is and electron and a photon. The final state? A different electron and photon (different, remember because they have different energies from the initial state versions). How does this compare with the dance of an electron and photon in the atomic absorption case? Apart from the energetics (Compton scattering would be X-rays, while the excitation of Hydrogen would be likely UV or visible wavelengths), they are the same physical process and so the same Feynman Diagram!

We literally just add the pieces together following the story: in the beginning, we had an electron and a photon and at the end, likewise. In the middle we had that solo show-off electron which connects the two pairs in the beginning and the end of the dance move. Figure **??** shows the whole thing.

Notice that the electron-ness "flows" in time with the arrows pointing the way. One of the cardinal rules of Feynman Diagrams is that electron (and sister particles we'll come to...I know, I keep saying that) lines are continuous. No breaks from beginning to end. That's not required of photons.

Remember that I've hinted at the fact that we know of four different forces in Nature, three of which are quantum mechanical and can be described in Feynman cartoon-language. In fact we can uniquely characterize every interaction by what I'll call a **Primitive Diagram** (PD). Primitive Diagrams are the starting points for any theoretical physicist who's creating a model for particles—he or she must work in terms of these pieces and then possible scatterings or decays of the particles in their candidate model putting the PDs together in order to create reactions that can be tested in an experiment. So the diagrams that are created include the new ones that he or she are trying out along with whatever "regular ones" that we already know are respected in Nature.

We now have our first Primitive Diagram shown on the largely empty table. As we move along, we'll fill in each box with a new PD for the forces that we know now, and love.

Don't try this at home, as the full implementation of the Feynman calculus is done by Professionals on a closed track. For a physicist, when a diagram is properly pieced together, it represents an algorithm for doing an important calculation: we build an equation that gets a term from each line and each dot.

Sometimes these equations can be quite involved when there are many lines and dots (interactions) and the Feynman Rules allow us to skip many, many steps and use the constructed equation to get to an actual prediction that can be checked in an experiment. For example, in the Compton Scattering marathon that I described above, 15 of those 25 pages of notes are eliminated by starting with the Feynman Rules, drawing the graph of Fig. **??** , and then building the formula and turning the crank. What one gets out of the calculation is the probability that a reaction might occur and even what the momenta of the final state particles would be. A direct prediction that can be tested. So, soup to nuts: Feynman Diagram → Probability → final state particles and their momenta.

18.4.2 Antimatter

But we know more now than just electrons. We've got to take into account its antimatter partner and we'll see that antimatter dances differently from matter. Buckle up your seatbelt because we're about to explore one of the more notorious interpretations of antimatter due to Feynman and his unusual thesis advisor.

Remember that we got into the antimatter game by imagining how Dirac's negative energy electrons might be liberated by being kicked by a photon into a positive energy electron leaving behind a positive energy positron. The process for this is

$$\gamma \to e^+ e^-.$$

We will take on the convention of referring to antimatter with a bar on top of the symbol:

$$\bar{e} \equiv e^+,$$

so that the liberation formula can be written as

$$\gamma \to \bar{e} + e$$

showing that when we do that, we dispense with the electric charge for the matter particle (electron). Given what we said above, we can draw the Feynman Diagram for this process as shown in Fig.**??**. This is a famous physical process called Pair Production, which we'll see motivates particular kinds of particle detection devices.

I've been careful to drawn the photon as going in positive time, left to right, and to label the electron and positron as likewise going in the positive time direction. Here's where it gets interesting because Feynman noticed that the energy and time variables in the equations of Dirac's always arranged themselves in the following ways. For positive energy electrons, they always appeared as products like:

$$(E)(t)$$

while for negative energy electrons they always appeared as

$$(-E)(t).$$

Being a really skilled mathematician [12] he rewrote it as

[12] And a really sneaky one.

$$(E)(-t).$$

Breathtaking, right? This is how Nobel Prizes are made! Moving a negative sign in a way you might have done in 8th grade. What he did was effectively turn a negative energy particle moving forward in time into a positive energy particle moving backwards in time.

Yes, you heard it first here. If we take Dirac seriously and take Feynman seriously—which is sometimes hard to do—we have a brand new interpretation of what antimatter is: it's regular matter moving backwards in time.

Now we don't actually detect backwards-moving electrons...electrons coming at us from the future. Rather we regularly see and manipulate positrons behaving as they should—moving forwards in time. But there's no way to ignore the possibility that antimatter is even stranger than you realized. But if we also take Einstein seriously, then if space and time are on equal footing and if I can walk as easily East as I can West (that move in space in a positive direction or a negative direction), then why can't I also do the same thing in time? Now don't get excited. Macroscopic objects are not coming at us from the future. Why? Well because we have not macroscopic antimatter objects since the annihilation of antimatter and matter happens in the blink of an eye and so macroscopic anti-objects are not possible in our Universe. But it's fun to think about it and frankly if you don't like Feynman's interpretation, you don't need to worry too much about it. But we'll use it and it's necessary in order to actually build a self-consistent RQFT.

So what about our diagram for Pair Production? If we take that positron leg which si going from the past into the future and simply reverse the time direction we turn it into an electron that's coming from the future, into the past (or here, present). The two are equivalent representations of the physical process, but it's important to remember that what we actually measure in the laboratory is always a particle or antiparticle going from the past into the future.

Manipulating Spacetime

Notice that I've drawn our QED PD in a manner that's not consistent with Relativity. Vertical lines in spacetime don't make any sense since they imply motion in space in zero time. But I've done this intentionally as I mean for the Primitive Diagrams to be puzzle pieces, except that instead of there being only one way to use a piece of a jigsaw puzzle, our PDs can be manipulated. Again, since we believe in Einstein and that space and time are on the same footing, any diagram that we draw in one configuration relative to the vertical (space) and horizontal (time) axes...we can rotate around into another process. One diagram actually can represent many different processes, and in fact we can predict physical processes by doing so.

That's how we'll use our PDs. When a new theory is on the table, I'll tell you the PD and fill in the diagram in the table. From them and the Feynman Rules (which I'll strictly enumerate in a bit), we can literally use all of the PDs as puzzle pieces that we can put together to predict reactions that we can expect to detect in the laboratory. If we find them, then great: we've just confirmed the model represented by the PDs. If we don't, then great! We've just disconfirmed a model and that's always fun for an experimenter (unless the theorist is one of your friends).

So let's complete the story by now predicting a number of QED reactions which are all embodied in that one, lowly QED PD.

18.5 The pieces

18.5.1 Spacetime Arrangements

18.5.2 Primitive Diagrams

18.6 How Do We Know?

18.6.1 The Atom Feels the Vacuum

18.6.2 The Vacuum Exerts Pressure!

18.6.3 The Whole Mess Works. Really Well.

18.7 Summary

18.7.1 Here are the important points from Newton's Laws and momentum:

- Physics quantities come in two kinds: numbers (scalars) and vectors. Vectors are double-valued and incomplete if only one of the two pieces is ignored. The two values are the magnitude of the vector and the direction which it points. Drawing arrows is a useful way to represent them in which the length of the vector can be scaled on paper to be the magnitude piece.
- The addition and subtraction of vectors will be an important tool through the whole of the book. In most cases, these manipulations can be done graphically. In some cases, it might be useful to make use of component combinations.
- Many times we will need to find the single vector that makes a combination of vectors add to zero. When this applies to forces, you have the image of them being all in balance.
- The essential quantity from this chapter for us is momentum which is defined as $\mathbf{p} = m\mathbf{v}$. When a force acts on an object for a finite period of time, it changes the momentum as seen by Newton's 2nd Law: $\Delta\mathbf{p} = \mathbf{F}\Delta t$.

18.7.2 Key Concepts

A force applied to a body will cause it to accelerate. *Key Concept 1*

Mass is a measure of an object's resistance to being accelerated. The measure of that is "inertia." *Key Concept 2*

a vector = a magnitude × a direction *Key Concept 3*

In order to make the negative of a vector, turn it around and reverse its direction. *Key Concept 4*

Force is equal to the rate of change of momentum. *Key Concept 5*

18.7.3 Key Questions

What is the nature of Mass? *Key Question 1*

ith a discussion of Collisions and Energy. Neither were very well described by Newton since he didn't have all of
ivals.

18.7.5 Key Definitions

Definition: Kinematics.
The study of motion without regard to cause.

Definition: Dynamics.
The study of forces which cause accelerations.

Definition: Momentum.
$\mathbf{p} = m\mathbf{v}$
Momentum is proportional to speed and mass. This is specifically "linear momentum."

Definition: "little g".
Near the surface of the Earth, the acceleration due to gravity is nearly constant and called:
$g = 32 \text{ ft/s}^2 = 9.8 \text{ m/s}^2$

Definition: unit vector.
A vector having length of 1 which is tied to a particular coordinate system. Usually two or three unit vectors are chosen to point along the directions of the particular coordinate system axes. Unit vectors are distinguished by the use of a little "hat" on the top of them.

Bibliography