Energy & Matter: Our Quantum World

Rob Knop

March 31, 2012

Contents

Preface **vii**

1 Units and Dimensionality **1**

1.1 SI Units 1

1.1.1 SI Prefixes 4

1.2 Arithmetic with Dimensional Quantities 5

1.3 The Unit Factor Method 5

1.4 Significant Figures 7

1.5 Dimensional Analysis 10

2 Energy **13**

2.1 The Units of Energy 13

2.2 Forms of Energy 14

2.2.1 Kinetic Energy 15

2.2.2 Potential Energy 16

3 Momentum and Angular Momentum **21**

3.1 Vectors 21

3.2 Momentum 23

3.2.1 The Units of Momentum 24

3.3 Angular momentum 25

3.3.1 The Units of Angular Momentum 27

3.3.2 The Direction of Angular Momentum 27

4 The Spin-$\frac{1}{2}$ particle 29

4.1 Particles in Quantum Mechanics 29

4.2 Measuring Electron Spin: the Stern-Gerlach Experiment 31

4.2.1 The Stern Gerlach Machine 33

4.3 Repeated Measurements of Spin 34

5 Dirac Notation 39

5.1 The State of a System 39

5.2 The Ket Vector 40

6 Amplitudes and Probabilities 45

6.1 Complex Numbers 46

6.2 Amplitudes 47

6.2.1 Calculating Probabilities from Amplitudes 48

6.3 Bra Vectors and the Inner Product 48

6.4 Normalization and Orthogonality 49

6.5 Interpreting the Inner Product 51

6.5.1 Propagating Amplitudes 52

7 The Collapse of the Wave Function 55

7.1 Summary of Rules for Manipulating Ket Vectors 56

7.1.1 Calculating Experimental Predictions 58

7.2 The “Collapse” Rule 59

7.2.1 What is a measurement? 60

7.2.2 Schrödinger’s Cat 63

7.3 Interpretations of Quantum Mechanics 64

8 Operators and Eigenvectors 67

8.1 Operators 67
8.2 Eigenstates 68
8.3 Linear Operators 69
8.4 Operators on Non-Eigenstates 70
8.5 The Hamiltonian 71

9 Vectors and Matrices 73

9.1 Column Vectors 73
9.2 Row Vectors 75
  9.2.1 The Inner Product 75
  9.2.2 Nothing is New! 76
9.3 Matrices 77
  9.3.1 Linear Operations on Matrices 77
  9.3.2 Multiplying a Matrix and a Column Vector 78
  9.3.3 The Identity Matrix 78

10 Pauli Spin Matrices 81

10.1 Spin Operators 81
10.2 Expectation Values 82
10.3 Total Angular Momentum 84

11 Noncommuting Operators and Uncertainty 87

11.1 Eigenstates and Commuting Operators 87
11.2 Non-Commuting Operators 88
11.3 Quantifying Uncertainty 90
  11.3.1 Mean and Variance 90
  11.3.2 Uncertainty in Quantum Mechanics 92
11.4 The Heisenberg Uncertainty Principle 94
  11.4.1 An Alternate Formulation 95
  11.4.2 Vacuum Energy 96

12 Multiple Particle States 99

12.1 Indistinguishable Particles . . . 99

12.2 Notating Multiple Particle States . . . 99

12.3 The Exchange Operator . . . 101

12.4 Fermions and Bosons . . . 102

12.5 The Pauli Exclusion Principle . . . 103

12.6 Entangled Particles . . . 104

13 The Schrödinger Equation 107

13.1 Where we are so far . . . 107

13.2 Stating the Equation . . . 109

13.3 Free Particles & the de Broglie Wavelength . . . 110

13.4 Quantized Energy Levels in Bound Systems . . . 112

13.5 The Simple Harmonic Oscillator . . . 113

13.6 The Hydrogen Atom . . . 114

13.7 Interpretation of the Wave Function $\psi(x)$ . . . 116

14 Atomic Orbitals 119

14.1 The Schrödinger Equation . . . 119

14.2 The Orbitals . . . 120

14.3 Visualizing Orbitals . . . 123

14.3.1 s Orbitals . . . 124

14.3.2 p Orbitals . . . 126

14.3.3 d Orbitals . . . 127

15 The Periodic Table of the Elements 129

15.1 Interacting Electrons, Energy Levels, & Filled Shells . . . 130

15.2 Filling Up Orbitals . . . 132

15.3 Reading a Periodic Table . . . 133

15.3.1 Electronic Configuration . . . 134

16 Matter **137**

16.1 The Standard Model of Particle Physics 137

16.2 Nuclei and Atoms 140

16.3 Molecules 140

16.4 Solids 141

16.5 Liquids, Gasses, and Plasmas 144

16.5.1 Quantum Gasses 145

16.6 Planets, Stars, Galaxies, and Clusters 146

16.7 Dark Matter and Dark Energy 146

Preface

This text is designed as an introduction to the theory of quantum mechanics for college students who have not had calculus, nor who have had any prior college-level courses in physics. The majority of quantum texts you can find out there either assume that the student is already familiar with calculus (and perhaps linear algebra), or are popular-level treatments of the topic that have nearly no rigorous mathematical content at all. This text tries to straddle the difference. The target audience are students who might take an algebra-based introductory physics course. No calculus is assumed, nor is any linear algebra. (The text will eventually use a very small amount of linear algebra, but it will introduce that, so the reader need not know it coming in.) However, it does give a rigorous introduction to quantum mechanics, and does not shy away from showing the mathematics of the theory where that is accessible to students with this sort of mathematical background.

Because quantum theory represents a way of looking at the world that is completely at odds with our intuition— so much so that physicists still debate how properly to interperet such things as the "measurement problem"— anybody who has had no prior exposure to quantum physics will find this material conceptually challenging. That is as it should be! The goal of a university course, particularly an introductory university course, should be to expand your mind, to make you work out parts of your brain that you might not even have known that you had. However, this does mean that if a student comes into the course uncomfortable with the concepts behind algebra at the advanced high-school level, they may be overwhelmed. If solving two systems of equations is something that you're shaky on, and if you're not comfortable with the *meaning* of an algebreic variable as a stand-in for something that we may or may not know, then you will find yourself at a disadvantage as you struggle with those concepts while also facing the new concepts of quantum physics. I want to emphasize that this text is *not* designed for physics students or math students in particular; it's designed for all liberal arts university students. I simply expect that those students will take seriously the on-paper prerequisite found at most universisties of having matered high school algebra through the grade 11 level.

The course does require students to try to deal with mathematics at an abstract level. Students tend to be much more comfortable with math when it is concrete.

I have found in all introductory physics courses that when faced with a problem requiring algebraic manipulations, students like to plug values into variables as soon as they can, and then manipulate the numbers. Those of us with more experience recognize this as a trap, for the resulting process is much more error prone and hard to follow than if one had solved the equation symbolically first. Students, however, seem to prefer to remove any abstractions as soon as possible. It's worth trying to train students to work with the algebra at an abstract level, only plugging in numbers when they absolutely can't avoid it any more (e.g. to determine a numerical result). At a higher level, this text *does* introduction the notion of operators, but doesn't always describe exactly the mechanics of those operators. Students will all be familiar with the square root; they know how it behaves on a number. The square root, of course, is an operator. If students aren't intimidated by it, it's simply because they're familiar with it, and because there's a button on their calculator that will perform the operation on a number. To explore quantum physics, this text will introduce operators as "something that does something to something", perhaps leaving the second "something" completely abstract. One need not know the detailed numerical representation of an operator in order to know that a given operator will extract the eigenvalue from one of its eigenstates. Students will often find this a difficult concept to grasp, but because it is so powerful (allowing you to solve and do things often without having to learn the details), the text does not shy away from it.

While many "modern physics" courses designed for second-year physics majors start with the Planck spectrum, and move quickly to the one-dimensional Schrödinger Equation, this text starts, after a quick background in basic physics (as no prior university level physics is assumed), with the spin-1/2 system. Of course, dealing with the Planck spectrum and with the Schrödinger equation requires calculus, so that wouldn't be an appropriate place to start for an algebra-based course. However, I believe that starting with the spin-1/2 system may well be a better way to introduce students to the concepts behind quantum mechanics. It's a very simple system, as there are only two states available. The concepts, while counter-intuitive, may be explained and understood, and the mathematics behind them may be explored with only algebra in your background. Concepts such as orthogonality and eigenstates are easily obscured amongst the slog through integrals that happens when solving the differential Scrhödinger equation as your first introduction to quantum physics. With the spin-1/2 system, however, there is much less to distract you.

This text does eventually inroduce the Scrhödinger Equation, but because students can't be assumed to know any differential equations it leaves the kinetic energy operator entirely abstract. As such, students won't be solving the Schrödinger Equation. However, they will explore some of the consequences of some of the famous known solutions, including the square well and the simple harmonic oscillator. The text then introduces the Hydrogen Atom. After having focused for so long on spin angular momentum (in the spin-1/2 system), it describes how orbital angular momentum in the Hydrogen Atom solutions to the Schrödinger Equation naturally gives rise to the sturcture of the Periodic Table of the Elements.

This text was written for the foundation course *Energy and Matter* at Quest University Canada. That course has several different versions; “Our Quantum World” is one of them.

A note about commas

You may be offended to find me placing commas and periods outside of quatation marks. I realize that I’m naughty, but as a computer programmer I can’t help but notice that standard usage is *wrong*, and leads to a pandemic misquoting. Consider the following sentence: many people have read “A Tale of Two Cities”, a novel by Charles Dickens. Standard usage would have me put the comma inside the quotation marks, but the comma is *not part* of the title. If the purpose of the quotation marks here is to set off the title from the rest of the text, then you’re misrepresenting the title by including things inside the quotation marks that aren’t part of that title. When I’m forced to, I use the standards. However, I would dearly love to see the standards change to something more logical and reasonable.

End of rant.

License

Chapter 1

Units and Dimensionality

If somebody asked me how tall I am, I might respond "1.78". But what do I mean by that? 1.78 feet? 1.78 miles? In fact, my height is 1.78 meters. Most physical measurements have *dimensionality* to them. That is, they are meaningless unless you attach a unit to them. *Dimensionality* means the type of unit. For example, inches, meters, miles, and light-years are all length units; something measured in those units have dimensionality of length. Kilograms, grams, and solar masses are all units of dimensionality mass. Measurements of different dimensionalities cannot be meaningfully compared. How many kilograms are there in a meter? The question does not even make sense.

There are some dimensionless quantities. For example, ratios are nearly always dimensionless. How many times older than my nephew am I? I am seven times older; that seven doesn't have any units on it, as it's a ratio of two ages (42 years and 6 years, respectively). For any other number you report, it's essential that you report the units of the number along with the number itself. Otherwise, you haven't completely specified what you're talking about.

1.1 SI Units

There is a "standard international" system of units. You may ask, why does anybody ever use anything other than these? SI Units are a good set of units for everyday measurements. However, they are very clumsy when dealing with the very small or the very large. When talking about atoms, or about stars, it's often convenient to use other units that are better matched to the scale of the system. What's more, some places historically use other units; for instance, the United States still uses the British Imperial system of units.

There are a finite number of dimensionalities. For purposes of this course, there are only four dimensionalities that you need to know about. They are, with their SI units, listed below:

Dimensionality	**SI Unit**
Length	m
Mass	kg
Time	s
Electric Charge	C

The four core dimensionalities are length, mass, time, and electric charge.[1] For each dimensionality there can be a lot of different units. Something of dimensionality length can be measured in *any* length unit, but cannot be measured with a (say) time unit. It doesn't make sense directly to compare quantities of different dimensionalities. So, I could measure my height in feet— 5.84 feet is my height— or in meters. While clearly the number 1.78 does not equal 5.84, 1.78 meters *does* equal 5.84 feet. A measurement with dimensionality is clearly different from a pure number; the units on the number affect what that number means.

You are already familiar with the meter, kilogram, and second. (Indeed, because of these three base units, the SI system is sometimes called the "MKS" system.) You may or may not have heard of the Coulomb before. All other units that we will deal with are derived from these base units. For instance, consider velocity. The *dimensionality* of velocity is length over time (sometimes written L/T). Any unit that corresponds to a length divided by a time is a valid velocity unit; that could be kilometers per hour, miles per hour, or furlongs per fortnight. The dimensionality of velocity is neither length nor time, but is composed of those two dimensionalities. The SI unit for velocity is meters per second, or m/s. Sometimes derived units have their own names. Below is a table of some of the more important derived units in the SI system:

Dimensionality	Unit	Definition
Force	Newtons N	$\mathrm{kg\,m\,s^{-2}}$
Energy	Joules J	$\mathrm{N\,m} = \mathrm{kg\,m^2\,s^{-2}}$
Power	Watts W	$\mathrm{J/s} = \mathrm{kg\,m^2\,s^{-3}}$

(Remember that something raised to the negative power is in the denominator. Thus, one Newton is "one kilogram times meter per second squared", or $\mathrm{kg\,m/s^2}$.) While we can say that "force" is the dimensionality of force— as in the table above— that

---

[1] In fact, in the SI system, electric current rather than electric charge is considered a core dimensionality. However, it's conceptually more simple to consider charge as the core unit, and current as a derived unit, so I'll use that in this document.

is exactly the same as saying it is a dimensionality of mass times length divided by time squared, or $M\,L/T^2$, or $M\,L\,T^{-2}$.

Some people will "just always work with SI units", and then not write down units to go with their numbers during intermediate calculations. The idea is that since you're always using the standard, the final result of any series of calculations will be in the SI unit for whatever it is that you calculated. Even though, if you are careful, you can get away with this, it would still be wise to write down the units that go with numbers every time you write down those numbers. There are two primary reasons for this. First, it makes it much clearer what you are doing and what these intermediate numbers actually are. Without that, anybody reading your calculations may have a hard time following them, and you have not communicated as effectively as you might have. Second, by keeping track of your units throughout your calculation, you provide yourself with a cross-check: does your final answer have the units that it's *supposed* to have? If it doesn't, then that's a sign that you've made a calculation mistake somewhere along the way.

For example, suppose I told you that the density of water is 1 gram per cubic centimeter, and I wanted you to tell me how much mass there is in a spherical drop of water with radius 0.2 cm. First, let's convert to SI units; if you do it right, you can figure out that 1 $\mathrm{g/cm^3}$ equals 1000 $\mathrm{kg/m^3}$. Also, 0.2 cm is equal to 0.002 m. If you say that the volume of a sphere is πr^2, you could calculate the volume from this number:

$$Vol \;=\; \pi\,(0.002)^2 \;=\; 1.257\times10^{-5}$$

Then, multiply the volume by the density to get the mass:

$$m \;=\; (1000)(1.257\times10^{-5}) \;=\; 0.013$$

Figuring that you've done everything in SI units, you should get an answer in the SI unit for mass, so you could write down and box $\boxed{m = 0.013\ \mathrm{kg}}$. However, **this answer is wrong**. Did you see where it went wrong? Let's redo the problem, this time keeping track of units:

$$\begin{aligned} m &= (Vol)(dens) \\ &= (\pi\,(0.002\,\mathrm{m})^2)\left(\frac{1\,\mathrm{kg}}{\mathrm{m}^3}\right) \\ &= (1.257\times10^{-5}\,\mathrm{m}^2)\left(\frac{1\,\mathrm{kg}}{\mathrm{m}^3}\right) \\ &= 0.0127\,\frac{\mathrm{kg}}{\mathrm{m}} \end{aligned}$$

Notice in the last step we cancelled the meter^2 in the numerator with *two* of the three meters in the denominator's meter^3. But, wait! This doesn't leave us with an answer that has dimensionality mass, it has dimensionality mass per length! Clearly we've done something wrong. In this case, the mistake was in our formula for volume. In fact, the volume of a sphere is of radius r is $\frac{4}{3}\pi r^3$. We caught this error because, by keeping track of the units as we were putting numbers into the calculation, we saw that the units didn't work out right. If you put in the right formula for volume, you discover that there are only 3.4×10^{-5} kg of water in a droplet that's 2 mm in radius.

1.1.1 SI Prefixes

Some "derived" units are just a prefix in front of a regular unit. There is a standard set of SI prefixes that can be prepended to any unit in order to make another unit of the same dimensionality but of a different size. The ones you are probably most familiar with are milli and kilo. A millimeter is 1/1000 of a meter, and a kilometer is 1000 meters. You could do the same thing with seconds; a millisecond is 0.001 seconds, and a kilosecond is 1000 seconds (about 17 minutes). Indeed, the SI mass unit, the kilogram, is itself 1000 grams. In this class, we will frequently talk about things that are much smaller, such as nanometers and microseconds. If you are in an astronomy class, you might find yourself using the mega or giga prefixes more often. The table below summarizes the prefixes.

Prefix	**Abbreviation**	**Multiplier**
terra	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}

Notice that case matters. There is very big difference between a Mm and a mm—a factor of a billion, in fact! The letter used to indicate micro is the Greek letter mu. There are a million μs in one second. The prefixes deci and centi are not used very often, and generally only with meters. While you will talk about centimeters, nobody generally talks about centigrams or centiseconds.

1.2 Arithmetic with Dimensional Quantities

When you put together numbers that have dimensions on them, you have to keep track of the units as you are doing your arithmetic. You can do algebra with numbers that have dimensions on them. **However, it is not a good idea in general to do algebra with numbers.** Solve things symbolically first, and only put in the numbers at the end. When you do this, you will have various quantities with different units.

When adding or subtracting numbers with units, you need to make sure that they have the *same* units. First of all, it doesn't make sense to add numbers with different dimensionality. One meter plus one kilogram isn't even meaningful. One meter plus one inch *is* meaningful, but it is not equal to two anything. You need to convert one of the two units to the other before adding the numbers. You could write one meter as 39 inches, and then say that one meter plus one inch is equal to 40 inches.

Multiplying and dividing units is more interesting. In this case, you treat the units just as if they were algebraic variables. If you multiply meters by meters, you get meters squared (or m^2). If you divide seconds cubed (s^3) by seconds, you get seconds squared (s^2). If you raise a quantity with units to a power, you have to remember to raise every part of that quantity's units to the same power. For example, you may be calculating the kinetic energy of a car massing 1,500 kg moving at 20 meters per second:

$$KE = \frac{1}{2}m\,v^2 = \frac{1}{2}(1500\,\mathrm{kg})\left(20\,\frac{\mathrm{m}}{\mathrm{s}}\right)^2 = \frac{1}{2}(1500\,\mathrm{kg})\left(400\,\frac{\mathrm{m}^2}{\mathrm{s}^2}\right) = 3.0\times10^5\,\frac{\mathrm{kg\,m}^2}{\mathrm{s}^2}$$

Notice that the squared on the velocity is applied to the number, to the meters, *and* to the seconds.

1.3 The Unit Factor Method

Sometimes you will need to convert one unit to another unit. The trick for doing this: **multiply by one as many times as necessary**. You can always multiply a number by 1 without changing that number. The secret is writing the number 1 in a particularly clever way. Here are some ways you can write the number 1:

$$1 = \left(\frac{60\ \mathrm{min}}{1\ \mathrm{hr}}\right)$$

$$1 = \left(\frac{2.54\ \mathrm{cm}}{1\ \mathrm{in}}\right)$$

$$1 = \left(\frac{1\mathrm{M}_\odot}{2.0\times10^{30}\ \mathrm{kg}}\right)$$

(The $M_\odot$ in the last example is the standard symbol for the mass of the Sun.)

If you have an expression in one set of units and you need them in another set of units, you just multiply by one as many times as necessary. Cancel out units that appear anywhere on *both* the top and bottom in your huge product, and you will be left with a number and another set of units. A simple example: convert the length 2.500 yards into centimeters:

$$2.500\ \mathrm{yd} = (2.5\ \mathrm{yd})\left(\frac{36\ \mathrm{in}}{1\ \mathrm{yd}}\right)\left(\frac{2.54\ \mathrm{cm}}{1\ \mathrm{in}}\right) = (2.5 \times 36 \times 2.54)\ \mathrm{cm} = 228.6\ \mathrm{cm}$$

Notice that yards (yd) appear in the numerator and the denominator, and so get canceled out, as does inches. We're left with just cm. All we did was multiply the value 2.5 yd by 1, so we didn't change it at all; 228.6 cm is another way of saying 2.500 yd.

Another example: suppose I tell you that the surface area of the Sun is 2.4×10^{19} square meters. How many square miles is that?

$$\left(2.4 \times 10^{19}\ \mathrm{m}^2\right)\left(\frac{100\ \mathrm{cm}}{1\ \mathrm{m}}\right)^2\left(\frac{1\ \mathrm{in}}{2.54\ \mathrm{cm}}\right)^2\left(\frac{1\ \mathrm{ft}}{12\ \mathrm{in}}\right)^2\left(\frac{1\ \mathrm{mi}}{5280\ \mathrm{ft}}\right)^2$$

Two things to notice about this. First, notice how all the unit factors are *squared*. That's because we started with meters squared at the beginning, which is meters times meters. If we're going to get rid of both of them, we have to divide by meters twice. The same then goes for all of the other units. Next, notice that everything except for the left-over miles squared cancel out. We're left with a bunch of numbers we can punch into our calculator (remembering to square things) to get:

$$\frac{(2.4 \times 10^{19})(100^2)}{(2.54^2)(12^2)(5280^2)}\ \mathrm{mi}^2 = 9.3 \times 10^{12}\ \mathrm{mi}^2$$

One more example. Sometimes you have more than one unit to convert. If I tell you that a car moves 60 miles per hour, how many meters per second is it going? (Notice here that instead of arduously multiplying out the conversion between meters and miles as I did in the previous example, I've looked up that there are about 1609 meters in one mile.)

$$\left(60\ \frac{\mathrm{mi}}{\mathrm{h}}\right)\left(\frac{1609\ \mathrm{m}}{1\ \mathrm{mi}}\right)\left(\frac{1\ \mathrm{h}}{60\ \mathrm{min}}\right)\left(\frac{1\ \mathrm{min}}{60\ \mathrm{s}}\right) = 27\ \frac{\mathrm{m}}{\mathrm{s}}$$

Note that since hours was originally in the *denominator*, we had to make sure to put it in the numerator in a later unit factor to make it go away (since we didn't want any hours in our final answer).

With this simple method, you can convert any quantity from one set of units to another set of units, keeping track of all the conversions as you do so.

1.4 Significant Figures

Suppose I tell you that one stick is 1.0 meters long, and that it is 4.7 times longer than another stick. How long is the second stick? Writing the words as equations (see previous section), you might write:

$$l_1 = 4.7\ l_2$$

l_1 is what you know (1.0 meters), and l_2 is what you're looking for, so solve the equation for l_2:

$$l_2 = \frac{l_1}{4.7}$$

plug in the numbers and solve for the answer:

$$l_2 = \frac{1.0\ \mathrm{m}}{4.7} = 0.212765957447\ \mathrm{m}$$

That answer is wrong! Why? Because it is expressed with too many significant figures.

Think about the original problem. I told you a stick was 1.0 meters long. Notice that I didn't say 1.00 meters long; only 1.0 meters long. That means that I was only willing to commit to knowing the length of the stick to within a tenth of a meter. It might really be more like 1.04 meters long, or perhaps 0.98 meters long, but I've rounded to the nearest tenth of a meter. Since I only know the length of the stick to about ten percent, and since I used that number to calculate the length of the second stick, I can't know the length of the second stick to the huge precision that I quote above— even though that is the "right" number that my calculator gave me. Given that I only know that the first stick is 1.0 meters long, and it is 4.7 times the length of the second stick, *all that I can say I know* about the length of the second stick is:

$$l_2 = 0.21\ \mathrm{m}$$

By saying this, I'm implicitly saying that I don't know the length of this second stick to better than the hundredths place. . . and I don't! Implicitly, I'm saying that I know the length of the second stick to about one part in 21. That's actually a bit better than I really know it (which is just to one part in 10, or to 10%, as that's all the better I know the length of the first stick), but this is the best you can do with just significant figures. (To do better, you have to keep track of not just units, but also uncertainties on every number. Doing so is an important part of the analysis of data in physics experiments. However, propagating uncertainties is beyond the scope of this course.)

How well you know a given number you write down is the reasoning behind significant figures. The basic idea is that you shouldn't report a number to more significant figures than you know are right. The rules can sometimes seem arbitrary, but if you think about them in terms of the basic idea behind them, they can start to make sense. There are four basic rules of significant figures:

1. When multiplying or dividing numbers, the answer has as many significant digits as that member of the product or quotient that has fewer significant digits. So, if I multiply 3.14159 by 2.0, the answer is 6.3; I round the answer to two significant figures, because 2.0 (the member of the product with fewer significant figures) only has two. This rule is an expression of the *percent uncertainty* in the figures that are going into your result. If you only know a number to within (say) 5%, then you will generally only have two significant figures on that number. You can't know the result of anything you multiply or divide by that number to better than 5% either, so the result won't have more significant figures than the number that went into it.

   Sometimes, it makes sense to report your result to one more or one less significant figure than what went into the calculation. This will make sense if you understand the "percent uncertainty" reasoning behind the number. For instance, if I tell you one stick 95 meters long, and another stick is exactly 1/9 as long as the first stick, the significant figure rule would suggest that you should only keep two figures, and report the answer as 11 meters long. However, the two significant figures on the first number means that you know it to about one part in 95. It would be better to report the answer as 10.6 meters long, since a result that is implicitly good to one part in 106 is much closer to your true precision than a result that is implicitly good only to one part in 11.

2. When adding or subtracting numbers, the answer is precise to the decimal place of the *least precise* member of the sum. If I add 10.02 meters to 2.3 meters, the answer is 12.3 meters. The second number was only good to the first decimal place, so the sum is only good to the first decimal place. Notice that the *number* of significant figures here is different from either number that went into the sum. When *multiplying*, it is the *number* of significant figures that is important; when *adding*, it is the *decimal place* that is important.

   Note that if I were to add 10.02 meters to 2.30 meters, the answer would be 12.32 meters; in this case, both members of the sum are significant to the hundreds place. It *is* possible to gain significant figures doing this. If you add 6.34 meters to 8.21 meters, each significant to three figures, the result is 14.55 meters, now significant to four figures.

   This rule makes sense again if you remember that significant figures represent the precision of a number. To what decimal place do you know all the things that you are adding or subtracting? You can't know the result to better than that decimal place.

3. A number which is *exact* should not go into considerations of significant figures. For example, suppose you're doing a unit factor conversion, and you multiply by the factor (12 in/1 ft). Your answer need not be limited to two significant figures because of this; there are *exactly* 12 inches in one foot. That's a *definition*; there is no uncertainty associated with it. In the first rule above, when I told you that the second stick was *exactly* 1/9 as long as the first stick, the 9 in 1/9 was a "perfect" number: you were told it was exact. Thus, that there is only one significant figure in the number 9 did *not* come into consideration for the number of significant figures in the answer.

4. **Always keep at least two or three more figures during intermediate calculations than you will report as significant figures in your final answer.** This is one of the two most common mistakes I observe in student work. (The other is thoughtlessly reporting your answer to however many digits your calculator gave you.) Otherwise, "round-off" errors will accumulate, and you may get the final answer wrong even though your general method and equations were correct. Consider, for example, summing the numbers 6.1 and 5.3, and multiplying the overall result by 4.1. The sum will be good to the first decimal place, and the final number will only be good to two significant figures because of the two significant figures in 4.1. The result is:

$$\begin{aligned} 5.3 + 6.1 &= 11.4 \\ (11.4)\,(4.1) &= 46.74 = 47 \text{ to two sig figs} \end{aligned}$$

   If, however, you round too soon, and don't keep the .4 at the end of the 11.4:

$$(11)\,(4.1) = 45.1 = 45 \text{ to two sig figs}$$

   In fact, you're now wrong! Even though both 11 and 4.1 are good to two significant figures, your result is incorrect to two significant figures. This is an example of "roundoff" error, where you lose precision by rounding numbers too soon.

You don't always have to get the number of significant figures *exactly* right. Significant figures are, after all, just an approximation of correctly taking into account and propagating your uncertainties, which is a topic that those who do more advanced studies in physical science will have to address. Just be reasonable, and make sure you understand the rationale behind why an answer might have a limited number of significant figures. It will often be acceptable to report an answer to one too many significant figures. However, it is technically incorrect to report a number that obviously has too many significant digits; in that case, you're misrepresenting your knowledge. By the same token, don't report a number with too *few* significant figures either, as in that case you're underselling what you know!

1.5 Dimensional Analysis

You can sometimes figure out something about a physical quantity just by considering its dimensionality. If you know what sorts of things *might* affect that quantity, and you have good reason to believe that it is just powers of those things multiplied together to give you that quantity, you may be able to figure out (up to a dimensionless constant) the equation that relates that quantity to the things that might affect it just by figuring out what makes the units work.

Consider the example of a simple pendulum: a small mass (the "bob") hangs at the end of a string. The other end of the string is fixed. The bob may oscillate back and forth. We want to figure out what is the equation for the period P (i.e. the length of time it takes to go through one oscillation). If we think about things that could affect that, there are three obvious possibilities. The first is the mass m of the bob at the end of the pendulum, the second is the length l of the string connecting the bob to the point from which the pendulum hangs, and the third is g, the acceleration due to gravity. For each of these quantities, we'll write down the dimensionality in terms of mass (M), length (L), and time (T). (Note that M here means mass, not meters!)

$$\begin{aligned} [m] &= M \\ [l] &= L \\ [P] &= T \\ [g] &= L/T^2 \end{aligned}$$

The "bracket" notation, here, means "dimensionality of". So, the dimensionality of the period is time; the dimensionality of acceleration is length divided by time squared.

If the period is a product of various powers of the different quantities, then we can write:

$$[P] \ = \ [m]^a \, [l]^b \, [g]^c$$

The period itself wouldn't be equal to this, as there may well be (and, in fact, there is) a dimensionless quantity multiplying everything else. However, even if we don't get the right formula, we can figure out how the period depends on these other things.

Now, put in the dimensions for each quantity:

$$T \ = \ M^a \, L^b \left(\frac{L}{T^2}\right)^c$$

$$T \ = \ \frac{M^a \, L^{b+c}}{T^{2c}}$$

Matching up the powers of each dimensionality on the left— which is simple, there is only T to the first power— to the powers on the right, we get these three equations:

$$\begin{aligned} a &= 0 \\ b+c &= 0 \\ -2c &= 1 \end{aligned}$$

In this case, the equations are easy to solve. The bottom equation gives us $c = -1/2$, and that together with the second equation gives us $b = 1/2$. So, we now know that:

$$P \propto l^{1/2} g^{-1/2}$$

$$P \propto \sqrt{\frac{l}{g}}$$

Without doing any of the actual physics to figure out the period of the pendulum, but *only* by considering the units on each quantity, we've figured out that the period must be proportional to $\sqrt{l/g}$. (If you want to figure out the dimensionless constant in front of $\sqrt{l/g}$, then in fact you do need to consider the full physics.)

Chapter 2

Energy

“Energy” is an extremely loaded term. It is used in everyday parlance to mean a number of different things, many of which bear at most a passing resemblance to the term as used in physical science.

At its core, energy is a mathematical construct that has turned out to be extremely useful. It shows up always with the same dimensionality, but in different forms. In a physical system, you can identify the forms of energy that are present, and calculate a number that represents the amount of energy there is for each of these forms. Ultimately, though, energy is just a mathematical construction that we calculate. What makes it so useful, however, is the observation that in all successful theories of physics thus far, *energy is conserved*. We could just as easily name and calculate an unending variety of other quantities for physical systems, but few are quite so useful as energy. If you take into account all of the various forms of energy in a complete system, you neither create nor destroy it. That is, in any interaction, the total amount of energy afterwards is exactly the same as the total amount of energy beforehand. Any energy lost by any part of the system must have been gained by another part of the system, and vice versa.

2.1 The Units of Energy

The SI unit for Energy is the joule, usually abbreviated J. One joule is equal to one kilogram meter squared per second squared:

$$1\,\mathrm{J} \;=\; 1\,\frac{\mathrm{kg\,m^2}}{\mathrm{s^2}}$$

Anything that is energy can be written as a number of joules. However, this isn't the only unit for energy. You are probably more familiar with another unit, the kilocalorie. (A kilocalorie is what is reported as mere *Calories* in food. The name is unfortunate, for there are 1000 calories in one Calorie; you can easily see how this would lead to confusion.) There are 4,184 joules in a kilocalorie; you can use this with the unit factor method (Section 1.3 to convert between the two forms of energy.)

A unit for energy that will be used more often in this course is the *electron Volt*, abbreviated eV. The conversion rate to joules is:

$$1\,\mathrm{eV} \;=\; 1.6022 \times 10^{-19}\,\mathrm{J}$$

The electron volt is a unit of energy well suited to the processes that happen in atoms. For example, it takes 13.6 electron volts to rip the electron off of a Hydrogen atom. This is a far more convenient number to use than 2.18×10^{-18}, the corresponding number of joules. It is important to remember, however, that the electron volt *is* the same dimensionality as joules, and that you can freely convert back and forth between the two. The eV is *not* a unit of voltage, nor, despite its name, is it specific to the electron; you can measure the energy of *anything* in eV. For example, if you consume 2,000 kilocalories worth of nutrition each day, it would be true, if not terribly illuminating, to say that you consume 5×10^{25} eV worth of food energy every day.

2.2 Forms of Energy

There are a number of forms of energy, but *most* of them can be reduced to either *kinetic energy* (energy of motion) or *potential energy* (energy of relative position). These two are discussed in greater detail below.

Heat energy, more accurately called *thermal energy*, is a form of energy that a bulk substance can have. As the temperature of an object goes up, its thermal energy content also goes up. Ultimately, however, thermal energy is just a form of kinetic energy. It is the vibrations of the molecules that make up the substance, or, in the case of the gas, the motions of the molecules zipping about that make up this thermal energy. When you heat water up, it gets hotter because the average speed at which water molecules are vibrating goes up. Indeed, that is what it *means* to say that water is hotter.

Internal energy is a catch-all term sometimes used to indicate energy that you're not keeping track of. As it sounds, this is energy that is, somehow, stored inside an object. In reality, this energy is made up of kinetic and potential energy. It may be that things inside the object are moving around, and thus the internal energy you're talking about is in the form of kinetic energy. You can, if you insist on painting with a broad brush, treat thermal energy as a form of internal energy. As another example, it may be that inside your object there are springs or other things that, as they move around, acquire potential energy as a result of their relative positions.

Chemical energy, sometimes called *chemical potential energy*, is, as the latter name suggests, just a special form of potential energy. It represents the energy that you could get out of a substance by performing chemical reactions with it. The chemical energy stored in gasoline may be treated as a form of internal energy, which you can extract and turn into other forms by burning that gas. On the microscopic level, what you're doing is rearranging the atoms into different molecules. That is, you're putting all of the atoms into different positions relative to each other.[1] Because potential energy is the energy of relative position, this means that you're changing the potential energy of all of these atoms.

Mass is, itself, a form of energy, leading to the term *mass energy*. Using Einstein's famous equation $E = mc^2$, you can convert from mass to other forms of energy. In chemical reactions, the amount of mass that is converted to or from energy is tiny— roughly one part in a billion. This is tiny enough that chemists will talk about the "conservation of mass", even though this is not strictly true. In nuclear reactions, however, the amount of mass that is converted to energy can be appreciable, approaching a percent. In matter-antimatter reactions, it is possible to convert *all* of the mass of reactants into other forms of energy.

Light energy, or more generally *radiation*, is energy in particles that are moving so fast (up to as fast as possible, in the case of light!) that their kinetic energy is much higher than their mass energy, if any.

Dark energy is in fact not energy in the classic sense of the word, but is the name given to the mysterious substance that fills the Universe and is driving its expansion to accelerate. We know next to nothing about dark energy, and we certainly don't know how to convert it to other forms of energy.

2.2.1 Kinetic Energy

If an object with mass m is moving with speed v, then the amount of kinetic energy that object has is

$$KE = \frac{1}{2} m v^2$$

. If you look at the dimensionality of this equation, you will see that on the right we have mass times length squared divided by time squared. In SI units, that would be $\mathrm{kg\,m^2\,s^{-2}}$, which is the Joule. It is comforting to see that this equation does give us the right units for energy. This equation only works as long as the speed v is a lot less than the speed of light. Once the speed approaches the speed of light, you have to take into account Relativity, and things become more complicated. Why the $\frac{1}{2}$? The

---

[1] Later, we'll see that talking about the position of particles in atoms is a bit troublesome, but for now this description is reasonable.

answer may not satisfy you: because that's what works. You can derive this from forces using a little bit of calculus, but even that derivation requires other definitions that may seem arbitrary. Ultimately, we've found that if we use this formula for kinetic energy, rather than something else times $m\,v^2$, the notion of conservation of energy works.

It is also possible to have kinetic energy if you are at rest: you can have *rotational kinetic energy* if you are rotating. However, at the microscopic level, ultimately this is the same thing. Imagine a ball that's at rest, but rotating. If you think about each little piece of the ball— each molecule in the ball, if you will— the ones that are not right on the axis of rotation are in fact themselves moving about the center of the ball. The ones closer to the axis are moving slower than the ones farther away. What we call rotational kinetic energy is just a way of summarizing this motion of all of the little pieces of the ball.[2]

2.2.2 Potential Energy

Potential energy is energy of relative position. Except in esoteric situations where general relativity and quantum field theory tentatively approach each other, the absolute value of potential energy doesn't matter. All that matters are the *differences* in potential energy as particles rearrange themselves into different relative positions. This means that you could add any constant (with energy units) you want to the potential energy of a system, and, as long as you don't change the constant you're using partway through a problem, all of your energy calculations will come out right. Frequently, but not always, we choose the constant such that the potential energy is zero for particles that are infinitely far away from each other. This convention makes sense; you don't want to have to think about having some energy to carry around for a particle that is so far away that it's not meaningfully interacting with any of the particles you do care about.

Technically, you can't talk about the potential energy of a single object. Really, the potential energy is in the interaction of that object and another object. To be proper potential energy, it must depend only on their positions relative to each other. It doesn't make sense to talk about "the potential energy of the Earth". However, it does make sense to talk about the potential energy of the Earth-Sun system.

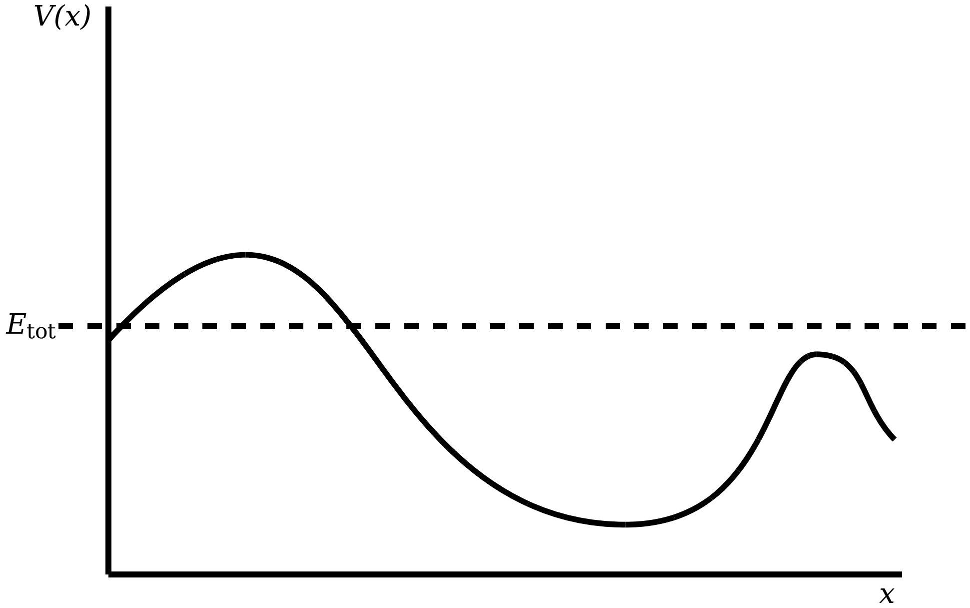
Sometimes, we can make the approximation that one particle is much smaller than everything else it is interacting with. In that case, we will talk about the potential energy *of* that particle. For example, if you lift a ball off of the ground, as that

---

[2] The formula for rotational kinetic energy is $\frac{1}{2}I\omega^2$. I is the *moment of inertia*; it depends on the mass, size, and geometry of the object. ω is the *angular velocity*, in radians per second. It is equal to 2π times the number of rotations per second the rotating object is making. You will read more about I and ω in Chapter 3.

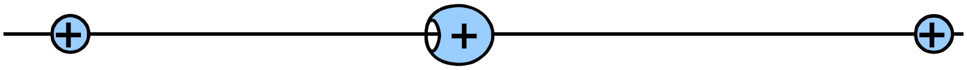
ball and the center of the Earth get farther away from each other there is more and more potential energy in the gravitational interaction between the ball and the Earth. However, the gravity of the ball on the Earth is extremely unimportant to the Earth, whereas the gravity of the Earth on the ball is extremely important to the ball. As such, we can treat the ball as a particle moving within the "fixed potential of the Earth". We then say that the ball has a certain amount of potential energy based on its height above the ground. Implicitly, this is really the potential energy in the interaction of the ball and the Earth, but it is more convenient to treat it as the potential energy of the ball, with the understanding that we're working in the (very valid) approximation that the ball is much smaller than the rest of the system (i.e. the Earth).

Different interactions (i.e. different forces) have different functional forms for potentials. For the moment, you won't need to use any of them. If you have had physics before, you may know some of them. For an arbitrary force or combination of forces, you could construct a potential energy function $V(x)$. It is useful to think of an analogy between a particle moving in a potential and a car rolling about on hilly ground. Suppose that $V(x)$ had the following form:

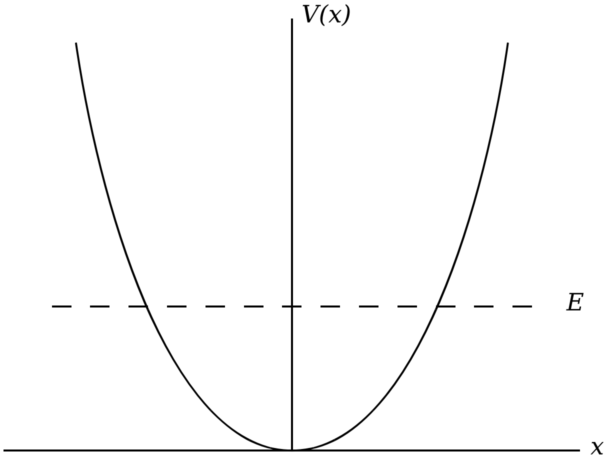


The dashed line on the plot indicates the total energy available to the particle. Imagine that instead of potential energy, the vertical axis where the height of hills, and imagine that the particle is a car. When the car is at a lower point, it has less potential energy, and thus more kinetic energy, and thus is moving faster. The car cannot get to places higher than the dashed line: it's not moving fast enough to make it that far up the hill. By thinking about potential energy in this manner, you can visually get an idea for how particles will move around in a given potential, even if you don't know all of the classical physics needed to work it out.

As another example, suppose you have a wire with two positive electric charges fixed to it. Sliding smoothly along the wire is a bead that also has a positive electric charge on it:



Positive electric charges will repel each other. As such, if the bead will be pushed away from the two positive charges at either end of the wire. Call x the position of the bead along the wire, with $x = 0$ the exact center of the wire. There will be a potential energy function $V(x)$ for the interaction between the bead and the two charges on either end of the wire. To make things more interesting, let's suppose that the bead has some total energy E that is *greater* than the minimum of the potential $V(x)$.



The minimum of the potential energy is where the bead "wants" to be. In this case, the bead is pushed away from the positive charge at either end. If you imagine a ball rolling in this potential, it would experience the same thing; it would want to move towards the center if it were up either side of the potential well. However, looking above at the picture of the bead on the wire, the bead makes no actual motion *down* in space; it's only moving to lower potential. Notice that we've chosen to make $V(x) = 0$ at the center of the wire. Remember that that is completely arbitrary; we could add a constant to the potential energy, and it wouldn't make any difference. (We would have to add the same constant to the total energy of the particle to keep things consistent, however!)

What happens if the bead is at $x = 0$? We can see that it's potential energy $V(x)$ is equal to zero. However, its *total* energy is something greater than that. That means that the bead must have some other form of energy. As we've defined the system, the only other form of energy the bead could have is kinetic energy. This means that if the bead really does have energy E as indicated on the plot, it *must* be sliding either to the left or to the right if it's at $x = 0$. Indeed, at any x, it will satisfy $\frac{1}{2}mv^2 \ + \ V(x) \ = \ E$.

Chapter 3

Momentum and Angular Momentum

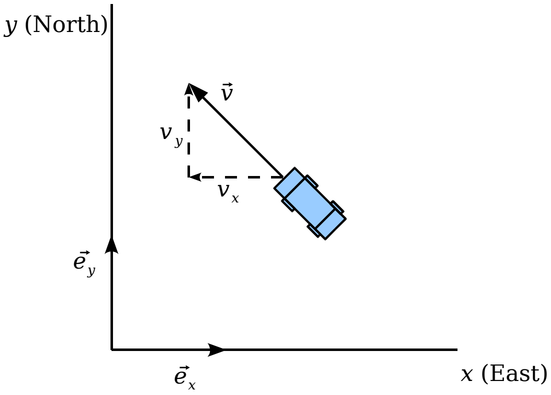
Chapter 2 introduced energy as a mathematical construct that has turned out to be very useful. There are two other conserved quantities that show up throughout all of our theories of physics. Both of those have to do with motion, but are different from kinetic energy. These other quantities are *momentum* and *angular momentum*.

3.1 Vectors

Both momentum and angular momentum are *vector* quantities. Later, we'll be talking about a more abstract form of vector used to represent the state of quantum systems. Here, we're talking about a special kind of vector, a vector in regular old 3d-space. Distinguish these from more general vectors, I shall call them 3-vectors, in reference to 3-dimensional space. A 3-vector is anything that has both a magnitude (size) and direction. For example, consider speed: the speed an object is moving is just a number. (We would call that a *scalar*.) Likewise, the kinetic energy of an object is a scalar; it's an amount of energy, and there is no direction associated with it. In contrast, the velocity of an object includes not just its speed, but also its direction. So, you could say that the speed of a car is 80 km/h. If you wanted to specify its velocity, you'd also have to give its direction. For example, you could say that the velocity of a car is 80 km/h due northwest.

3-vectors corresponding to different physical quantities will have different dimensionalites (and thus different units) associated with them. *Displacement* is a 3-vector form of distance. Distance just tells you how far apart two things are. Displacement tells you how far apart and in what direction. Just like distance, displacement comes in length units. So, you might say that one person is 1 meter due east of another person; in this case, the displacement from the other person to the first person is 1 m due east.. The *magnitude* of a 3-vector is just its size. Distance is the magnitude of displacement. If you consider that person whose displacement was 1 meter due east of the other person, you could also say that the distance between the two people was 1 meter. This is correct, even though it has less information.

A 3-vector can be visualized as an arrow in space. The length of the arrow represents the magnitude, and the direction the arrow points is the direction of the 3-vector. So, for example, let us consider a car going at 50 km/h due northwest.



The picture shows the x and y axes, representing East and North respectively. The z axis not drawn; it's up, straight out of the page. We notate vectors by drawing a little arrow on top of them; you can see the $\vec{v}$ in the diagram referring to the velocity of the car. Also shown are $\vec{e}_x$ and $\vec{e}_y$, the two *basis vectors*. In the case of 3-vectors, we can also call the basis vectors *unit vectors*, as they are 3-vectors whose length 1 (dimensionless), and that point right along the axes. The basis vectors define the coordinate system that we're using; here, they just define x and y for us. The car's velocity 3-vector $\vec{v}$ is represented by the direction and length of the arrow sticking out of the front of its picture. We can see that it points partially in the negative-x direction, and partly in the positive-y direction.

If you have a complete set of basis vectors, you can construct any other vector out of them. The three unit 3-vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$, all 3-vectors of length dimensionless 1 pointing (respectively) along the x, y, and z axes, form the most obvious and most generally useful set of basis 3-vectors. Any other 3-vector can be written as a sum of constants times those basis vectors. So, here, we could say that:

$$\vec{v} \;=\; v_x\,\vec{e}_x \;+\; v_y\,\vec{e}_y \;+\; v_z\,\vec{e}_z$$

From looking at the picture, we can see that in this case v_x is going to have to be negative, v_y is going to have to be positive, and v_z is going to be zero. How do we figure out where they are? Well, we know that the car is going due northwest, so we expect that the absolute value of v_x and v_y will be the same (it's got just as much north velocity as west velocity). For the total speed, that is the total length of the 3-vector, we recognize that there's a right triangle there, and use a generalization of the Pythagorean Theorem:

$$v^2 \;=\; |\vec{v}|^2 \;=\; {v_x}^2 \;+\; {v_y}^2 \;+\; {v_z}^2$$

$$v \;=\; \sqrt{{v_x}^2 \;+\; {v_y}^2 \;+\; {v_z}^2}$$

In this example, we know that $v = 50$ km/h. For this to work, we have to have $v_x = -35$ km/h and $v_y = 35$ km/h.

3.2 Momentum

Kinetic energy is a quantity that's associated with motion. However, kinetic energy itself is not always conserved. If a cue pool ball runs into another ball, and the cue ball stops dead, the other ball goes off with the same speed that the cue ball came in at. In this case, the two balls have the same mass, so $\frac{1}{2}mv^2$ is the same both before and after the collision; kinetic energy *is* conserved in the collision. However, if two cars hit each other in a head-on collision, and the tangled wreck of the two cars stops dead at the point of impact, kinetic energy is clearly not conserved, as the v of everything after the collision is zero. It's not kinetic energy that's conserved, but total energy. The kinetic energy that the cars had before the collision is, during the collision, converted into other forms of energy: heat, noise, and possibly some potential energy as the structure of the car is rearranged. So, *sometimes*, in some collisions, kinetic energy is conserved. However, in other collisions, kinetic energy is not conserved. Note that total energy is always conserved; it's just that there are forms of energy other than kinetic energy, and sometimes kinetic energy can be converted to or from those other forms.

However, there is a quantity of motion that is conserved in every collision. If it is to be conserved in both the examples above, it can't just be based on the speeds of the particles. While the speed would seem to be enough in the example of the pool balls, in the example of the cars there was a lot of speed to start with, but *no* speed after the collision. To work in both of these examples (and in general), this conserved quantity has to be something that takes into account both speed *and* direction. That quantity is momentum. It is traditional to use the letter p to represent momentum. The momentum of a particle is defined by:

$$\vec{p} \;=\; m\,\vec{v}$$

where m is the mass of the particle, and $\vec{v}$ is the *velocity* of the particle. The magnitude of $\vec{v}$ is traditionally written $|\vec{v}|$, but is often just abbreviated as v without the arrow. Magnitudes of 3-vectors are *always* positive or zero; it does not make sense to say a 3-vector has a negative magnitude. The magnitude of velocity is what we call speed. You can't have a speed of -50 km/h, but you *can* be moving at 50 km/h in the negative-x direction.

One way of dealing with 3-vectors is to break them into components— an x-component v_x, a y-component v_y, and a z-component v_z. For now, to keep things simple, we'll only consider motion in one dimension, so that particles will not have any component of velocity in the y or z directions. Therefore, we can say that the particle's velocity is v_x in the $+x$ direction. If v_x is negative, it means that the particle is moving to the left. The *speed*, however, the magnitude of the velocity, is still positive; that's just how fast it's going, without reference to direction.

Just like total energy, it turns out that momentum is a conserved quantity. If you take everything into account (which is occasionally tricky), the total momentum before and after a collision or interaction must be the same. Consider the example of the two pool balls above. If the pool balls have a mass m and an x-velocity v_x, then the initial momentum is just mv_x. After the collision, it's the other ball that's moving, but the speed is the same, so the final momentum is mv_x. The total momentum is conserved.

In the case of the two cars colliding, suppose that both cars have the same mass m and are approaching each other with speed v. The car that is moving to the right has x-momentum $mv_{x1} = mv$, and the car that's moving to the left has x-momentum $mv_{x2} = -mv$. Notice that the x-momentum of the car moving to the left is *negative*! In contrast, the kinetic energy of both cars is positive, and is the same: $\frac{1}{2}mv^2$. The *total* momentum in the system is the sum of the momentum of the individual particles. Thus, the total x-momentum is $mv_x + m(-v_x) = 0$. After the collision, the velocity is zero, so the total momentum is still zero. Momentum is, in fact, conserved in the collision.

3.2.1 The Units of Momentum

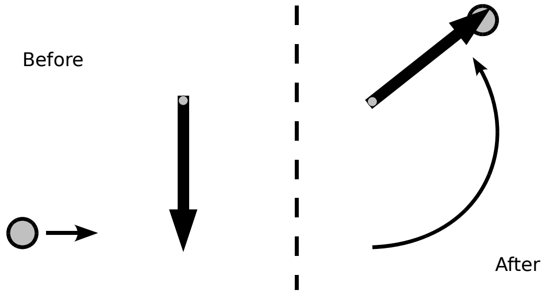
There isn't a special name for the units of momentum. If you look at the equation

$$\vec{p} = m\vec{v}$$

don't let the vector signs bother you. Velocity has dimensionality of length over time just like speed. If you multiply that by mass, you get a dimensionality of mass times length divided by time. Because the dimensionality must be the same on both sides of the equation, that is also the dimensionality of momentum. In the SI system, momentum comes in units of $\mathrm{kg\,m\,s^{-1}}$.

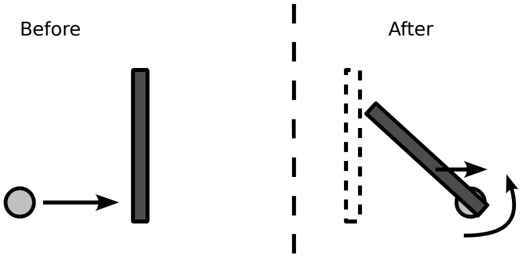
3.3 Angular momentum

Imagine the following experiment. You've got, somehow, a frictionless plane. (These frictionless planes are common in physics, but much more difficult to manufacture in the real world! If you wish, you can imagine it as an air hockey table, or a particularly smooth and slippery sheet of ice or teflon.) A hockey puck is sliding along the plane, where it hits a big clock hand, and *sticks* to the end of the clock hand. The other end of the clock hand is nailed into the ground, so that it's not going anywhere. After the hockey puck hits the clock hand, the clock hand starts spinning around.



On first glance, you might think, wait! Momentum isn't conserved here! The clock hand may be spinning around, but it's no longer moving off in one direction, whereas before there was clearly momentum in the x-direction! However, remember that the clock hand is nailed into the ground. That means when the puck collides with the clock hand, the clock hand will push on that nail, which pushes on the ground, and effectively the *whole earth* is pushed off (very, very slowly!) to the right. Momentum is conserved, but you have to consider *everything* that's interacting to keep track of all of it.

So, let's do another experiment. Let's collide the puck with a bar— and still have it stick— but not nail that bar to the ground. What happens now is that after the collision, the bar *does* move off to the right only not as fast. It's still rotating, though.

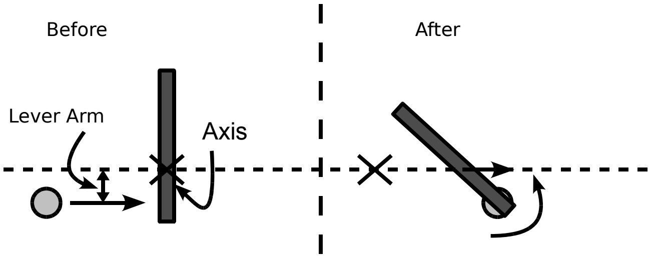


Here, momentum is conserved. The combined system moves off to the right at a lower speed than the puck came in because it's a more massive system. However, there's also the rotation. There's clearly some kinetic energy associated with that, as bits of the rod have motion about the center of the rod in addition to the bulk motion of the rod as a whole. It turns out that there is yet another quantity, called *angular momentum*, that is conserved in interactions. Be careful about the name! Angular momentum is not a "special kind" of momentum. It is a wholly different quantity, with different units, that is conserved separately. It bears some similarities to momentum, and thus the name is similar, but it is a different thing. You cannot mix momentum and angular momentum; again, remember that they are *two different things*.

Back to our example here: the question is, if there's angular momentum afterwards in the rotation of the rod, what angular momentum is there before the collision? It must be there, if angular momentum is a conserved quantity!

In order to measure angular momentum, you must choose an axis to measure it about. How can angular momentum be a conserved quantity if you can choose any arbitrary axis you may ask? The answer is that angular momentum is conserved about *any* axis, as long as you stick with the same axis all the way through the problem.

If a particle is moving in a straight line directly towards or directly away from your chosen axis, then it has *no* angular momentum. However, if it's motion is offset from the axis, even if it's moving in a straight line, it still has angular momentum. To figure out the angular momentum, you multiply the *lever arm* by the momentum of the moving particle. The lever arm is the *perpendicular* distance from the axis to the line of motion of the particle. For example:



I've cleverly chosen my axis to be on line with the motion of the center of mass of the system after the collision. That means that the after the collision, the linear motion of the center of mass of the system makes no contribution to the angular momentum; all of the contribution comes from whatever the rotation is doing. Before, however, the linear momentum of the puck does contribute angular momentum. The lever arm is, as drawn, the perpendicular distance from the axis to the line of motion of the particle. If we call that perpendicular distance d, then the angular momentum (for which we traditionally use the letter l) is:

$$l = dp = dmv$$

where m is the mass of the puck and v is the initial speed of the puck.[1]

How do you figure out the angular momentum of a rotating object? The hard way to do it is to consider the object as a collection of a lot of little pieces of object. For each small piece of that object, you multiply the small mass of that piece by the speed of that piece resulting from the rotation by the lever arm from the axis to that piece. Add up what you get, and you have the object's angular momentum. In practice, for most objects we're able to define a single number that we call the *moment of inertia*, which takes care a of a lot of that for you. This is a quantity that adds up all of the bits of mass and the distances of those bits of mass from a specified axis of rotation for an object. It takes into account the mass part of momentum and the lever arms for all those little bits of the object. The angular momentum of an object rotating about a given axis is then:

$$l = I\,\omega$$

where I is the moment of inertia of that object about the axis and ω is the *angular speed* of the rotation. To figure out angular speed, first figure how how long it takes for the object to make one complete rotation; call that the period T. The angular speed is then:

$$\omega = \frac{2\pi}{T}$$

ω then has dimensionality of one over time; the SI unit for ω is s^{-1}.

3.3.1 The Units of Angular Momentum

If you look at the equation

$$l = d\,p$$

where l is angular momentum, d is the lever arm to a moving particle, and p is the magnitude of the momentum of that particle, you can figure out the units of angular momentum. d has dimensionality of length, of course, and as we worked out in Section 3.2.1, the dimensionality of momentum is mass times length divided by time. Thus, angular momentum has dimensionality mass times length squared divided by time. The SI unit for angular momentum doesn't have a special name; it's just $\mathrm{kg\,m^2\,s^{-1}}$.

3.3.2 The Direction of Angular Momentum

Just like momentum (sometimes called "linear momentum" when you want to be clear that you're not talking about angular momentum), angular momentum is a 3-

---

[1] If you are familiar with vectors, in fact the real definition of angular momentum is $\vec{l} = \vec{r} \times \vec{p}$, where $\vec{r}$ is the displacement from the axis to the position of the moving particle and $\vec{p}$ is the particle's momentum.

vector. With regular momentum, it's pretty easy to figure out what the direction of the 3-vector is: it's the direction that the object is moving. What, however, is the direction of angular momentum? If an object is spinning, it assuredly has angular momentum. However, the bits of the object are all moving in different directions (the bits on one side of a rotating disk are moving in the opposite direction from the bits on the far side), and what's more, later any given bit of the object will be moving in a different direction.

It turns out there *is* a unique direction for rotation: the axis about which an object is rotating. As such, we can define the direction of the angular momentum 3-vector to be pointing along the axis of rotation. If a Frisbee is flying through the air, rotating, and is parallel to the ground, you would say that its angular momentum 3-vector points either up or down.

How do you figure out up or down? This is just a matter of convention. The convention we use is called the *right-hand rule*. What you do is curl the fingers of your right hand so that they point around in the direction of the rotation. Stick your thumb straight out, and it points along the direction of the angular momentum 3-vector. For example, if you're looking down on a Frisbee, and the Frisbee is rotating counter-clockwise, you would say that it's angular momentum 3-vector is pointing straight up. (Try using your right hand to see why that would be the case.)

Chapter 4

The Spin-$\frac{1}{2}$ particle

Moving electric charges, or currents, interact with magnetic fields; they both respond to them, and create them. You know from Section 3.3 that a spinning ball has angular momentum. If that spinning ball is also charged, that means that, effectively, there are currents associated with the ball. Suppose that the charge is spread uniformly throughout the ball. The charges right along the axis aren't moving, and so wouldn't respond to or create magnetic fields. However, all of the bits of ball that aren't right along the axis are making a circle around the axis. As such, they are moving charges, and they will respond to a magnetic field.

This may seem like a completely unfounded leap, or it may seem like an obvious leap, but from this observation, we can say that a particle that has both charge and angular momentum will respond to magnetic fields.

4.1 Particles in Quantum Mechanics

When we talk about a "particle" in quantum mechanics, we mean something that behaves as if it were just a single body. However, we are often also talking about a particle as it is understood in the Standard Model of Particle Physics. In the Standard Model, a fundamental particle is something that is effectively a mathematical point. As far as we can tell, the fundamental particles have *no* spatial extent. The most common everyday example of a particle from the Standard Model is the electron. You may be familiar with electrons if you have taken any chemistry classes in the past. Atoms are made of of electrons orbiting nuclei. Nuclei themselves are made up of protons and neutrons. Protons and Neutrons may be treated as particles in quantum mechanics, but in fact they are not fundamental particles. Rather, they are themselves made up of *quarks*, which are (at least as far as we understand) fundamental particles.

If a fundamental particle doesn't have a size, what can it have? Well, it can have a position, and it can have a momentum. Later, we will find out that there must be some uncertainty associated with one or both of these quantities for any given particle, but these are quantities that you can figure out for the particle. However, they aren't really fundamental to the particle; they just say where the particle is, or, effectively, how fast it's moving relative to something you've chosen to measure speeds relative to. Similarly, if the particle is an electron in an orbital in an atom, it can have angular momentum as a result of that orbit. Again, this isn't a fundamental property of the particle, but there result of its interaction with the atomic nucleus.

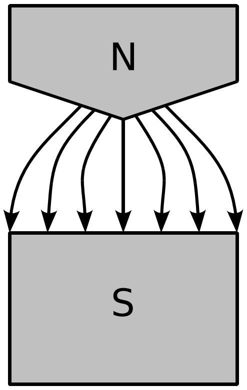
The mass of the particle *is* a fundamental property of the particle. Likewise, the electric charge of the particle. The electric charge on the electron, in SI units, is -1.602×10^{-19} C. In fact, often when we are dealing with atomic and subatomic particles, we'll measure charge in terms of the elementary charge e, which is defined as the absolute value of the charge on the electron: $e = +1.602 \times 10^{-19}$ C. (It is unfortunate that the notation for the elementary charge is the same letter as e, the natural exponential that shows up, for instance, in the mathematical model for radioactive decay. You need to be careful about the context whenever you see an e, so that you can figure out whether we're talking about the natural exponential, the charge on the electron, or something else..)

Another property of fundamental particles is their angular momentum. Because this is fundamental to the particle itself, we refer to it as the *spin* of the particle. As an analogy, consider the Earth orbiting the Sun. The Earth has *orbital* angular momentum as a result of the circle it makes yearly about the Sun. It also has *spin* angular momentum as a result of its daily rotation about its own axis. Where the analogy breaks down, however, is that the Earth is indeed an extended ball; the electron, on the other hand, is a point particle, and has no spatial extent. As such, there really isn't anything spinning around anything else to create this angular momentum. This is conceptually difficult; how, then, can the electron have angular momentum? Alas, the best answer we can give is that it just does. Experiments have shown that indeed electrons behave as if they have angular momentum, and that they can transfer angular momentum to other particles and systems when they interact with them.

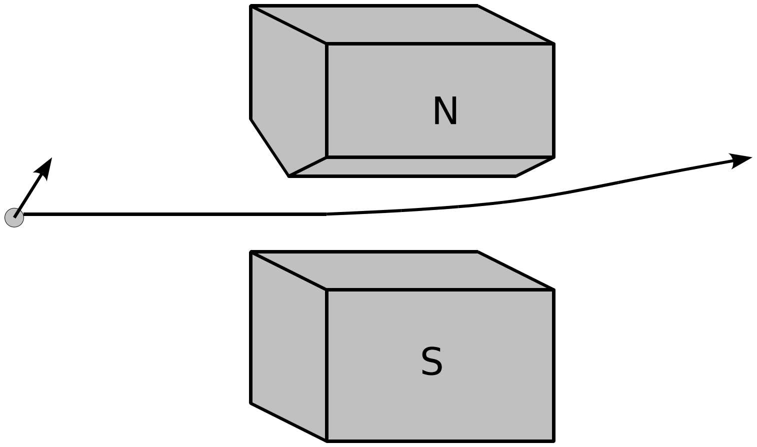
Just as every electron has exactly the same mass and exactly the same electric charge, every electron has exactly the same total angular momentum. (We will see later what the value of that angular momentum is.) You can't cut off a piece of an electron to leave behind a particle that is a part of an electron, with a lower mass and possibly a lower electric charge. Similarly, you can't speed up or slow down the spin of an electron, the way you can get a top spinning faster or slower. All electrons are effectively spinning at the same rate— only, remember, they're not really little balls spinning at all, but rather angular momentum is just one of the properties associated with those quantum particles we call electrons.

4.2 Measuring Electron Spin: the Stern-Gerlach Experiment

If a particle that has both charge and angular momentum interacts with magnetic fields, and if we know what that charge is through other experiments, then we ought to be able to figure out the angular momentum of that particle by some sort of experiment involving magnetic fields. If a particle with charge and angular momentum moves through a *nonuniform* magnetic field, it will be pulled along the direction of the nonuniformity based on the projection or component of its angular momentum along the direction of the magnetic field nonuniformity.



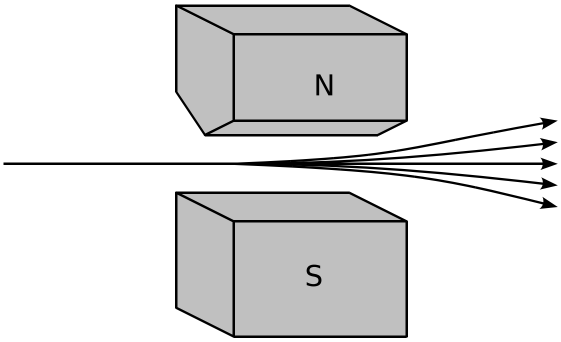
A nonuniform magnetic field, as seen by a particle that will be shot into the page through it.



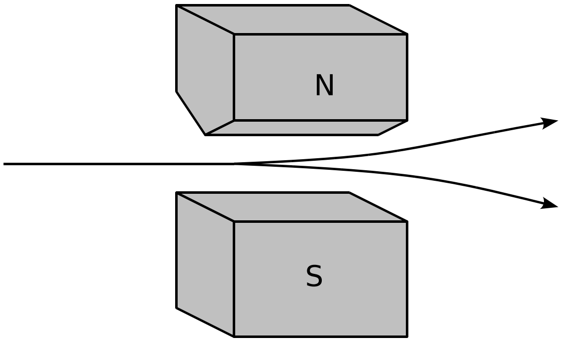
A charged particle with some component of angular momentum along the direction of the nonuniform magnetic field will have its path bent by that field. Whether the path bends up or down depends on the charge of the particle and the direction of the angular momentum.

Remember that angular momentum is a 3-vector. For a spinning object, the angular momentum 3-vector is oriented along the axis about which the object is spinning. To figure out which direction along that axis the angular momentum points, you use the right-hand-rule: orient your *right* hand so that if you curl your fingers, they point along the sense of rotation. Then, your thumb points along the direction of the angular momentum 3-vector. For a classical spinning object like a top or a planet, that angular momentum 3-vector can point in any direction. Indeed, the angular momentum 3-vector of the Earth's rotation is pointed at an angle of 23.5° with respect to the angular momentum 3-vector of the Earth's orbit; they're not perfectly aligned.

Let's imagine what a classical physicist, having accepted (somehow) that all electrons have exactly the same angular momentum, would expect to see if he sent a beam of electrons through a nonuniform magnetic field that bent electrons along the z-axis. If an electron's angular momentum happened to be oriented entirely along the $+z$-axis, its path would be deflected upwards the maximum amount. If its angular momentum happened to be oriented entirely along the $-z$ axis, its path would be deflected downwards the maximum amount. Most of the electrons would have their angular momentum 3-vector randomly oriented somewhere in between, and so the beam should spread out into a vertical smear as it passed through the nonuniform magnetic field.



In the early 1920's, two physicists, Otto Stern and Walther Gerlach, performed this experiment.[1] What they observed was not a continuous smear, but rather that the beam split into two different beams.



Think about what this means. This means that when you take a beam of electrons whose angular momenta are all randomly oriented, if you measure the z component of angular momentum you get one of *only two different values*. The component of spin angular momentum of an electron along the z-axis is either $5.27 \times 10^{-35}\,\mathrm{kg\,m^2\,s^{-1}}$, or $-5.27 \times 10^{-35}\,\mathrm{kg\,m^2\,s^{-1}}$. The z-component of the spin angular momentum of the

---

[1] Stern and Gerlach *did* measure the spin of the electron, but at the time they *thought* they were measuring quantized orbital angular momentum! For the history of this experiment, see Bernstein (2010).

electron is *quantized*. These values of angular momentum relate Planck's constant $\hbar$ (pronounced "h-bar"), which has the value $\hbar = 1.055 \times 10^{-34}\,\mathrm{kg\,m^2\,s^{-1}}$. When the z-spin of an electron is measured, it comes out to either $+\hbar/2$ or $-\hbar/2$. Indeed, it becomes much more convenient to measure angular momentum in units of $\hbar$ in quantum mechanics, so we refer to the electron as a "spin-$\frac{1}{2}$ particle". Remember, however, that whenever somebody says that an electron is measured to have z spin of $1/2$, they really mean that the z component of its angular momentum is $+\hbar/2$.

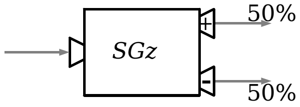
We define an *observable* as a quantity that we could, at least in principle, measure. The position of a particle is an observable, as is its momentum. The z component of the spin angular momentum of an electron is an observable. One of the primary features of quantum physics is that many observables have the same property that we see for electron spin: when they are in fact observed, they take on one of a finite number of values. They are quantized. It is this property from which quantum mechanics takes its name.

In Quantum Mechanics, many observables are quantized. That is, when measured, they take on one of a finite number of possible values.

It's tempting to think of the electrons whose z spins are $+\hbar/2$ as having their angular momentum oriented entirely along the $+z$-axis, and those whose z spins are $-\hbar/2$ as having their angular momentum oriented entirely along the $-z$-axis. Indeed, physicists will often refer to "spin up" and "spin down" particles. However, the *total* angular momentum of an electron is actually $(\sqrt{3}/2)\hbar$. That means that you *never* observe an electron with its spin oriented entirely along the z axis! There must always be some component of spin oriented along another axis.

4.2.1 The Stern Gerlach Machine

As we continue to explore electron spins in quantum physics, we're going to use a measuring device that repeats the Stern-Gerlach experiment so often that it's worth describing an imaginary "Stern-Gerlach machine". Such a machine has a single input, into which you send a beam of electrons (or even just a single electron). It has two outputs, one for electrons whose angular momentum has been measured as positive along the axis of the machine, the other whose angular momentum has been measured as negative along the axis of the machine. There's no reason why the Stern-Gerlach experiment has to measure the z component of electron spin. By rotating the magnets used in the device, you could measure the x component or y component of the spin. (It's trickier to measure component of spin along the direction of motion of the particle, but that can be done as well.) We will draw an SG machine as follows:

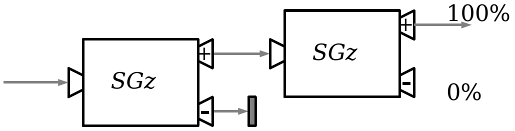


Each time we have an SG machine, it will be named such that the third letter tells you the axis along which its measuring the angular momentum. Thus, an SGz machine measures the z spin of an electron, and an SGx machine measures the x spin of an electron. You could also imagine an SG machine that has its axis oriented at some other angle θ with respect to the z axis. (If that angle is 90°, then it's an SGx machine.) In that case, we will call it an SGθ machine.

4.3 Repeated Measurements of Spin

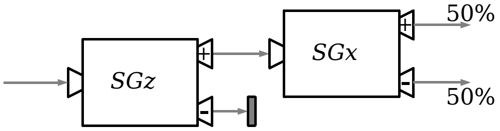
If you have a beam of electrons with randomly oriented spins, when you measure the z spin of the beam you get half of the electrons showing a spin of $\hbar/2$ and half showing a spin of $-\hbar/2$.

Suppose that you block off the beam with negative z spin. Send the beam with positive z spin into a second Stern-Gerlach machine. What do you get?



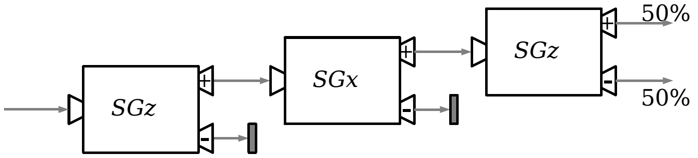
Unsurprisingly, the second SGz machine shows that every electron that comes into it has $+z$ spin. You wouldn't expect anything else. After all, we divided up the electrons that went into the first SGz machine based on their z spin, and threw out the ones that didn't have $+z$ spin.

What if you put the $+1/2$ out put of the SGz machine into an SGx machine?



When the x component of the spin angular momentum of an electron is measured, just as with the z component it only takes on values of $+\hbar/2$ or $-\hbar/2$. Because the x axis is perpendicular to the z axis, you wouldn't expect knowing whether the z spin of the electron was along the $+z$ or $-z$ direction to tell you anything about the whether the x spin was positive or negative. And, indeed, that's what's observed. For each electron with spin $+z$ that goes into an SGx machine, there's a 50% chance it will come out the $+x$ output, and a 50% chance it will come out the $-x$ output.

Things get interesting when you add one more SG machine to the mix. Take the electrons that were first measured to have $+z$ spin, and were then measured to have $+x$ spin. That is, at the first SG machine (an SGz machine), we're throwing out the electrons with $-z$ spin, and at the second SG machine (an SGx machine), we're throwing out the electrons with $-x$ spin. What happens if you send these electrons through another SGz machine? You might expect all of them to come out through the $+z$ output; after all, we already know from a previous measurement that all of these electrons have a positive z component of spin angular momentum. In fact, however, this is *not* what's observed! If you construct this experiment, you find that the final SG machine, an SGz machine, puts out electrons through either output with a 50% chance for each!



The fact that angular momenta were quantized was the first thing about quantum mechanics that was completely at odds with our intuition and our experience with classical physics. This is the second thing. It seems that, somehow, by measuring the x spin of the electrons, we *lost* information about the z spin of the electrons. To explain this and similar experiments, the theory of quantum mechanics includes formalism that shows that it is impossible to know certain pairs of observables at the same time. This is related to the famous Heisenberg Uncertain Principle, about which we will say more in a later chapter. If you know the z spin of an electron, you know nothing about its x spin; were you to measure the x spin, you have a 50% chance of measuring either $+1/2$ or $-1/2$. Likewise, if you know the x spin, you know nothing about its z spin.

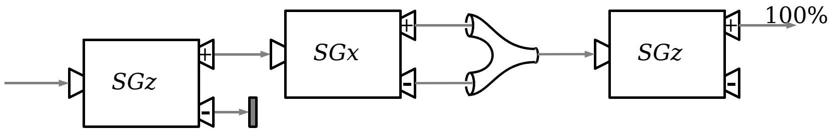
The same result is observed if, instead of the $+x$ output, we take the $-x$ output of the second machine. We have a beam of electrons who all were first measured to have positive z spin, and were then measured to have *negative* x spin. As before, if we measure the z spin again, we find that we have a 50% chance of measuring $+z$, and a 50% chance of measuring $-z$.

The quantum weirdness goes deeper than that. It turns out that it's not just that you *don't know*. The particles themselves do not have a definite state! If you've measured the z spin of an electron, the electron *does not have a definite x spin*! The jargon we use to describe this is to say that the electron is in an "indefinite state", or that it is in a "mixture of states". In this case, the x spin state of the electron is a mixture of the $+1/2$ and $-1/2$ states.

In Quantum Mechanics, certain observables may not be known— do not even take on definite values— at the same time as certain other observables.

At this point, you might object, reasonably so, that we could have neglected an effect of our measuring devices. Charged particles with angular momenta interact with magnetic fields. Could it not be that our devices aren't only deflecting the electrons' paths, but also rotating those electrons? That is, after the first SGz machine, the electrons coming out of the $+z$ output do have z spin of $+1/2$. But when they go through the SGx machine, perhaps it's not just *measuring* the x spin, but also rotating the electrons so that their angular momenta no longer as up along the z axis as they were before. Indeed, it's clear that the state of the system is changed when the x angular momentum is measured. Must it really be something particular to quantum mechanics?

To answer that question, suppose that after we've sent the beam through the SGx machine, dividing it into a beam of electrons with positive x spin and a second beam with negative x spin, we recombine those two beams. Take the recombined beam and put that into the third SGz machine. What do we observe?



The beam *did go* through the SGx machine, so any effect it has on the beam has happened. Remember that the beam coming out of the $+1/2$ output of the SGx machine had an indeterminate z spin; likewise for the beam coming out of the $-1/2$ output of the SGx machine. Yet, somehow, by recombining the beams, we are able to restore the information about the z spin of the electrons! Again, if we make it so that the beam has a very low intensity, and only one electron is going through the apparatus at a time, exactly the same result is observed. In a sense, by recombining the beams, we never really did measure the x spin of the electron. Sure, the SGx machine measured it…but we never let the measurement go beyond that, we never let it go into any other experimental apparatus, we didn't let any physicists know about it, we didn't record the spin of any given electron.

There is something peculiar about *measurement* that changes the state of a system. Yet, exactly what is a measurement is not entirely clear. Indeed, the "measurement problem" in quantum mechanics has troubled physicists for nearly a century, and remains a point of active debate today. We will discuss this in greater detail in a later chapter.

For the time being, however, review the results of the various experiments combining SG machines together. The set of observations that we see can not be explained by pure classical physics. The fact that particles have quantized values is already unfamiliar enough. Add to that the fact that for some pairs of observables, such as the z and x components of angular momentum, you can't know both observables at the same time. Finally, on top of all of that, you can destroy, but then somehow reconstruct, information about the state of a given observable based on whether there are multiple paths a particle could have followed, and how those paths are put together.

In future chapters, we will explore the mathematical formalism that physicists have developed to model this behavior.

Chapter 5

Dirac Notation

5.1 The State of a System

The *state* of a system means the collection of all properties that that system may have. For example, consider an electron. If you wanted to specify the state of that electron as completely as possible, you'd have to specify where it is and its momentum, as well as how well determined its momentum is. You would also have to specify the state of its angular momentum. Does its z spin have a definite or an indeterminate value? If it has a definite value, what is it? If it has an indeterminate value, is it a half and half chance that, if measured, you'll get +1/2 or -1/2, or is it more likely to be one or the other? If your system includes more than just one particle, you have to include all the information about other particles, as well as any information that arises as a result of the interaction between the particles. For instance, if this electron is moving in some potential, for instance because it's part of an atom, what (effectively) is the electron's potential energy?

We are going to introduce an abstract mathematical notation that will indicate "the state of a particle". The notation itself won't necessarily have all of the information above. However, what it will give us is a way to talk about the state of a particle. Because the state of a particle potentially includes a lot of information, it will be necessary to use a more abstract notation than you're used to for mathematical objects. However, remember that even the seemingly-concrete math that you're comfortable with itself is just constructed from abstract mathematical representations of reality.

Consider algebra. Suppose you have a variable, that may or may not be known. You use the name x to represent the state of that variable. Now, if we're dealing with algebra, and we're dealing with only real numbers, then it's possible to represent the full information about the state of this variable with just a single number. For instance, suppose you are given the following algebraic equation:

$$2x + 5 = 9$$

You could use the rules of manipulating algebraic equations to determine that $x = 2$. At that point, you know everything there is to know about the state of this variable. However, you could still represent it with the letter x if you wished to. Even if you don't, however, and if this equation is supposed to represent something from the real world (say, ages of children in a word problem), even the 2 is a mathematical representation of something in the real world.

Let's make it more abstract. Suppose I tell you that $2x + y = b$, and that y and b are known. However, I haven't given you a number to fully specify y, nor have I given you one for b. I then ask you what x is. You could solve this and tell me that

$$x = \frac{b - y}{2}$$

Now you would say that x is "known", even though you can't reduce it to the concrete representation of real numbers. However, you have given me a representation of x in terms of other things, including this letter y and this other letter b. Those two are stand-ins, abstract mathematical representations of "some number that we've decided to call y" and "some number that we've decided to call b".

We will use a similar abstract notation to represent "the state of a quantum particle", or, perhaps, just "the angular momentum state of a quantum particle" (if we don't care about things like position and momentum). The rules of quantum mechanics will give us mathematical operations we can perform on this representation, and then other things we can do to extract useful information out if it (such as the energy of a particle, or the probability that its z spin will be positive if measured).

5.2 The Ket Vector

To represent the state of a quantum particle, or a quantum system, we introduce the "ket vector"

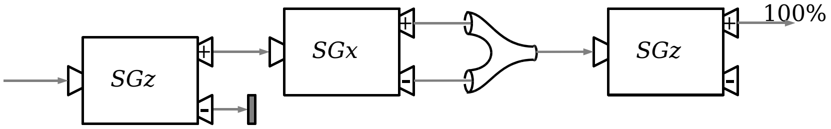
$$|\psi\rangle$$

This is just an abstract mathematical notation, a compact way of saying "the state of this particle". The name "ket" is the latter half of the word "braket", a misspelling of bracket. We will later learn about "bra" vectors, which are written as $\langle\psi|$. However, don't worry about those for now; let's focus on Schrödinger's kets.

Why the ψ inside the ket vector? It's traditional to use the Greek letter "psi" for the state of a system. However, you could put anything you want inside the vertical bar and right angle bracket. It's similar to the convention of using x as the variable for the horizontal axis of a graph in algebra or geometry; people do it a lot, but you can use any letter you want. Sometimes, we will use other Greek letters. Sometimes, we will use something that gives useful information about just what this state is. However, even if we just use something that doesn't tell you anything, like ψ, remember that this ket vector is just a way of representing the state so that we can talk about it, so that we can get a handle on it, and so that we can perform mathematical operations with it.

Why do we call it a vector? This is potentially a source of confusion. This is *not* a vector in there-dimensional space, the way angular momentum, velocity, momentum, or displacement are. In fact, it's a vector in an abstract mathematical space called a "Hilbert space". However, for now, don't worry about that. We will see later the ways in which the state vector behave sort of like three-dimensional vectors like velocity. For now, take it as an idiom that when we talk about the "state vector", we're just talking about a mathematical representation of the state of a quantum particle or a quantum system.

Consider, for example, the sequence of Stern-Gerlach machines shown below:



Consider an electron going into the first SGz machine. If its angular momentum can be oriented in any direction, we would say that the electron is "unpolarized". We don't know the state of the electron, so we will just pick a name for it, and call it $|\psi\rangle$.

Now consider an electron coming out of the first SGz machine. If it is measured to have a z spin of $+\hbar/2$, then it will come out of the + output. At that point, we know the z angular momentum of the electron, so let's wisely choose a representation for the state that will make it easy for us to remember: $|{+z}\rangle$. Note that the "$+z$" inside the notation doesn't mean anything about adding any variable named z, nor have we defined a variable z. It's *just a name.* I could just as easily have named the state vector $|\text{Fred}\rangle$. That would have allowed us to carry it around in equations, talk about it, and perform mathematical operations with it. However, for us humans reading the equations, it's convenient of the name is something that reminds us what we know about the state. So, we'll choose $|{+z}\rangle$ as our name so that we remember that, aha, this is an electron whose z spin is known to be positive.

Similarly, an electron coming out of the $-$ output from the SGz machine will be in the state $|{-z}\rangle$. Again, there's no subtraction, or any multiplying by negative one going on here. It's just a name, the same way x is just a name for a (possibly unknown) variable in algebra.

Now move on to the second SG machine, the SGx machine. An electron going into this machine is in state $|{+z}\rangle$; we know that, because all of these electrons are coming out of the + output of an SGz machine. However, as this electron goes through the SGx machine, its state changes. If it comes out the + output of the SGx machine, it's in a state we shall choose to call $|{+x}\rangle$. If it comes out of the − output of the SGx machine, it's in a state we shall choose to call $|{-x}\rangle$.

Remember what happened in the previous chapter when we then took an electron in state $|{+x}\rangle$— that is, an electron whose x spin was measured to be positive— and put it back into a second SGz machine. It had a 50% chance of being measured with positive z spin, and a 50% chance of being measured with a negative z spin. In other words, the electron is no longer in state $|{+z}\rangle$, nor is it in state $|{-z}\rangle$; the state $|{+x}\rangle$ is different from both of those z states.

It turns out, however, that you *can* describe the state $|{+x}\rangle$ in terms of the states $|{+z}\rangle$ and $|{-z}\rangle$. Remember the algebraic equation $2x + y = b$, which allowed us to figure out that $x = (b - y)/2$. The variables b and y are abstract representations of numbers, and x is an abstract representation of another number. The equation $x = (b - y)/2$ tells us that the thing that x represents is not independent from b and y; were b or y to change, x would have to change along with it. It also tells us how to figure out x in terms of b and y. If we have rules for doing things to b and y, we can then apply those rules to the right side of the equation to figure out how x changes.

Bearing that in mind, it turns out that you can represent the x spin states of an electron in terms of the z spin states as follows:

$$\begin{aligned} |{+x}\rangle &= \tfrac{1}{\sqrt{2}}|{+z}\rangle + \tfrac{1}{\sqrt{2}}|{-z}\rangle \\ |{-x}\rangle &= \tfrac{1}{\sqrt{2}}|{+z}\rangle - \tfrac{1}{\sqrt{2}}|{-z}\rangle \end{aligned}$$

In a future chapter, we will see why this seemingly odd combination, with all of its square roots of two, makes sense. We will also see why being able to represent the x states in terms of the z states is useful. For now, however, recognize this as the first rule about performing mathematical operations on these state vectors: **you can multiply a state vector by a number**. Don't worry about how you would actually calculate something from that. In algebra, you can write down $x/2$, it has meaning even if you don't know x and can't calculate a number for $x/2$. Or, if you're doing algebra with 3-vectors in space, you could write down $\vec{v}_1 = \vec{v}_2/2$. Even if you don't have numbers for all three components of $\vec{v}_2$, and thus can't calculate numbers for all three components of $\vec{v}_1$, this equation is still meaningful. Similarly, it's a valid mathematical operation to multiply a ket vector by a number. For now, we'll just leave it written out as that number multiplied by the ket vector, and bear in mind that when you multiply a number (a *complex* number— it doesn't have to be real!) by a ket vector, you get another ket vector as a result.

You can also add two ket vectors together, and the result of that addition is yet a third ket vector. Here, we see that particular combinations of constants times $|{+z}\rangle$ and $|{-z}\rangle$ turn out to be equal to $|{+x}\rangle$. This is a mathematical expression, within the theory of quantum mechanics, that represents a truth about reality. It's very similar to the Pythagorean Theorem, $a^2 + b^2 = c^2$, which expresses a truth about triangles (if c is a representation of the length of the hypotenuse of the triangle, and a and b are representations of the lengths of the legs of the triangle).

Let's go through the next step of our sequence of devices above. After the SGx machine, the two beams (one in state $|{+x}\rangle$ and one in state $|{-x}\rangle$) are recombined together into a single beam. Let us pretend that we don't know what the state of the electron coming out of the recombining apparatus is. Just as in algebra, when we have a quantity we don't know, we'll give this state a name; let's call it $|\psi_{\mathrm{RC}}\rangle$, with "RC" standing for "recombined". If we knew the all of the rules for figuring out how quantum states evolve as they pass through these SG machines, we could figure out what this state is by performing calculations on the previous states, based on where the beams have been. However, we don't yet know these rules. Instead, what we do is perform one more thought experiment: we send this electron, in state $|\psi_{\mathrm{RC}}\rangle$, through an SG$z$ machine. We discover that 100% of the time, the electron comes out of the + output of the SGz machine. From this experiment, we've figured out that

$$|\psi_{\mathrm{RC}}\rangle \;=\; |{+z}\rangle$$

In upcoming chapters, we'll learn how to figure out theoretically that this is the state of that electron.

Chapter 6

Amplitudes and Probabilities

The classical physics represented by Newton's Laws is deterministic. The equations tell you that if a particle is *here* and its speed is exactly just *this much*, then it *will* be over here and moving this fast later. It gives us a picture of a clockwork universe, where everything future possible measurement is completely determined by the current state of the system.[1]

In quantum physics, as we have seen, this is not the case. If you have an electron whose spin has been measured to be pointing along the $+z$ axis, then the best statement you can make about the x projection of the electron's spin angular momentum is a probabilistic one: there is a 50% chance you'll measure x spin along the $+x$ direction, and a 50% chance you'll measure x spin along the $-x$ direction. What's more, this probabilistic nature is not simply due to our lack of knowledge. Statistics is an entire branch of mathematics used to estimate what we know and determine our confidence in what we know when we have imperfect information. While statistics does apply to quantum mechanics, most of the time statistics is employed in practice the probabilities come not from a fundamental probability, but from lack of perfect knowledge about the state of the system, or because the system itself contains individuals who vary. In quantum mechanics, this probabilistic nature runs more deeply, even though each and every electron is identical. Whereas in classical physics, we may never be able to make perfect measurements, but the theory underneath them is able to presume perfectly determined quantities. In quantum mechanics, the theory needs to be able to handle the calculation and propagation of these probabilities.

---

[1] In fact, chaos theory has shown us that nonlinearities even in classical physics place a limit on the predictability of those systems. However, the laws themselves *are* deterministic.

6.1 Complex Numbers

Before we begin, however, we need briefly to review complex numbers. Complex numbers are intrinsic to quantum mechanics, and indeed the entire theory wouldn't work if we didn't use complex numbers as part of it.

A complex number is a number that has both a real and an "imaginary" part. The name "imaginary" is perhaps unfortunate, because it suggests there's something less tangible about imaginary numbers than there is about real numbers. Remember, however, that even real numbers, when used in science, are abstract mathematical representations of the systems that they are standing in for. Even real numbers are imaginary, in that sense of the word.

All imaginary numbers can be constructed from i, sometimes called "the" imaginary number, which is defined as:

$$i = \sqrt{-1}$$

you may remember from math that you can't take a square root of a negative number. In fact, you can, but you don't get a real number as a result; you get an imaginary number. By the same token, you may remember that the square of any number is positive. That only applies to real numbers; the square of any real number is positive. However, square both sides of the equation above and you can see that:

$$i^2 = -1$$

You can construct any other imaginary number by just multiplying i by a real number. So, $3i$, πi, and $-2.9 \times 10^{21} i$ are all imaginary numbers.

You can then write any complex number as the real part plus the imaginary part. So, $2+3i$ is a complex number. You can't simplify it any further than that. Remember that i is not a variable here, but a number, just as concrete as any other number. It's not a number that you could place on a numberline, because a numberline only has the real numbers on it. But it's just as…well, just as real as a real number. The expression $2 + 3i$ is fundamentally different from the expression $2 + 3\pi$. You can view $2 + 3\pi$ as being completely reduced, as there's no need to reduce it further (as there would be with the expression $2 + (3)(4)$, which can be reduced to 14). However, you could, if you wished, reduce $2 + 3\pi$ with your calculator, and write down an imperfect single-valued representation of it: 11.424778. No such further reduction may be done with the number $2+3i$. The two parts of this number, 2 and $3i$, are like two components of a vector; they both have an independent identity. However, when we get to using vectors to represent spin states of particles don't confuse the real and imaginary parts of a complex number with components of those vectors. The value $2 + 3i$ represents a *single* complex number. You can reduce the real part and the imaginary part of a complex number down so that the first is represented by a single real number, and the second is represented by a second real number multiplied by i.

For every complex number, there is a partner number called the *complex conjugate*. Along with the noun complex conjugate there is a verb complex conjugate. In order to complex conjugate a number, you replace every instance of i with $-i$. So, the complex conjugate of $2 + 3i$ is $2 - 3i$. In algebra, we use the symbol * to indicate the complex conjugate of a quantity. If you have a complex number a (i.e. a variable in algebra that may not just have a real value, but which may have a fully complex value), the complex conjugate of a is represented as a^*. Thus, if $a = 2 + 3i$, then $a^* = 2 - 3i$.

6.2 Amplitudes

What makes quantum mechanics so different from the propagation of uncertainty in classical physics is that it's not directly the probabilities that propagate, but rather these things called *amplitudes*. Suppose you constructed something like a Stern-Gerlach machine, and propagated the system through it using the rules of classical physics. Suppose the path of the particle has two places where there are two possibilities. Suppose that at each of these branches, the probability of each branch is 1/2. That would leave you with four possibilities in the end. The rules of probability tell you that the chance that a particle will take a certain branch at the first choice, *and* a certain branch at the second choice, means that you have to multiply the probability of each branch at each choice. In this example, that would leave you with an overall 1/4 probability of the particle having gone through a given path.

In quantum mechanics, however, the situation may be entirely different. The probabilities you get at the end cannot be simply calculated from the probabilities you would get if you evaluated each choice in isolation.

There is another realm of mechanics where paths taken by the system depend more directly on amplitudes than on probabilities, and that realm is wave mechanics. If two waves pass each other, it's possible to get destructive or constructive interference, possibly giving you wave intensities that are the sum of the two individual intensities, but also possibly giving you wave intensities of zero. Indeed, it is from the amplitudes of waves that quantum mechanics gets the term amplitudes for the thing that it propagates in order ultimately to calculate probabilities. Quantum mechanics bears a lot of similarities to more general wave mechanics, and indeed we often refer to the state of the system $|\psi\rangle$ as the "wave vector" or the "wave function." Although we will not explore this statement in great detail in this course, it is correct to say that in quantum mechanics, particles often (but not always) behave more like waves than like particles. Different quantum states may *interfere* with each other in the same way that waves can interfere with each other. From this interference arises much of the non-intuitive nature of quantum mechanics.

6.2.1 Calculating Probabilities from Amplitudes

Suppose that you have the amplitude A for a particle to be in a given state. Sometimes, this is all you want. You may need to use it to calculate the interference of this state with another state. However, often, what you really want is the probability P for that particle to be found in that given state. You can calculate P by A by taking the *absolute square* of A, written as $|A|^2$. This is different from squaring A, in that you don't multiply the number A by itself, but rather you multiply A by its complex conjugate. So, if A is the amplitude for a particle to be in a given state, then the probability P for that particle to be in that state is:

$$P \;=\; |A|^2 \;=\; A^*A$$

As an example, suppose that you've calculated that the amplitude for a particle in state $|\psi\rangle$ to be subsequently measured to have $+z$ spin (and thus go into the state $|{+z}\rangle$) is $(2+i)/3$. If we wanted to calculate the probability, we'd need to multiply this by its complex conjugate:

$$\begin{array}{rcl} P &=& \left(\frac{2+i}{3}\right)^* \left(\frac{2+i}{3}\right) \\ &=& \left(\frac{2-i}{3}\right)\left(\frac{2+i}{3}\right) \\ &=& \frac{(2-i)(2+i)}{9} \\ &=& \frac{4+2i-2i-i^2}{9} \\ &=& \frac{4+1}{9} \\ &=& \frac{5}{9} \;=\; 0.55555\ldots \end{array}$$

6.3 Bra Vectors and the Inner Product

For each ket vector $|\psi\rangle$, there is a corresponding bra vector $\langle\psi|$. We haven't yet looked into any specific representations of ket vectors beyond just the ket vector itself, so at the moment that's all you need to know. However, when we do get into specific representations, the rules for converting ket vectors to bra vectors are generally very easy. You always take the complex conjugate of any numbers in the representation going from the ket vector to the bra vector. (You may also turn a column vector into a row vector, if you're using column vectors to represent ket vectors; much more about that later.) $\langle\psi|$ is *something like* the complex conjugate of $|\psi\rangle$, although that's not really right. However, just as a number and its complex conjugate are associated with each other, each ket vector $|\psi\rangle$ is uniquely associated with a bra vector $\langle\psi|$.

With the introduction of bra vectors, it becomes possible to define a new operation you can do on these things. You can always stick a bra vector on to a ket vector. The notation is meant to help suggest this; where there is a straight side, you can stick two of them together. The result is called the *inner product.* The specific rules for how you calculate the inner product again depend on the detailed representation of the ket vector, so for now we'll keep them abstract. As an example, suppose you have two different quantum states represented by the ket vector $|\psi\rangle$ and the ket vector $|\phi\rangle$. The bra vector corresponding to the latter is $\langle\phi|$, and the inner product of that bra vector with the ket vector $|\psi\rangle$ is:

$$\langle\phi\,|\,\psi\rangle$$

When you see a bra-ket pair combined like that, the result is a **scalar**! It may well be a complex number, but it is just a number. At that point, you can manipulate it in algebraic equations the way you would manipulate any other complex number.

The inner product of a bra and a ket is the first way we've seen to multiply two of these state vectors together. We've talked about multiplying the state vectors by a scalar, but before we didn't know how to multiply them together. Notice, however, that this is a different sort of multiplication than multiplying two scalars. When you multiply two scalars, you get another scalar out— the same sort of thing as the things you multiplied together. However, when you take the inner product of two state vectors, you get a scalar out, something different from the two things that went into the inner product.

Note that you can only take the inner product between two quantum states if they are the same sort of state. That is, they must be the same kind of state for the same particle or system. For instance, you could take the inner product between two angular momentum states for the same electron, but you couldn't take the inner product between an angular momentum state and a position state.

6.4 Normalization and Orthogonality

Although we aren't yet going to learn rules for doing general inner products between state vectors, there are two cases where the inner product of two state vectors produces a simple answer. The first is not intrinsic to the mathematical representation, but rather something we will insist for state vectors that properly represent real physical states. For a complete state vector $|\psi\rangle$ to be a proper quantum mechanical state, it must satisfy the condition

$$\langle\psi\,|\,\psi\rangle \;=\; 1$$

We say that this means that the state vector is *normalized*. It is possible to have non-normalized state vectors. For instance, in the equation

$$|{+x}\rangle \;=\; \frac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \frac{1}{\sqrt{2}}\,|{-z}\rangle$$

the two parts of the sum on the right side are themselves ket vectors. However, because they are valid state vectors multiplied by a constant, they are not normalized themselves. We will show later that this definition of $|{+x}\rangle$ is, however, normalized.

The second rule is that state vectors that represent different possible states corresponding to different possible measurements of a given observable must be *orthogonal*. Mathematically, this is expressed as:

$$\langle \phi_1 \,|\, \phi_2 \rangle \;=\; 0$$

if $|\phi_1\rangle$ and $|\phi_2\rangle$ are two different states corresponding to definite states for a given observable. For example, the states $|{+z}\rangle$ and $|{-z}\rangle$ correspond to two states of the same observable, specifically, the z component of angular momentum. The first corresponds to that component being measured along $+z$, the second to it being measured along $-z$. The orthogonality condition is then:

$$\langle {+z} \,|\, {-z} \rangle \;=\; 0$$

As an example of doing these calculations with a more complicated state, consider the state $|{+x}\rangle$. If this state is properly normalized, then we should have $\langle {+x} \,|\, {+x} \rangle = 1$. Do we? Well, first, we have to construct the bra vector that goes along with the ket vector:

$$\begin{array}{rcl} |{+x}\rangle & = & \frac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \frac{1}{\sqrt{2}}\,|{-z}\rangle \\ \langle{+x}| & = & \frac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \frac{1}{\sqrt{2}}\,|{-z}\rangle \end{array}$$

Notice that in the case of a compound ket vector, to get the bra vector we just turn all ket vectors on the right side into bra vectors, and replace all the numbers with their complex conjugates (which is trivial here, since all the numbers are real). Now we have what we need to figure out the inner product. Just substitute in our expressions for $|{+x}\rangle$ and $\langle{+x}|$:

$$\begin{array}{rcl} \langle {+x} \,|\, {+x} \rangle & = & \left(\frac{1}{\sqrt{2}}\,\langle{+z}| \;+\; \frac{1}{\sqrt{2}}\,\langle{-z}| \right) \left(\frac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \frac{1}{\sqrt{2}}\,|{-z}\rangle \right) \\ & = & \frac{1}{2}\,\langle {+z} \,|\, {+z} \rangle \;+\; \frac{1}{2}\,\langle {+z} \,|\, {-z} \rangle \;+\; \frac{1}{2}\,\langle {-z} \,|\, {+z} \rangle \;+\; \frac{1}{2}\,\langle {-z} \,|\, {-z} \rangle \end{array}$$

That looks very complicated, but now we can use the orthogonality condition we know is true for the z states, as we've defined them as good states corresponding to the z component of z angular momentum. We know that $\langle {+z} \,|\, {+z} \rangle = 1$, $\langle {-z} \,|\, {-z} \rangle = 1$,

$\langle -z \,|\, +z \rangle = 0$, and $\langle +z \,|\, -z \rangle = 0$ from normalization and orthogonality. Substitute these in:

$$\begin{aligned} \langle +x \,|\, +x \rangle &= \tfrac{1}{2}(1) \;+\; \tfrac{1}{2}(0) \;+\; \tfrac{1}{2}(0) \;+\; \tfrac{1}{2}(1) \\ \langle +x \,|\, +x \rangle &= 1 \end{aligned}$$

So the state is properly normalized! I leave it as an exercise for the alert reader to show that $|{+x}\rangle$ and $|{-x}\rangle$ are orthogonal.

6.5 Interpreting the Inner Product

So far, all we know about the inner product is that for a properly normalized quantum state, the inner product of that state with itself is 1, and that the inner product between two different states corresponding to definite states of the same observable must be zero. But what about the inner product between two arbitrary states? Consider:

$$\langle \phi \,|\, \psi \rangle$$

The interpretation of this is that it is the **amplitude for a particle in state $|\psi\rangle$ to subsequently be measured in state $|\phi\rangle$**. As an example, suppose that we have an electron in the following state:

$$|\psi\rangle \;=\; \frac{3}{5}\,|{+z}\rangle \;+\; \frac{4i}{5}\,|{-z}\rangle$$

Suppose we send this electron through an SGz machine. If this state is properly normalized (is it?), then we could work out the amplitude for it to be measured in the $|{-z}\rangle$ state (i.e. the amplitude for measuring its z-spin to be $-\hbar/2$) as follows:

$$\begin{aligned} \langle -z \,|\, \psi \rangle &= \langle -z|\, \left(\tfrac{3}{5}\,|{+z}\rangle \;+\; \tfrac{4i}{5}\,|{-z}\rangle\right) \\ &= \tfrac{3}{5}\,\langle -z \,|\, +z \rangle \;+\; \tfrac{4i}{5}\,\langle -z \,|\, -z \rangle \\ &= \tfrac{3}{5}(0) \;+\; \tfrac{4i}{5}(1) \\ &= \tfrac{4i}{5} \end{aligned}$$

This tells us the amplitude for the electron to be found in the $|{-z}\rangle$ state. Remember that the *probability*, what we really care about, is the absolute square of the amplitude. That probability is:

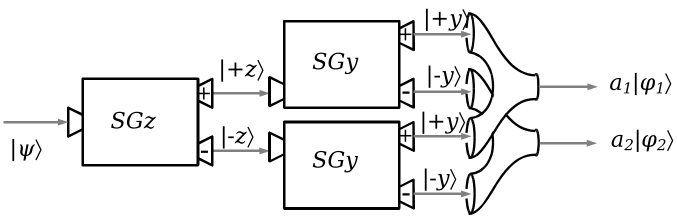
$$\begin{aligned} |\langle -z \,|\, \psi \rangle|^2 &= \langle -z \,|\, \psi \rangle^* \, \langle -z \,|\, \psi \rangle \\ &= \left(\tfrac{-4i}{5}\right)\left(\tfrac{4i}{5}\right) \\ &= \left(\frac{-16\,i^2}{25}\right) \\ &= \tfrac{16}{25} \;=\; 0.64 \end{aligned}$$

If the electron was in the state $|\psi\rangle$ defined above upon entering a SGz machine, there's an 64% chance it will come out the $-z$ output of the machine, being measured with a z-spin of $-\hbar/2$.

6.5.1 Propagating Amplitudes

We have seen that the amplitude for a given quantum state $|\psi\rangle$ to later be found in another quantum state $|\phi\rangle$ is $\langle\phi\,|\,\psi\rangle$. Physically, when would you apply this amplitude? You would apply it when the system went through a device that measured whatever quantity $|\phi\rangle$ is associated with. For example, if you have an electron beam in state $|\psi\rangle$ going into an SGz machine, you'd associate the amplitude $\langle +z\,|\,\psi\rangle$ with the state $|+z\rangle$ emerging from the positive output of the machine. What do you do, however, if the electron beam then goes through another machine? How do you deal with amplitudes when there is more than one process that might have a state change associated with it? The answer is that to get the overall amplitude for a starting state to end up in some final state, you multiply the individual amplitudes of each step the system went through.[2]

As an example, consider the following sequence of SG machines:



Yowza.[3] An electron in some state $|\psi\rangle$ goes into the beginning of this system. There are two possible places it may come out. It may emerge from the upper output in state $|\phi_1\rangle$ (which is currently unknown, but we will figure it out); the amplitude for it to emerge here is a_1. It may also emerge from the lower output in state $|\phi_2\rangle$ (which we will also figure out); the amplitude for it to emerge from the lower output is a_2.

Ultimately, what we're interested in is the amplitude for the electron emerging from this whole thing with state $|+y\rangle$, and the amplitude for it emerging with state $|-y\rangle$. To figure those out, we need to trace the electron through all of the possible

---

[2] This is different from classical physics, where you'd multiply the probabilities. You might wonder what the difference is, since you are going to square the whole thing at the end anyway. The differences comes from the fact that the quantum amplitudes may be *complex*, so the products of the individual amplitudes could end up with terms canceling each other.

[3] The reason why we have this complicated a collection of SG machines and beam combiners is that it's important that we *not be able to figure out* which output from the SGz machine the electron went through, for subtle reasons that will be discussed in the next chapter.

paths. At the input to the first machine, the electron is in the state $|\psi\rangle$. At the positive output of the first machine, the electron is now either in the state $|{+z}\rangle$, with amplitude $\langle{+z}\,|\,\psi\rangle$, or in the state $|{-z}\rangle$, with amplitude $\langle{-z}\,|\,\psi\rangle$.

Let's consider the possible paths for the electron if it comes out of the $+z$ output of the first machine. If the electron goes this way, it will go into the upper SGy machine, with state $|{+z}\rangle$. It will emerge from either the + output, with amplitude $\langle{+y}\,|\,{+z}\rangle$, or from the − output, with amplitude $\langle{-y}\,|\,{+z}\rangle$. The *overall* amplitude for the electron to make it from the very beginning to the + output of the upper second machine is the *product* of the amplitudes for each step: $\langle{+z}\,|\,\psi\rangle\,\langle{+y}\,|\,{+z}\rangle$. Likewise, the overall amplitude for the electron to make it from the very beginning to the − output for the lower second machine is $\langle{+z}\,|\,\psi\rangle\,\langle{+y}\,|\,{+z}\rangle$.

Next, consider the possible path of the electron emerging from the − output of the first machine. The amplitude for it to get this far is $\langle{-z}\,|\,\psi\rangle$. The overall amplitude, then, for it to come out of the + output of the lower machine is $\langle{-z}\,|\,\psi\rangle\,\langle{+y}\,|\,{-z}\rangle$, and the overall amplitude for it to come out of the − output of the lower machine is $\langle{-z}\,|\,\psi\rangle\,\langle{-y}\,|\,{-z}\rangle$.

What do you do at a beam combiner? There, you just add the two states together, each multiplied by their respective amplitudes. Let's first consider the upper beam combiner. The two states coming into this system, with their respective amplitudes, are:

$$\langle{+z}\,|\,\psi\rangle\,\langle{+y}\,|\,{+z}\rangle\,|{+y}\rangle$$

and

$$\langle{-z}\,|\,\psi\rangle\,\langle{+y}\,|\,{-z}\rangle\,|{+y}\rangle\,.$$

Therefore, the final output amplitude and state is:

$$a_1\,|\phi_1\rangle \;=\; \left[\langle{+z}\,|\,\psi\rangle\,\langle{+y}\,|\,{+z}\rangle \;+\; \langle{-z}\,|\,\psi\rangle\,\langle{+y}\,|\,{-z}\rangle\right]\;|{+y}\rangle$$

By looking at this, we can see that the state $|\phi_1\rangle$ is in fact just $|{+y}\rangle$. Hopefully, that does not come as a surprise to you, as the state of the two electron beams going into this beam combiner was just $|{+y}\rangle$. The amplitude a_1 is then just

$$a_1 \;=\; \langle{+z}\,|\,\psi\rangle\,\langle{+y}\,|\,{+z}\rangle \;+\; \langle{-z}\,|\,\psi\rangle\,\langle{+y}\,|\,{-z}\rangle$$

The two state going into the lower beam combiner, with their respective amplitudes, are:

$$\langle{+z}\,|\,\psi\rangle\,\langle{-y}\,|\,{+z}\rangle\,|{-y}\rangle$$

and

$$\langle{-z}\,|\,\psi\rangle\,\langle{-y}\,|\,{-z}\rangle\,|{-y}\rangle\,.$$

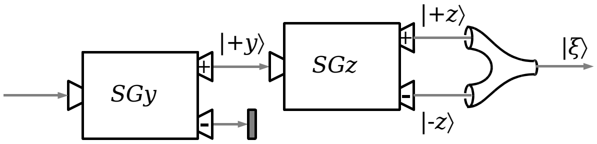
Therefore, the final output amplitude and state for the lower output from this whole system is:

$$a_2\,|\phi_2\rangle \;=\; \left[\langle{+z}\,|\,\psi\rangle\,\langle{-y}\,|\,{+z}\rangle \;+\; \langle{-z}\,|\,\psi\rangle\,\langle{-y}\,|\,{-z}\rangle\right]\;|{-y}\rangle$$

The state $|\phi_2\rangle$ is just $|{-y}\rangle$, and the amplitude for the lower output is:

$$a_2 \;=\; \langle +z\,|\,\psi\rangle\,\langle -y\,|\,+z\rangle \;+\; \langle -z\,|\,\psi\rangle\,\langle -y\,|\,-z\rangle$$

As another example, consider the following collection of SG machines:



We know from the previous chapter that the final state of this system should be $|{+y}\rangle$. Is that what we get?

In order to analyze this, you're going to need to know how to express the states $|{+y}\rangle$ and $|{-y}\rangle$ in terms of $|{+z}\rangle$ and $|{-z}\rangle$:

$$\begin{aligned} |{+y}\rangle &= \tfrac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \tfrac{i}{\sqrt{2}}\,|{-z}\rangle \\ |{-y}\rangle &= \tfrac{i}{\sqrt{2}}\,|{+z}\rangle \;+\; \tfrac{1}{\sqrt{2}}\,|{-z}\rangle \end{aligned}$$

Consider the electron going into the input of the second machine. It is in state $|{+y}\rangle$. We won't worry about the amplitude for the initial electron to get into this state, because we'll just consider the ones that happen to come out the positive output of the first machine. (The purpose of that first machine is to make sure that we know the electrons are in fact in the $|{+y}\rangle$ state when they go into the second machine.) The amplitude for an electron to come out of the upper terminal is $\langle +z\,|\,+y\rangle$, and the amplitude for an electron to come out of the lower terminal is $\langle -z\,|\,+y\rangle$. Call the final state coming out of the beam combiner $|\xi\rangle$. To figure out what this state is, combine together the two states going into it, each multiplied by their respective amplitudes:

$$\begin{aligned} |\xi\rangle &= \langle +z\,|\,+y\rangle\,|{+z}\rangle \;+\; \langle -z\,|\,+y\rangle\,|{-z}\rangle \\ &= \langle +z|\,\left(\tfrac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \tfrac{i}{\sqrt{2}}\,|{-z}\rangle\right)|{+z}\rangle \;+\; \langle -z|\,\left(\tfrac{1}{\sqrt{2}}\,|{+z}\rangle + \tfrac{i}{\sqrt{2}}\,|{-z}\rangle\right)|{-z}\rangle \\ &= \left(\tfrac{1}{\sqrt{2}}\,\langle +z\,|\,+z\rangle \;+\; \tfrac{i}{\sqrt{2}}\,\langle +z\,|\,-z\rangle\right)|{+z}\rangle \;+\; \left(\tfrac{1}{\sqrt{2}}\,\langle -z\,|\,+z\rangle \;+\; \tfrac{i}{\sqrt{2}}\,\langle -z\,|\,-z\rangle\right)|{-z}\rangle \end{aligned}$$

Once again, we just have inner products of z states with themselves. We can use normalization (e.g. $\langle +z\,|\,+z\rangle = 1$) and orthogonality (e.g. $\langle -z\,|\,+z\rangle = 0$) to substitute in the numbers from the inner products in the expression above, yielding us:

$$\begin{aligned} |\xi\rangle &= \left(\tfrac{1}{\sqrt{2}}(1) \;+\; \tfrac{i}{\sqrt{2}}(0)\right)|{+z}\rangle \;+\; \left(\tfrac{1}{\sqrt{2}}(0) \;+\; \tfrac{i}{\sqrt{2}}(1)\right)|{-z}\rangle \\ &= \tfrac{1}{\sqrt{2}}\,|{+z}\rangle \;+\; \tfrac{i}{\sqrt{2}}\,|{-z}\rangle \\ &= |{+y}\rangle \end{aligned}$$

Sure enough, the mathematical rules for propagating amplitudes has given us the answer that we know is supposed to be right for the final state.

Chapter 7

The Collapse of the Wave Function

At this point, it's worth taking a step back and reviewing where we are. We started with some observations about how electron spins function, and how it's very different from what you'd expect for little spinning balls operating under the laws of classical physics. These observations are:

- Every single electron has exactly the same total angular momentum $(\sqrt{3}/2)\hbar$. In contrast, classical spinning balls can be spinning at pretty much any rate (limited only by the speed of light for very fast rotation rates).

- Every time you measure the component of angular momentum along a given axis (for example, the z axis), you get one of only two values: $+\hbar/2$ and $-\hbar/2$. This is different from classical spinning balls in that even if they all have exactly the same rate of rotation, you could still orient them so that the z component of angular momentum is anything between the total (if the angular momentum is pointing in the $+z$ direction), on down to 0 (if the angular momentum is pointing in the $x-y$ plane), on down to minus the total (if the angular momentum is in the $-z$ direction).

- You can only know one component of angular momentum at a time. That is, if you measure the z spin of an electron and it comes out $+\hbar/2$, next time you measure it you will still get $+\hbar/2$. If you then measure x spin, you have an even chance of $+\hbar/2$ or $-\hbar/2$. This may not be surprising, as you hadn't measured the x component yet, so you didn't now anything about it. However, after measuring the x spin, if you go on to measure the z spin again, you have an even chance of measuring $+\hbar/2$ or $-\hbar/2$. Electrons *can not have a definite angular component of angular momentum along more than one axis at a time.* From a classical point of view, this is extremely bizarre. In classical physics, angular momentum is a vector. Thus, a spinning ball has an x, a y, and a z

component of angular momentum, and you can in principle measure all of them at once.

In order to explain this observed behavior, we've been constructing a mathematical model that operates on state vectors. We've been using Dirac notation, with objects written similar to $|\psi\rangle$ used to represent this state vector. These are abstract mathematical objects, different from algebraic variables, different from (but with some similarities to) vectors in 3-d space.

7.1 Summary of Rules for Manipulating Ket Vectors

As with algebraic variables or vectors in 3-d space, there are rules for manipulating ket vectors. It's important to remember that these rules exist, and that they are specific to ket vectors. Some of them look and behave exactly like the rules for manipulating algebraic objects, and indeed, you use ket vectors in algebraic equations. However, this does not mean that you can do everything with a ket vector that you can do with algebraic objects. For instance, there is no way you can *divide* by a ket vector; that's just not a defined operation. Also, multiplication with ket vectors does not match terribly well to the algebraic counterpart, except when you're multiplying a ket vector by a scalar (i.e. something that represents just a plain complex number).

The two most basic things you can do with a ket vector are summing them together and multiplying them by a scalar. These are also things that you can do with vectors in 3-d space, or with any other vector for that matter. If you multiply a ket vector by a constant, you get another ket vector. if you add together two ket vectors, you get a third ket vector. This rule can be summarized by:

$$|\xi\rangle \;=\; a\,|\psi\rangle \;+\; b\,|\phi\rangle$$

where $|\xi\rangle$, $|\psi\rangle$, and $|\phi\rangle$ are all state vectors, and a and b are scalars (i.e. something that could just be a complex number). All of the usual rules for scalars still apply to multiplying *scalars*. Thus, for example, you can use the distributive property, $a(|\psi\rangle + |\phi\rangle) = a\,|\psi\rangle + a\,|\phi\rangle$.

You can turn any ket vector $|\psi\rangle$ into a corresponding bra vector $\langle\psi|$. The detailed rules for how you do that will depend on how you represent a ket vector. In general, if a ket vector is built from other ket vectors

$$|\xi\rangle \;=\; a\,|\psi\rangle \;+\; b\,|\phi\rangle$$

then the corresponding bra vector is

$$\langle\xi| \;=\; a^*\,\langle\psi| \;+\; b^*\,\langle\phi|$$

where a^* represents the complex conjugate of a (i.e. replace all instances of i with $-i$). Notice what we did here on the right: replace all scalars with their complex conjugates, and replace all ket vectors with their corresponding bra vectors.

You can take an *inner product* between a bra vector $\langle\phi|$ and a ket vector $|\psi\rangle$, which is notated as

$$\langle\phi\,|\,\psi\rangle$$

The notation is meant to suggest this; you can "stick these vectors together on their straight sides." The result of this inner product is a *scalar*. While you can't divide by a ket vector, if you have something that's closed off (i.e. an inner product), it becomes just a scalar, so you can do anything with it in equations that you can do with scalars (including divide by it).

The meaning of an $=$ sign is the same as always. That means that if you have, for example, $|\psi\rangle$ in one expression, and you have an equation that sets $|\psi\rangle$ equal to something else, you can substitute what $|\psi\rangle$ is equal to back into the first expression. You will usually want to make sure to put parentheses around what you're substituting in, to make sure that you don't (for instance) multiply by just a *piece* of $|\psi\rangle$ when you mean to multiply by all of $|\psi\rangle$. As an example, suppose you wanted to evaluate $\langle\psi\,|\,{+z}\rangle$, and you know:

$$|\psi\rangle \;=\; \frac{i}{\sqrt{3}}\,|{+z}\rangle \;+\; \sqrt{\frac{2}{3}}\,|{-z}\rangle$$

Well, first, we know how to build $\langle\psi|$

$$\langle\psi| \;=\; -\frac{i}{\sqrt{3}}\,\langle{+z}| \;+\; \sqrt{\frac{2}{3}}\,\langle{-z}|$$

and now we can substitute that into $\langle\psi\,|\,{+z}\rangle$:

$$\langle\psi\,|\,{+z}\rangle \;=\; \left(-\frac{i}{\sqrt{3}}\,\langle{+z}| \;+\; \sqrt{\frac{2}{3}}\,\langle{-z}|\right)\,|{+z}\rangle$$

If we wanted to reduce this further, we could distribute the $|{+z}\rangle$ to the left through the parentheses, and then substitute the known results $\langle{+z}\,|\,{+z}\rangle = 1$ and $\langle{-z}\,|\,{+z} = 0\rangle$ to get out just a single number.

There is one important thing to realize about $=$ signs, however: an equation is only meaningful if you have the same types of objects on both sides of the equation. You've actually seen this before, with dimensionalities. It doesn't make sense to set a certain number of meters equal to another numbers of kilograms. Meters and kilograms are different sorts of things (one is length, the other is mass), and so they can't be equal. Similarly, you can't set different kinds of mathematical objects equal to each other. Bras, kets, and scalars are all different kinds of mathematical objects. You can't add a scalar to a ket vector, and you can't set a ket vector equal to a scalar. Nor can you set a ket vector to a bra vector. (Remember, however, that a scalar times a ket vector is a ket vector, and so forth. Thus, you *can* set a ket vector equal to a scalar times another ket vector, because the latter is just a ket vector itself. This is similar to saying that you can set a speed equal to a number of meters divided by a number of seconds, because when you divide length by time you get speed.)

One thing that you can **not** do with inner products is change the order of them. The *commutative* property of multiplication applies to scalars, but does not apply necessarily in general to other kinds of mathematical objects. So, while $ab = ba$, it's important to remember that $\langle \psi \,|\, \phi \rangle \neq \langle \phi \,|\, \psi \rangle$. (In fact, it turns out here that $\langle \psi \,|\, \phi \rangle = \langle \phi \,|\, \psi \rangle^*$. That works for bras and kets, but also is not going to be generally true for other mathematical objects.) You *can* change the order when you're multiplying by a scalar, however. Therefore, if you have:

$$\langle \psi | \frac{1}{\sqrt{2}} | \phi \rangle$$

it is the same as

$$\frac{1}{\sqrt{2}} \langle \psi \,|\, \phi \rangle \,.$$

Here, we didn't reorder any bras and kets; we just moved a scalar around.

For any given set of ket vectors (e.g. the set of all ket vectors that could potentially represent an electron spin state), you can identify a set of *basis vectors* from which all the other vectors can be built. For vectors in 3d space, the unit vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$ form the basis vectors. For electron spins, $|{+z}\rangle$ and $|{-z}\rangle$ form the basis vectors. These two basis vectors represent, respectively, a particle whose z component of angular momentum is $+\hbar/2$ and a particle whose z component of angular momentum is $-\hbar/2$.

7.1.1 Calculating Experimental Predictions

Because quantum mechanics is stochastic rather than deterministic, often the results we expect from our theoretical calculations are probabilities of certain observations. We interpret the inner product

$$\langle \phi \,|\, \psi \rangle$$

as the *amplitude* (sometimes called "probability amplitude" to distinguish it from other sorts of amplitudes) for a particle in state $|\psi\rangle$ to be found in state $|\phi\rangle$ given a measurement of the observable for which $|\phi\rangle$ is a definite state. To calculate the probability, you take the absolute square of the amplitude, i.e.:

$$Pr \;=\; |\langle \phi \,|\, \psi \rangle|^2 \;=\; \langle \phi \,|\, \psi \rangle^* \, \langle \phi \,|\, \psi \rangle$$

If you want to calculate the overall probability for a particle to go through two different subsets of a path (e.g. if that particle is going through two different SG machines), you multiply the *amplitudes* for each subset of the path to get the overall amplitude for that path. When two possible paths for a particle to have traversed are combined together, you add the states of the particle at the end of each path, multiplied by their respective amplitudes. You only take the absolute square of amplitudes to find a probability when an actual measurement is made. More about that in Section 7.2.1.

We say that a ket vector describing a quantum state is *properly normalized* if $\langle\psi\,|\,\psi\rangle = 1$. Additionally, we generally want to choose basis states that are orthogonal, i.e $\langle +z\,|\,-z\rangle = 0$. If you square these two amplitudes to get probabilities, you see that this makes sense with the interpretation. If the electron is in state $|\psi\rangle$, the probability of finding it in state $|\psi\rangle$ is obviously 1. If the electron is in state $|-z\rangle$, then the probability of subsequently finding it in state $|+z\rangle$ is 0.

7.2 The "Collapse" Rule

There is another important more rule for manipulation of ket vectors in order to represent quantum systems. We've been using this all along, but haven't explicitly identified it yet. That rule is that when a *measurement* is made of an observable, the state of the system being measured *changes* to become a state that corresponds to a definite value of that observable. *Which* value of that observable the state adopts is random. It will be one of the ones that are possible, and quantum mechanics allows us to calculate the probabilities for each state to be adopted, but it does not allow us to predict with certainty which specific state the system will fall into. For example, if an electron is in state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\,|+z\rangle\,+\,\frac{1}{\sqrt{2}}\,|-z\rangle$$

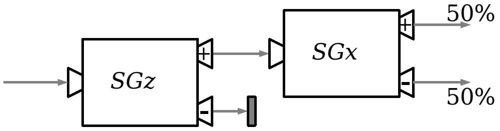
then after a measurement of the z component of angular momentum, it will either switch into the state $|+z\rangle$ or $|-z\rangle$. The *amplitude* for it to switch into $|+z\rangle$ is $\langle +z\,|\,\psi\rangle$ (in this case, $\frac{1}{\sqrt{2}}$), and thus the probability is $|\langle +z\,|\,\psi\rangle|^2$ (in this case, $\frac{1}{2}$).

This process of the state of a particle changing from an indeterminate state into a state that has a definite value for a given observable is often described as "collapse". The state vector of the system "collapses" to one of the definite state for that observable. Sometimes (although not always) you can use a function (e.g. a function of position $[x,\,y,\,z]$) to represent a quantum state $|\psi\rangle$. In quantum mechanics, these functions are called "wave functions" because the equations that govern their evolution are very similar to standard wave equations. As such, you will hear the term "the collapse of the wave function" to describe what happens to a quantum state when a measurement is made on it.

7.2.1 What is a measurement?

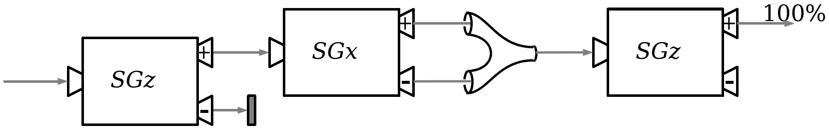
When you start to dig into exactly how to decide what happens with a particle going through various parts of a system, it turns out that it's not exactly obvious what it means to "make a measurement." This is the source of a lot of debate within the scientific community, and has led to various different (often bizarre) interpretations of quantum mechanics. It's also the source of a lot of the dubious and downright wrong things that are said about quantum mechanics, including much of "quantum mysticism". As such, it's worth putting some thought into the measurement problem.

Consider the following sequence of Stern-Gerlach machines:



The particle going into the second machine is in state $|+z\rangle$. While this is a definite state for the z component of angular momentum, it's *not* a definite state for the x component of angular momentum, which is what the second SG machine measures. Thus, we would say that upon the making of the measurement the quantum state collapses into either $|+x\rangle$ or $|-x\rangle$, each with a probability of 0.5. Evidently, the SGx machine has performed this process of "making a measurement," whatever that is.

However, now consider this sequence of SG machines:



In the previous example, when the electron goes through the SGx machine, its state changes. It changes into either $|+x\rangle$ or $|-x\rangle$. Both of those x states do not have a determined value of the z spin. In both cases, if you subsequently measure the z spin, you find an 0.5 probability for measuring the z spin as either positive or negative. Because that's true for both $|+x\rangle$ and $|-x\rangle$, you would then expect that if you combined paths together each that represented one of those two states, you'd still have a 0.5 probability of either $|+z\rangle$ or $|-z\rangle$. That is, if the electron follows the top path, it collapses into state $|+x\rangle$, and thus you'd think it has a 50/50 chance of being measured with either positive or negative z-spin. Likewise, if it follows the bottom path, it collapses into state $|-x\rangle$, and thus you'd think it has a 50/50 chance of being measured with either positive or negative z-spin. However, that's not what's observed! Somehow, when the two paths are recombined, the state $|+z\rangle$ that the electron was in before it entered the second SG machine is reconstructed. We have said previously that when you make a measurement, the state of the system changes. We have also said that x spin and z spin can't be known at the same time, so if you've measured x spin, the particle can't have a definite value of z spin. And, we saw previously that evidently an SGx machine makes this measurement of x spin, because you can figure out the x spin of the electron by seeing which output of the SGx machine the electron emerges from. Yet, here, it looks like a measurement of x spin wasn't made after all!

So how do you know what to do? Does the state vector collapse, or doesn't it? How do you know if you've made a measurement?

The mathematics of quantum mechanics are clear. Despite the interpretational difficulty, it's very important to realize that the *predictive power* of quantum mechanics is strong. If you followed the rules for propagating amplitudes through the series of *SG* machines above, what you'd find is that all of the amplitudes on the $|-z\rangle$ parts of the x states coming out of the SGx machine subtract out when you combine the two beams at the recombiner. So, while there is an *interpretive* mystery as to exactly what's going on here, there's no mystery as to what the result of either set of SG machines is. Too many of those who want to argue for some form of quantum mysticism seem to lose track of this distinction. Too many seem to say that because of debates about the interpretation of quantum mechanics, there are debates about what quantum mechanics says can happen. While you may find some *qualitative* and *interpretive* similarities between some sorts of radical post-modernist philosophies and the uncertain interpretations of quantum mechanics, it's simply wrong to say or imply that quantum mechanics tells us that we can't know the results of experiments. Those results may be *probabilities*, but even in that case they are rigid probabilities defined by the nature of physical reality. It's incorrect to claim that quantum mechanics points to a physical reality that doesn't fully exist without our own perception of it, and that reality itself can somehow be a "social construction". Rather, the success of quantum mechanics simply tells us that on the smallest scales, physical reality is simply something that is deeply unintuitive to us with our brains that evolved to deal with huge numbers of atoms at one time, where the laws of quantum mechanics in bulk give rise to the much more deterministic laws of classical physics.

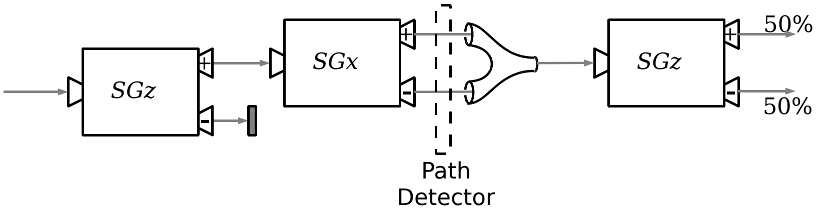
Cautions aside, let's return to the interpretive difficulties that this whole notion of collapse gives us. The physical observation is that we experience the world in definite states. Sure, there is always *measurement* uncertainty, meaning that we don't know things perfectly. (For example, when you measure your height, do you know it to the millimeter? To the micrometer? If your ruler is marked to centimeters, and perhaps millimeters, you probably haven't measured your height to better than a few millimeters.) But while we do measure things with experimental uncertainty, we *never* directly observe something to be both something to be two different ways at the same time, we never see this "mixture of states" that quantum mechanics tells us particles can be in.[1] Our devices that measure spin will measure a value of angular momentum, with some experimental uncertainty, they will not return a "maybe" for two different discrete values for the measurement on a given particle. It is from this observation that we know *something* like the collapse of the state vector happens. But what is it that makes something into a measurement? The answer is not obvious, and has led to various different interpretations of quantum mechanics (see Section 7.3). Some even suggest that it must require a conscious observer to cause this collapse. After all, the argument goes, if it is *our* experience of the world that tells us that things are found in definite states for observables, then it must be something in *our act of observation* that causes the wave function to collapse. The unthinking SGx machine in the example above wasn't able to fully and irrevocably collapse the state vector; however, if an experimentalist looks at the output of the SGx machine, if that experimentalist figures out which output an electron came out of, then the state vector does change.

Many physicists, however, are very uncomfortable with requiring a conscious observer to change the state of the system, for it is not obvious exactly what "consciousness" is in this context. Indeed, modern neuroscience models all the thought processes of our brain as the material interaction of atoms and ions in our neurons, which are themselves ultimately governed by the laws of quantum mechanics. Where, and how, then, does this different "consciousness" manage to arise? Or is it an illusion, something that looks like it's there to us the same way a liquid appears to have a temperature even though what we call temperature is really just a measurement of the average speed at which the molecules in that liquid are bouncing about? The whole notion of temperature doesn't then feed back and somehow affect the molecules in ways that couldn't be derived from the laws governing the direct molecular interaction. If "consciousness", whatever that is, arises just as a property of a whole lot of neurons working together, then there is nothing particularly special there from a physics point of view that could somehow cause the collapse of the wave vector. So how does it happen? The question is not fully answered.

It is worth revising the SG machine one more time. Consider the sequence of SG machines we looked at last, but add one wrinkle. We aren't going to *capture* the electron out of the second machine; we'll let it go on into the recombiner unhindered. However, we are going to put some sort of detector that allows us to figure out *which* of the two outputs of the second machine the electron came out of before it goes into the beam recombiner.

---

[1] Greg Egan's science fiction novel *Quarantine* plays with the notion that there might be creatures who can, somehow, directly perceive these mixtures of states.



In this case, if you *could* tell which of the two outputs of the SGx machine the electron emerged from, you do *not* reconstruct the $|{+z}\rangle$ state after the recombiner. Without the path detector, effectively a measurement has not been made. If there is no way that you could know which path the electron went through, effectively it goes sort of goes through both, and the two paths *interfere*. Mathematically, what happens is that when you combine together the amplitudes, the amplitudes on the $|{-z}\rangle$ state from the expansion of the $|{+x}\rangle$ and $|{-x}\rangle$ states cancel each other out, just leaving you with $|{+z}\rangle$. But, if the path detector is there, if somewhere data is recorded that somebody *could* look at and see which path the electron took, then the electron takes *only* that path.

7.2.2 Schrödinger's Cat

In the sequence of SG machines without the path detector, in a very real sense each individual electron goes through *both* the $+x$ and $-x$ paths in the SGx machine. This is one example of a particle acting more like a wave. You can divide a wave up and recombine it (adding together amplitudes), but a particle is either *here* or *there*. The notion that particles like electrons could somehow follow both paths if you don't measure exactly which path it takes— even though, if you do measure the path, you never see that happen— seems absurd. Yet, that's what quantum mechanics tells us happens, and the predictions of quantum mechanics have been confirmed by countless experiments.

Schrödinger's cat is a thought experiment that tries to point out the absurdity of what quantum mechanics seems to be saying. Put a cat in a closed box. With the cat, put in a single radioactive nucleus, that is attached to a thread holding a hammer over a vial of poison. If the radioactive nucleus decays, the thread will break, releasing the hammer, breaking the vial, releasing the poison, and killing the cat. (Poor kitty!)

Put the cat and everything else in the box. Then wait enough time that there is a 50% chance that the radioactive nucleus has decayed. Is the cat still alive or is it dead? You don't know, because it's completely random exactly when any individual radioactive nucleus will decay. You can predict probabilities, but you can't predict anything about an individual decay. Indeed, we've now seen that if you don't make the measurement as to whether the nucleus is still there or not, in a very real sense it's neither decayed nor undecayed, but just like an electron whose z spin hasn't been measured, it's in an indeterminate quantum state, in a state that is a mixture of the decayed state and the undecayed state.

However, whether or not the state is decayed determines whether or not the cat is still alive. Is the cat still alive? Or is it dead? It's not just that you don't know, the argument goes. In fact, the cat is in a sense *both* alive and dead. It's in an indeterminate state. The jargon we use is to say that the cat's state has become entangled with the radioactive nucleus' state, since whether or not the nucleus has decayed determines whether the cat has died. But the cat doesn't take on a determinate state, being either just alive or just dead, until you open the box and make the observation to find out whether it's alive or dead.

Most physicists would argue that in reality a cat would function as an observer, and as such the cat makes the "observation" of the nucleus' decay by dying (or by its failure to decay by staying alive). Indeed, the vial of poison is itself a macroscopic enough system that once the radioactive nucleus becomes entangled with the states of the huge numbers of particles in the vial, wave function collapse has already happened; you don't even need the cat. However, Schrödinger's cat remains as a thought experiment that points out the very non-intuitive nature of quantum measurement and quantum mixtures of states.

The largest objects for which quantum interference has been directly observed is C-60 molecules, or buckyballs (Arndt et al., 1999). Physicists refer to hypothetical states where the interference of amplitudes of quantum states for macroscopic objects can be observed as "Schrödinger's Cat States." While these show up in science fiction (such as in some stories by Greg Bear and Greg Egan), they have yet to be observed in reality.

7.3 Interpretations of Quantum Mechanics

Does the wave function *really* collapse? What does it mean to say that? And what really is a measurement? Quantum mechanics is a great physical theory. It gives us a mathematical model that allows us to predict results for a wide range of experiments. It explains phenomena that could not be explained with classical physics. Quantum mechanics explains the structure of chemistry's periodic table of the elements. Practically speaking, it provided the understanding of nature that allowed us to develop, among other things, the laser and the solid-state transistor. All of today's digital technology is based on an understanding of semiconductors given to us by quantum mechanics. It is a tremendous misrepresentation of quantum mechanics to say that it brings mysticism into science, to say that it shows us that nothing is real and nothing is tangible or definite. The reality of today's human society would bear absolutely no resemblance to what we all know without the reality of quantum mechanics and our understanding of it.

However, while quantum mechanics provides us for clear rules for manipulating its mathematical model of atomic and subatomic reality, some questions it leaves unanswered have led to numerous "interpretations" of quantum mechanics. By and large, these interpretations struggle with the measurement problem. Practically speaking, the measurement problem is not a serious problem. We know when we've made a measurement. Whether or not consciousness is really involved (something, again, that most physicists are extremely uncomfortable including in their models), we're able to design things based on quantum mechanics with the knowledge that once macroscopic things are affected by the results of quantum processes, measurements effectively have been made. However, if you want to understand what it really means, what quantum mechanics is saying about the nature of reality, then you have to grapple with the various different interpretations.

These interpretations include the standard Copenhagen interpretation, which says the wave function does in fact collapse. Practically speaking, however, most physicists go through their days behaving as they accepted the instrumentalist interpretation, which N. David Mermin summarized as "shut up and calculate!" (citation needed). This is the interpretation described above: practically speaking, we know ☑ how it works. So, just accept that it's a mathematical model that is useful and don't worry too much about what it means beyond what it tells you about the outcome of experiments.

A second variety of the instrumentalist interpretation is the statistical interpretation. This interpretation is based on the fact that in order to actually measure real probabilities, you have to perform experiments a large number of times. Otherwise, the statistics of counting random events tells you that you cannot make all but the roughest estimates of what your experiment tells you those probabilities are. In the statistical interpretation, quantum mechanics ultimately only talks about ensembles, groups of particles in enough numbers that you could compare the results of experiments to the predictions of quantum mechanics. In this interpretation, it's over-interpreting the theory to talk about what it says about the behavior of individual particles. The author of this text thinks that the statistical interpretation doesn't hold water. In the various quantum systems where different paths interfere and give us results that would be surprising to a classical physicist, yes, it's true that we can't practically compare those results to the numerical predictions of quantum mechanics until we've put multiple particles through the system. However, quantum interference happens even if you send only one particle *at a time* through the system; therefore, individual particles do in some way interfere *with themselves.*

Perhaps the most interesting interpretation of quantum mechanics is the Many Worlds interpretation, sometimes called (in an attempt, perhaps, to make it sound less outlandish) the "relative states" formulation of quantum mechanics. Before an electron spin is measured, it's in a state that is in a sense half up and half down. When you measure that state, you see the electron as being (say) spin up. What happened to the spin downness? In the Copenhagen interpretation, it's just gone; the wave function has collapsed. In the Many Worlds interpretation, the universe splits, and thereafter there are two universes. In one, you measured spin up; in the other, you measured spin down. Every time a measurement of a quantum system happens that requires that system to take on a definite value, and there are multiple possibilities for that value, the universe splits, one universe for each possibility of the value.

An interpretation that has been gaining a lot of favor recently is decoherence (Schlosshauer, 2004). As quantum particles interact with other quantum particles, there states become entangled. In reality, it's difficult (or impossible) to so isolate a system that you can do much for long without that system interacting, and thus having its state become entangled with other systems. Indeed, the act of measurement itself represents a quantum state becoming entangled with the state of the measuring device (or, perhaps more properly, with the quantum states of all of the particles in the measuring device). The decoherence idea states that as particles become entangled with more and more other particles— effectively, as the system becomes more and more macroscopic— interference terms become highly suppressed, leading to the practical appearance of wave function collapse. However, while decoherence indisputably happens, it's not clear that the decoherence paradigm actually addresses the measurement problem or not. (Only Schrödinger's cat can probably answer for sure!)

As you consider interpretations of quantum mechanics, it is important to remember that none of the valid interpretations of quantum mechanics lead to quantum mysticism. Much quantum mysticism— unfortunate parts of popular culture such as the movie *What The #$*! Do We Know?* or the book *The Secret*— is based on a misreading of the measurement problem. Two things are true: first, that quantum particles can be in a mixture of states, where multiple outcomes are possible and consistent with the laws of physics. Second, when *we* make a measurement, somehow that act of measurement causes one of the outcomes to be realized. This leads many people to conclude that we are affecting the state of the system, and that therefore somehow we can influence these outcomes. This is not the case. The probabilities for the outcomes are rigidly dictated by the probabilities that you can calculate from the mathematical model that we call quantum mechanics. Countless experiments have given us extremely good confidence that this is a good mathematical model. Nowhere in that model is there anything that allows the observer to *influence* or *choose* which particular outcome will be observed when an experiment with multiple probable outcomes is observed. Nowhere has a valid, reproducible quantum experiment been performed to demonstrate this effect (despite what you will hear in things such as the aforementioned movie).

Chapter 8

Operators and Eigenvectors

8.1 Operators

In quantum mechanics, we associate observables with *operators*. There is a position operator, an energy operator, a spin-z operator, etc. The exact form of the operators and the rules for how you work with them vary depending not only on what the operator is, but on the representation we're using for the type of states that the operator operates on. In this chapter, we're concerned with operators in general, so we're going to keep things abstract. In the next two chapters, we'll see how actually to do calculations with the operators and ket vectors that represent angular momentum states.

On the most level, an operator is just something that does something to something else. Whether you know it or not, you're already familiar with a wide variety of operators in regular mathematics. For example, the addition operator, usually denoted with the + symbol, is a *binary operator*. It takes two arguments, and returns a third argument of the same type. For example, $3+2$ is the notation we use to apply the addition operator to the numbers 3 and 2. Our understanding of the rules for applying this operator tells us that the result is 5.

The operators we're going to use in quantum mechanics just operate on one thing, a quantum state vector. An analogy to arithmetic would be the negation operator. If you see $-a$ written, you recognize that whatever number a represents, $-a$ is the negative of that number. The $-$ sign just denotes that one should apply the operator that performs this negation to the number a. Another operator you're familiar with is square root. When you see $\sqrt{4}$, that means that you should apply the square root operator to the number 4; the result in this case is 2.

Here, we will follow the convention of putting a "hat" on top of letters that represent quantum operators. So, $\hat{\mathcal{O}}$ would be a particular operator. Operators operate on ket vectors[1], and the result of the operation is another ket vector. So, if we write:

$$|\phi\rangle \;=\; \hat{\mathcal{O}}\;|\psi\rangle$$

then we're saying that when you apply the $\hat{\mathcal{O}}$ operator to the state $|\psi\rangle$, the result is the state $|\phi\rangle$.

8.2 Eigenstates

An *eigenstate* is a state that corresponds to some observable having a definite value. So, $|{+z}\rangle$ and $|{-z}\rangle$ are eigenstates of the z component of angular momentum. Likewise, $|{+y}\rangle$ and $|{-y}\rangle$ are eigenstate of the y component of angular momentum. Quantum systems do not necessarily have to be in an eigenstate of anything. However, the act of measurement causes a quantum state to "collapse" to an eigenstate of the observable that was measured.

If a quantum state is an eigenstate of a given observable, then we can describe the state vector that represents that state as an *eigenvector* of the operator corresponding to that vector. For example, suppose that $\hat{S}_z$ is the spin-z operator. Then, $|{+z}\rangle$ is an eigenvector of $\hat{S}_z$. (We will also later talk about *eigenfunctions*. This is when you are using mathematical functions as a way of representing quantum states. Before we get to that, we will use column vectors to represents $|\psi\rangle$, so there will be no need to invent a term beyond eigenvector.)

Eigenvectors of an operator have a special relationship with that operator. If a given vector is an eigenvector of a certain operator, then the following applies:

$$\hat{\mathcal{O}}\;|\psi\rangle \;=\; v\;|\psi\rangle$$

where v is just a scalar constant. In other words, when an operator operates on one of its own eigenvectors, the result of that operation is the same vector that went into it, only multiplied by a constant. In quantum mechanics, there is an additional constraint: these eigenvalues must be *real* (i.e. their imaginary part must be zero). What's more, there's a very direct interpretation to them: the are the measured values that corresponds to the respective eigenstate. So, for the spin-z operator:

$$\hat{S}_z\;|{+z}\rangle \;=\; \frac{\hbar}{2}\;|{+z}\rangle$$

---

[1] They can also operate on bra vectors, but there are some wrinkles to how they work. In the standard notation, an operator always has to operate through one of the straight sides of a vector, be it bra or ket. For our present purposes, we'll only allow operators to operate to the right on ket vectors.

$$\hat{S}_z \left|-z\right\rangle \;=\; \frac{-\hbar}{2} \left|-z\right\rangle$$

This knowledge right here will allow us to do calculations with operators and ket vectors, even if we don't actually know the detailed mechanics for just what it is that the operator is doing, nor what the detailed mathematical representation we're using for our ket vectors. Just as with the probability calculations we've done before, even though we're keeping ket vectors abstract, we can still get results out of them.

The different eigenstates of a given operator are orthogonal. This was already discussed in Section 6.4, although there we hadn't yet introduced the concept of operators. There, we said that the different states corresponding to definite states for a given observable were orthogonal. Now, we know that another way to say "the states corresponding to the definite states for a given observable" is "the eigenstates for the observable's operator". This means that if you have an operator with eigenstates $\left|\phi_n\right\rangle$, then

$$\left\langle \phi_n \,\middle|\, \phi_m \right\rangle = 0 \qquad \text{if } n \neq m.$$

Also, the eigenstates for a given operator need to be normalized, that is,

$$\left\langle \phi_n \,\middle|\, \phi_n \right\rangle = 1.$$

The set of eigenstates for a given operator form a set of "basis states". If you put all of them together, you can express *any* state in terms of those basis states (as long as you're talking about the same general kind of state). So, for example, the spin-z operator $\hat{S}_z$ has two eigenstates, $\left|+z\right\rangle$ and $\left|-z\right\rangle$. Those two eigenstates form a basis, and you can write *any* electron spin state as a sum of constants times those two states. (For example, $\left|+x\right\rangle$ can be written $\frac{1}{\sqrt{2}}\left|+z\right\rangle + \frac{1}{\sqrt{2}}\left|-z\right\rangle$.) This is why we call the complete set of eigenstates for the operator a "basis": it's a base upon which all other states can be built. By considering different different operators from the same general type, you may be able to come up with a different basis. For example, all electron spin states could just as well be written as sums of the eigenvalues of $\hat{S}_x$; in that case, $\left|+x\right\rangle$ and $\left|-x\right\rangle$ form a different set of basis states. By convention, we generally use $\left|+z\right\rangle$ and $\left|-z\right\rangle$ as the basis states for spin angular momentum, but we don't have to.

8.3 Linear Operators

In order to go any further, there are two properties of these operators that we have to know. These operators are what we call *linear* operators. Do not confuse this with a straight line on a graph, for when we say "linear" here we do not mean anything so concrete. Rather, to say that an operator $\hat{\mathcal{O}}$ is linear is to imply the following two properties, where k is a scalar constant:

$$\hat{\mathcal{O}}\ (|\psi_1\rangle\ +\ |\psi_2\rangle)\ =\ \hat{\mathcal{O}}\ |\psi_1\rangle\ +\ \hat{\mathcal{O}}\ |\psi_2\rangle$$

$$\hat{\mathcal{O}}\ (k\ |\psi\rangle)\ =\ k\,\hat{\mathcal{O}}\ |\psi\rangle$$

The action of an operator on a sum of vectors is just the sum of the actions of that operator on the two vectors. Second, if a constant multiplies a vector and you want to operate on that constant times that vector, you are free to pull the constant out in front of the operator, and do the multiplication of by the constant after you perform the operation of the operator on the vector. (Notice that you can't do the multiplication by the constant first once you've pulled it out before the operator. All we've defined for using these vectors is to have one of them operate on a ket vector that is to its right. We haven't defined any way for one of these operators to interact with something to its left.)

8.4 Operators on Non-Eigenstates

What happens when an operator works on a ket vector that is not one of its eigenstates out? You get a different ket vector out, one that can't be expressed as a constant times the ket vector that the operator did its work on. As an example, let's figure out the result of the spin-z operator on an x eigenstate:

$$|\psi\rangle\ =\ \hat{S}_z\ |{+x}\rangle$$

We don't know enough to perform the calculation on the right side of the equation. However, we can substitute for $|{+x}\rangle$ in terms of the z eigenvectors:

$$|\psi\rangle\ =\ \hat{S}_z\ \left(\frac{1}{\sqrt{2}}\ |{+z}\rangle\ +\ \frac{1}{\sqrt{2}}\ |{-z}\rangle\right)$$

We can use the fact that $\hat{S}_z$ is linear to make the next step:

$$|\psi\rangle\ =\ \frac{1}{\sqrt{2}}\,\hat{S}_z\ |{+z}\rangle\ +\ \frac{1}{\sqrt{2}}\,\hat{S}_z\ |{-z}\rangle$$

Now, we have the $\hat{S}_z$ operator just operating on its own eigenstates, and we know how to handle that! We can replace $\hat{S}_z\ |{+z}\rangle$ with $(\hbar/2)\ |{+z}\rangle$, because $\hbar/2$ is the eigenvalue that goes together with the $|{+z}\rangle$ eigenstate of $\hat{S}_z$. We can do the equivalent thing with $|{-z}\rangle$. Performing these substitutions:

$$\begin{aligned}|\psi\rangle\ &=\ \tfrac{1}{\sqrt{2}}\left(\tfrac{\hbar}{2}\right)\ |{+z}\rangle\ +\ \tfrac{1}{\sqrt{2}}\left(-\tfrac{\hbar}{2}\right)\ |{-z}\rangle\\ &=\ \tfrac{\hbar}{2}\ \left(\tfrac{1}{\sqrt{2}}\,|{+z}\rangle\ -\ \tfrac{1}{\sqrt{2}}\,|{-z}\rangle\right)\end{aligned}$$

You may recognize the thing in parentheses here as $|{-x}\rangle$. That means that

$$|\psi\rangle \;=\; \frac{\hbar}{2}\,|{-x}\rangle$$

which is the operation of $\hat{S}_z$ on the state $|{+x}\rangle$:

$$\hat{S}_z\,|{+x}\rangle \;=\; \frac{\hbar}{2}\,|{-x}\rangle$$

Clearly, $|{+x}\rangle$ is not an eigenstate of $\hat{S}_z$, as the operation of $\hat{S}_z$ on that state doesn't return the same state, but a different state. We expected this; if eigenstates correspond to states that have definite values for the observable of a given operator, then states that *don't* have definite values for that observable can't be eigenstates for that operator. And, we've seen before that $|{+x}\rangle$ doesn't have a definite value for z-spin.

8.5 The Hamiltonian

There is one operator in quantum mechanics that is so important it has its own name, and the eigenvalue equation for it in turn has its own name. That is the operator corresponding to energy. Energy is an observable for a quantum particle or a quantum system; it is something that you could measure. The operator that corresponds to the observable energy is called the *Hamiltonian*, and is usually denoted with the symbol $\hat{H}$. If a state $|\psi\rangle$ is an energy eigenstate, then the usual eigenvalue equation applies

$$\hat{H}\,|\psi\rangle \;=\; E\,|\psi\rangle$$

where E is the eigenvalue that corresponds to the eigenvector $|\psi\rangle$. This equation only works if $|\psi\rangle$ is in fact an eigenvector for the Hamiltonian! That means that it is a state that has definite energy, and E is the value of that energy. This equation is called the time-independent Schrödinger equation.

The eigenstates for the Hamiltonian are usually expressed as functions of position. Thus, we might write

$$\hat{H}\,\psi(x) \;=\; E\,\psi(x)$$

and call that the one-dimensional time-independent Schrödinger equation. The function $\psi(x)$ is a "wave function", and is a representation of a quantum state that we have previously been calling $|\psi\rangle$. It is important to remember that both are just representations, and are quite abstract. The wave function or wave vector is an abstract mathematical object, to which you can do things to in order to make predictions about the system that the theory is modeling. The wave function is not a function like $\vec{v}(t)$, the velocity of a particle as a function of time. If you have a full expression for $\vec{v}(t)$, you can interpret it fairly directly; just plug in a time t, work out what the function is, and you've got the velocity at that time. Not so for the wave function $\psi(x)$. Even if you do have the functional form, the interpretation is not so direct. One thing that you can do to it in order to get something physical out is operate the Hamiltonian on it. If you do that, you get back a constant times the function; that constant is the energy that corresponds to the state represented by the function. In later chapters, we will see ways to pull out other physical interpretations of the wave function.

Despite being called "the" Hamiltonian, it is actually not just one operator. Remember that the energy of a particle can come in two forms: kinetic energy and potential energy. The Hamiltonian operator has two parts to it, one for kinetic energy and one for potential energy. Unfortunately, we won't be able to look in detail into how we represent the Hamiltonian operator, because to do so would require calculus. However, even though kinetic energy for a given particle is only a function of its speed, potential energy can be entirely different depending on the environment that the particle is interacting with. For instance, if your particle is at the end of something that behaves just like a spring, then the potential energy for that particle (which is just a function of position) would be the simple harmonic oscillator potential. For an electron orbiting an atomic nucleus, the interaction is the electrostatic interaction and the potential is what we call the Coulomb potential. The Hamiltonians that are built from different potentials will, of course, have different eigenfunctions. We call those eigenfunctions the "solutions" of the Schrödinger equation.

Chapter 9

Vectors and Matrices

Up to now, we've treated each ket vector as an abstract mathematical object unto itself. In the next chapter, we will see a way in which we can represent the ket vectors for a spin-1/2 system and do calculations with them. In order do that, we first have to lay some groundwork for the mathematical objects and operations we will be using.

9.1 Column Vectors

In Section 3.1, we were introduced to the concept of vectors in 3-d space, or 3-vectors. We talked about visualizing them as an arrow in space, and we also talked about representing them as a sum of scalars (i.e. just numbers) times the three basis vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$. Those three basis vectors are vectors with unit magnitude (so that the constants in front of them end up giving you the vector's real magnitude) that just point along the three cardinal axes.

There is a second way that we could represent a 3-vector: as a column vector. Take again our example of $\vec{v}$, the velocity of our car going due northwest, from Section 3.1. From the equation

$$\vec{v} \;=\; v_x\,\vec{e}_x \;+\; v_y\,\vec{e}_y \;+\; v_z\,\vec{e}_z$$

we see that it takes three numbers— v_x, v_y, and v_z— to represent this vector. Instead of representing it as an arrow on a drawing, and instead of writing out the equation above, we could come up with a more compact notation that just lists those three numbers. One such way to do that is to list those three numbers one above each other in a column vector, as such

$$\vec{v} \;=\; \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

In the specific example of our car, we could even numerically represent its velocity as a column vector:

$$\vec{v} = \begin{bmatrix} -35\ \mathrm{km/h} \\ 35\ \mathrm{km/h} \\ 0 \end{bmatrix}$$

If we're going to use this representation, we need to know how to apply the standard vector operators in this representation. For example, you can add together two vectors in order to get a third vector, $\vec{c} = \vec{a} + \vec{b}$. In the column vector representation, this is easy; you just add the individual components:

$$\begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \begin{bmatrix} a_x + b_x \\ a_y + b_y \\ a_z + b_z \end{bmatrix}$$

When you multiply a vector by a scalar, $\vec{b} = k\,\vec{a}$, you just multiply each component by that scalar:

$$\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = k \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} k\,a_x \\ k\,a_y \\ k\,a_z \end{bmatrix}$$

All of this started above when we went from writing the vector as a sum of components times the basis vector into a column listing those components. We can do the same thing for ket vectors! For the spin-1/2 system, you can write a ket vector in terms of the basis vectors; the basis vectors we're using here are $|{+z}\rangle$ and $|{-z}\rangle$. Any ket vector can be written as:

$$|\psi\rangle = a\,|{+z}\rangle + b\,|{-z}\rangle$$

where a and b are complex scalars. Just as the coefficients on the basis vectors for a 3-vector could become the elements of a column vector, we can represent a ket as a column vector. Here, there are only two basis vectors, so we can represent the ket with just a two-row column vector:

$$|\psi\rangle = \begin{bmatrix} a \\ b \end{bmatrix}$$

For instance, the column vectors corresponding to the eigenstates for angular momentum along the three axes are:

$$|{+z}\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |{+y}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{bmatrix} \quad |{+x}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$|{-z}\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad |{-y}\rangle = \begin{bmatrix} i/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad |{-x}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

9.2 Row Vectors

If you can represent ket vectors as column vectors, how about bra vectors? Bra vectors may be represented as a row vector. In order to find the bra vector $\langle\psi|$ that corresponds to a ket vector $|\psi\rangle$, you turn the column vector into a row vector, and take a complex conjugate of each component of the vector:

$$|\psi\rangle \;=\; \begin{bmatrix} a \\ b \end{bmatrix} \quad \langle\psi| \;=\; \begin{bmatrix} a^* & b^* \end{bmatrix}$$

This makes it straightforward to write out the bra vectors corresponding to the angular momentum eigenstates along the three axes:

$$\langle +z| \;=\; \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \langle +y| \;=\; \begin{bmatrix} 1/\sqrt{2} & -i/\sqrt{2} \end{bmatrix} \quad \langle +x| \;=\; \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$\langle -z| \;=\; \begin{bmatrix} 0 & 1 \end{bmatrix} \quad \langle -y| \;=\; \begin{bmatrix} -i/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \langle -x| \;=\; \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

One thing to be careful about with row vectors: remember that it's a sequence of numbers, each in a different column of the row. Don't multiply them together! They're in different spots in the row vector.

9.2.1 The Inner Product

We know we can take the inner product of a bra vector and a ket vector, $\langle \phi \,|\, \psi \rangle$. How do you do this with this column and row vector representation we're building? Let's do an example. First, define a couple of bra vectors:

$$|\phi\rangle \;=\; \begin{bmatrix} a \\ b \end{bmatrix} \quad |\psi\rangle \;=\; \begin{bmatrix} c \\ d \end{bmatrix}$$

The inner product is only defined between a bra vector and a ket vector, so we need to get the ket vector that corresponds to $|\phi\rangle$:

$$\langle\phi| \;=\; \begin{bmatrix} a^* & b^* \end{bmatrix}$$

Now, to do the inner product, we put the two together:

$$\langle \phi \,|\, \psi \rangle \;=\; \begin{bmatrix} a^* & b^* \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}$$

To evaluate this, you multiply the first column of the row vector by the first row of the column vector, and then add to that the second column of the row vector times the second row of the column vector:

$$\begin{bmatrix} a^* & b^* \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} \;=\; a^*c \;+\; b^*d$$

You can multiply a row vector with n rows by a column vector with the same number (n) of rows. You just multiply each component of the first with the corresponding component of the second, and add all of those products together to get the overall result. That overall result is just a scalar. You can *not* multiply a row vector by a column vector unless both have exactly the same number of components.

When you multiply a row vector by a column vector, it's standard always to write the row vector first. It's not a defined operation to multiply a column vector by a row vector if you write the column vector first (at least, as far as we are going to go for our present purposes).[1] This matches the inner product of a bra vector and a ket vector; you always write the bra vector first, as that's the way that you can make the flat sides of each vector fit together.

9.2.2 Nothing is New!

While this is a new formalism for calculations, in fact the addition of column vectors, and multiplying a row vector by a column vector, does exactly the same operations you have already performed previously just by writing out a ket vector in terms of the basis vectors $|{+z}\rangle$ and $|{-z}\rangle$. The column vector formalism makes it faster to perform certain calculations. For example, if you wanted to calculate $\langle +y \,|\, +x \rangle$, you would have to write out both vectors in terms of the z basis vectors, turn the y into a ket, and then work through the algebra. With row and column vectors, you can just start with:

$$\langle +y \,|\, +x \rangle \;=\; \begin{bmatrix} 1/\sqrt{2} & -i/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Calculating this out, you get

$$\begin{aligned} \langle +y \,|\, +x \rangle &= \left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) + \left(\frac{-i}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) \\ &= \frac{1-i}{2} \end{aligned}$$

If you take the absolute square of this, you get $\frac{1}{2}$, which is what we know is the probability for an electron in the $|{+x}\rangle$ state to subsequently be measured to have y spin along the positive y axis. This amplitude is exactly what you would get writing out the two vectors in terms of the z basis, and the ultimate calculation would be the same. By using this formalism, you get to skip writing out a bunch of terms involving $\langle +z \,|\, +z \rangle$, $\langle -z \,|\, +z \rangle$, and the like. You only end up multiplying together the terms

---

[1] In fact, you *can* multiply a column vector by a row vector with the column vector first. However, the result is not a scalar, but a 2×2 matrix! This operation is equivalent to the putting a ket vector and bra vector together in the order $|\phi\rangle\,\langle\psi|$. While such constructions do have their uses in more advanced quantum mechanics, we will not be using them here.

that won't go away due to the orthogonality of the z basis states, and you get fairly quickly to the calculations you need to do.

9.3 Matrices

There is one final bit of mathematical formalism we need to learn before we can get back to the business of applying this formalism to figuring out what will happen in quantum systems. A *matrix* is an extension of column and row vectors. Whereas a column vector has one column and multiple rows, and a row vector is the other way around, a matrix can have multiple columns *and* multiple rows. For our purposes, we need only concern ourselves with square matrices. In the case of the spin-1/2 system, these matrices will be 2×2 matrices. You could write out such a matrix M as:

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

Whereas a row vector or column vector only has two components, a matrix has four components.

What are matrices for? They can be used to represent operators. Remember that an operator, when applied to a ket vector, returns another ket vector. Thus, we need to have a way to apply a 2×2 matrix to a 2-element column vector, which is what we are using to represent a ket vector.

9.3.1 Linear Operations on Matrices

Just like column vectors, you can add together two matrices of the same size. To figure out the result, just add together the components:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

For example:

$$\begin{bmatrix} 2 & i \\ i & 2 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 3+i \\ 3+i & 1 \end{bmatrix}$$

Just as column vectors may be multiplied by a scalar, you may also multiply a matrix by a scalar. As before, the result is a matrix with each component multiplied by the same scalar:

$$k \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} k\,M_{11} & k\,M_{12} \\ k\,M_{21} & k\,M_{22} \end{bmatrix}$$

9.3.2 Multiplying a Matrix and a Column Vector

When using a matrix as an operator that operates on a column vector, the mathematical term for what you are doing is "matrix multiplication". In fact, this is a special case of matrix multiplication; you can learn more general matrix multiplication in a linear algebra course. Multiplying a matrix by a column vector is like repeatedly multiplying row vectors by column vectors. You start with the top row of the matrix. Treat that top row as a row vector, and multiply it by the column vector. That gives you the top row of the answer (which, remember, is itself a column vector). Then go down to the second row of the matrix, and treat that row as a row vector. Multiply it by the column vector. That gives you the second row of the answer. Thus, the result would be:

$$\hat{M}\,|\psi\rangle \;=\; \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} \;=\; \begin{bmatrix} M_{11}\psi_1 \,+\, M_{12}\,\psi_2 \\ M_{21}\psi_1 \,+\, M_{22}\,\psi_2 \end{bmatrix}$$

Notice that the result is *not* a 2×2 matrix. Rather, it's just a column vector!

As an example, we will see in the next chapter that the z angular momentum operator can be represented by the matrix:

$$\hat{S}_z \;=\; \frac{\hbar}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

We can verify that $|{+z}\rangle$ is in fact an eigenvector of $\hat{S}_z$ with the right eigenvalue by trying it out with this representation:

$$\begin{aligned} \hat{S}_z\,|{+z}\rangle &= \frac{\hbar}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{\hbar}{2}\begin{bmatrix} (1)(1)+(0)(0) \\ (0)(1)+(-1)(0) \end{bmatrix} \\ &= \frac{\hbar}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{\hbar}{2}\,|{+z}\rangle \end{aligned}$$

Sure enough, we get the answer that we expected. We get back exactly the same vector, multiplied by the measure of the z-spin corresponding to a system in the state represented by this vector.

9.3.3 The Identity Matrix

There is one special matrix called the identity matrix. If you multiply this matrix by any column vector, the result is exactly the same column vector. In a sense, the number 1 is the 1×1 identity matrix! Hopefully, you are very familiar with the notion that multiplying a number by 1 returns the same number that you started with. The 2×2 identity matrix is:

$$I \;=\; \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Use of this matrix would correspond to the "identity operator" in quantum mechanics, which is not terribly useful. It doesn't really correspond to any observable, and *every* state is an eigenstate of this operator with an eigenvalue of 1! All it does is keep states exactly the way they began.

Chapter 10

Pauli Spin Matrices

We can represent the eigenstates for angular momentum of a spin-1/2 particle along each of the three spatial axes with column vectors:

$$|{+z}\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |{+y}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{bmatrix} \quad |{+x}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$|{-z}\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad |{-y}\rangle = \begin{bmatrix} i/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad |{-x}\rangle = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

Similarly, we can use matrices to represent the various spin operators.

10.1 Spin Operators

We've been talking about three different spin observables for a spin-1/2 particle: the component of angular momentum along, respectively, the x, y, and z axes. In quantum mechanics, there is an operator that corresponds to each observable. The operators for the three components of spin are $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_z$. If we use the column vector representation of the various spin eigenstates above, then we can use the following representation for the spin operators:

$$\hat{S}_x = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \hat{S}_y = \frac{\hbar}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \hat{S}_z = \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

It is also conventional to define the three "Pauli spin matrices" σ_x, σ_y, and σ_z, which are:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Clearly, then, the spin operators can be built from the corresponding Pauli matrices just by multiplying each one by $\hbar/2$.

You can verify that this is a good representation of the spin operators by making sure that all all of the various observations about spin states are reproduced by using these operators and these vectors to predict them from the theory. For example, $|{+y}\rangle$ is an eigenstate for the y component of spin, so the column vector representation of $|{+y}\rangle$ needs to be an eigenvector of $\hat{S}_y$. Is it? Let's try it:

$$\begin{aligned}
\hat{S}_y\,|{+y}\rangle &= \frac{\hbar}{2}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{bmatrix} \\
&= \frac{\hbar}{2}\begin{bmatrix} (0)(1/\sqrt{2}) + (-i)(i/\sqrt{2}) \\ (i)(1/\sqrt{2}) + (0)(i/\sqrt{2}) \end{bmatrix} \\
&= \frac{\hbar}{2}\begin{bmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{bmatrix} \\
&= \frac{\hbar}{2}\,|{+y}\rangle
\end{aligned}$$

In at least this case, the matrix and column vector representations of $\hat{S}_y$ and $|{+y}\rangle$ are working.

10.2 Expectation Values

You can also use the matrix representation of operators to figure out *expectation values*. Suppose that you have an electron in the state:

$$|\psi\rangle = \sqrt{\frac{1}{3}}\,|{+z}\rangle + \sqrt{\frac{2}{3}}\,|{-z}\rangle$$

What are the expectation values of for spin along the x-axis??

First, we construct the column vector representation of this state $|\psi\rangle$:

$$|\psi\rangle = \begin{bmatrix} \sqrt{1/3} \\ \sqrt{2/3} \end{bmatrix}$$

The corresponding bra vector is represented by a row vector:

$$\langle\psi| = \begin{bmatrix} \sqrt{1/3} & \sqrt{2/3} \end{bmatrix}$$

To figure out the expectation value of x-spin, we sandwich the $\hat{S}_x$ operator in between the bra and ket vectors for this state:

$$\begin{aligned}
\langle s_x \rangle &= \left\langle \psi \,\middle|\, \hat{S}_x \,\middle|\, \psi \right\rangle \\
&= \begin{bmatrix} \sqrt{1/3} & \sqrt{2/3} \end{bmatrix} \left(\frac{\hbar}{2}\right) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{1/3} \\ \sqrt{2/3} \end{bmatrix} \\
&= \left(\frac{\hbar}{2}\right) \begin{bmatrix} \sqrt{1/3} & \sqrt{2/3} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{1/3} \\ \sqrt{2/3} \end{bmatrix}
\end{aligned}$$

(all we did between the last two lines was pull the scalar constant $\hbar/2$ out front). We've got a row vector times a matrix times a column vector. That may look intimidating, but we know how to do the matrix times the column vector, so let's do that first. That will leave us with a row vector times a column vector; we know how to work that out as well, leaving us with just a scalar. A scalar is what we need for an expectation value.

$$\begin{aligned}
\langle s_x \rangle &= \left(\frac{\hbar}{2}\right) \begin{bmatrix} \sqrt{1/3} & \sqrt{2/3} \end{bmatrix} \begin{bmatrix} \sqrt{2/3} \\ \sqrt{1/3} \end{bmatrix} \\
&= \left(\frac{\hbar}{2}\right) \left(\sqrt{\frac{1}{3}}\right) \left(\sqrt{\frac{2}{3}}\right) + \left(\sqrt{\frac{2}{3}}\right) \left(\sqrt{\frac{1}{3}}\right) \\
&= \left(\frac{\hbar}{2}\right) \frac{2\sqrt{2}}{3} \\
&= \frac{\sqrt{2}}{3}\hbar
\end{aligned}$$

That's a plausible expectation value. It's neither $\hbar/2$ nor $-\hbar/2$, which means that this is not a definite state for x spin. That's good, because the state is clearly not the same as $|{+x}\rangle$ when you write out that state in terms of $|{+z}\rangle$ and $|{-z}\rangle$. It's between those two. However, from just looking at the state, while you can fairly quickly see that $|{-z}\rangle$ has more amplitude than $|{+z}\rangle$, and thus a measurement of z spin will yield $-\hbar/2$ more often than $+\hbar/2$, it's not obvious at all just looking at the state which value of x spin would be more common, and thus whether the x expectation value should be positive or negative. In this case, you have to perform the calculation. The matrix formulation of the spin operators makes the calculations faster and easier than they would be when you explicit writing out everything in terms of the z basis states.

We could also quickly figure out what the amplitude for measuring positive x spin is with this formalism. Remember that for a particle in state $|\psi\rangle$, the amplitude for finding positive x spin is $\langle +x \,|\, \psi \rangle$. Putting together the $\langle +x|$ bra vector with the column vector for $|\psi\rangle$ above, we get:

$$\begin{aligned}
\langle +x \,|\, \psi \rangle &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{1/3} \\ \sqrt{2/3} \end{bmatrix} \\
&= \left(\sqrt{\tfrac{1}{2}}\right)\left(\sqrt{\tfrac{1}{3}}\right) + \left(\sqrt{\tfrac{1}{2}}\right)\left(\sqrt{\tfrac{2}{3}}\right) \\
&= \sqrt{\tfrac{1}{6}} + \sqrt{\tfrac{1}{3}} \\
&= 0.9856
\end{aligned}$$

That's a high positive amplitude, corresponding to a probability of 0.97 that positive x spin will be measured for this state. Again, without performing the calculations, this is not at all obvious. However, this high probability for positive x spin is consistent with the fact that the x spin expectation value $\langle s_x \rangle$ is positive and only a little bit less than $\hbar/2$.

10.3 Total Angular Momentum

In 3D space, if you have three components of a vector $\vec{v}$, then the magnitude of that vector squared is $v^2 = {v_x}^2 + {v_y}^2 + {v_z}^2$. Angular momentum is a vector, and so this rule would apply to angular momentum as well. However, in quantum mechanics, we see that angular momentum behaves very differently from how it does in classical physics. In particular, if an object has a definite z component of angular momentum, then it has an *indefinite* x component of angular momentum. Does that mean that total angular momentum must also be indefinite? In order to answer this question, we must ask it in a proper quantum manner.

In quantum mechanics, we associate each observable quantity with an operator. We can then use that operator on one of its eigenstates (i.e. a state where the observable has a definite value) to pull out the value of the observable as the eigenvalue a in the equation $\hat{A}\,|\phi\rangle = a\,|\phi\rangle$. If the system is not in an eigenstate, we can figure out the "expectation value" $\langle a \rangle$ (i.e. the weighted average of all values that could be observed) using the operator in the equation

$$\langle a \rangle = \left\langle \psi \mid \hat{A} \mid \psi \right\rangle$$

.

What then is the operator that corresponds to total angular momentum? By analogy to classical physics, we can guess that the operator for total angular momentum squared is:

$$\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$$

This begs the question as to what $\hat{S}_x^2$ means; we know how to square numbers and variables, but how do you square an operator? For $\hat{S}^2$, we will just treat that as an operator itself, the "total spin squared" operator. We'll not treat the superscript 2 as squaring, rather we'll just consider it part of the name. As for the others, what really matters about an operator is what it does to a vector representing a quantum state that it's supposed to operate on. In order to make this definition of the spin angular momentum squared operator to work, we need to interpret them as follows:

$$\hat{S}_x^2 \; |\psi\rangle \;=\; \hat{S}_x \hat{S}_x \; |\psi\rangle$$

In other words, first apply the $\hat{S}_x$ operator to the state $|\psi\rangle$, and then apply the $\hat{S}_x$ operator again to the vector that resulted from the first application. As an example, let's consider the $\hat{S}_x^2$ operator on the state $|{+z}\rangle$:

$$\begin{aligned}
\hat{S}_x^2 \; |{+z}\rangle &= \frac{\hbar}{2}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\frac{\hbar}{2}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}
\end{aligned}$$

All we've done in the first step is pulled the scalar constants out front. To perform this matrix multiplication, first we must multiply the rightmost matrix by the vector, and then we can multiply the first matrix by the result.[1]

$$\begin{aligned}
\hat{S}_x^2 \; |{+z}\rangle &= \frac{\hbar^2}{4}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}0\\ 1\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}1\\ 0\end{bmatrix}\\
&= \frac{\hbar^2}{4}\;|{+z}\rangle
\end{aligned}$$

Interestingly, it seems that $|{+z}\rangle$ *is* in fact an eigenstate of ${\hat{S}_x}^2$, even though it's not an eigenstate of $\hat{S}_x$!

Armed with these techniques, it is possible to show that *any* properly normalized spin-1/2 state $|\psi\rangle$ is an eigenstate of $\hat{S}^2$ with eigenvalue $\frac{3}{4}\hbar^2$. Although it may be surprising that $|{+z}\rangle$ is an eigenstate of $\hat{S}_x^2$, in retrospect it should not be surprising that all states are eigenstates of the total angular momentum operator. We've been saying all along that the *total* angular momentum of an electron is $\frac{\sqrt{3}}{2}\hbar$; what can be in an indefinite state is the components of that angular momentum along various axes.

---

[1] If you know how to multiply 2×2 matrices, you can do the matrix multiplication first if you wish. As we will see, the commutative property does not apply to matrix multiplication, but the associative property does.

Chapter 11

Noncommuting Operators and Uncertainty

11.1 Eigenstates and Commuting Operators

You are probably used to the idea that multiplication is *commutative*. That is, if you have a product ab where a and b are scalars, you can write the multiplication in either order (ab or ba), and the product is exactly the same. This is *not* necessarily the case for matrix multiplication! If A and B are matrices, then $AB \neq BA$ in general. Sometimes it will be true, but not always. Because we can use matrices to represent operators in quantum mechanics, this means that operators don't commute in general. That is, for example, $\hat{S}_x\hat{S}_y\,|\psi\rangle \neq \hat{S}_y\hat{S}_x\,|\psi\rangle$.

Sometimes, however, operators do commute. Suppose that you have two observables A and B with corresponding operators $\hat{A}$ and $\hat{B}$. Suppose also that you have a state $|\phi\rangle$ that is a definite state for *both* A and B. That means that in our mathematical formalism, $|\phi\rangle$ must be an eigenvector for both $\hat{A}$ and $\hat{B}$:

$$\begin{aligned}\hat{A}\,|\phi\rangle &= a\,|\phi\rangle \\ \hat{B}\,|\phi\rangle &= b\,|\phi\rangle\end{aligned}$$

Here, a and b are the eigenvalues for $\hat{A}$ and $\hat{B}$ respectively. In other words, $|\phi\rangle$ has a definite value of observable A, and a is that value; likewise, it has a definite value of observable B, and b is that value.

Let us now consider the application of both of these operators to this state $|\phi\rangle$:

$$\begin{aligned}\hat{A}\,\hat{B}\,|\phi\rangle &= \hat{A}\,b\,|\phi\rangle \\ &= b\,\hat{A}\,|\phi\rangle \\ &= ba\,|\phi\rangle \\ &= ab\,|\phi\rangle\end{aligned}$$

where in the last step we've used the fact that a and b are real numbers, so the product of the two of them does in fact commute. Let's now try this in the other order:

$$\begin{aligned}\hat{B}\,\hat{A}\,|\phi\rangle &= \hat{B}\,a\,|\phi\rangle \\ &= a\,\hat{B}\,|\phi\rangle \\ &= ab\,|\phi\rangle\end{aligned}$$

Here, we can see that in fact the operators $\hat{A}$ and $\hat{B}$ *do* commute if they are operating on a state that is an eigenstate for both operators.

Remember that in the case of spin, we argued that $|{+z}\rangle$ and $|{-z}\rangle$ form a complete basis set of vectors; that is, any spin state $|\psi\rangle$ can be written as a sum of scalar constants times those two vectors. In general, a complete set of eigenvectors for a given operator do form a basis set that can be used to construct any vector that is part of the overall scheme that that operator is part of. (For instance, the projection of spin along all three axes are part of the same scheme, as they are all the same kinds of states— that is, spin angular momentum states.) Therefore, we can write any state $|\psi\rangle$ as a sum of constants times the eigenvectors for that operator. If $\hat{A}$ and $\hat{B}$ are two operators that share the same eigenvectors, then $\hat{A}\hat{B}\,|\psi\rangle = \hat{B}\hat{A}\,|\psi\rangle$. That is, the operation of these two operators on *any* state commutes. For that reason, we generally just say that the operators commute.

11.2 Non-Commuting Operators

In the previous section, we saw that if a particle can be in a definite state for two observables, then the two operators associated with those observables will commute. The converse is therefore also true; if two operators do *not* commute, then it is *not possible* for a quantum state to have a definite value of the corresponding two observables at the same time.

We've already seen examples of this. A particle can't have a definite x spin and a definite y spin at the same time. If our theory is to be useful, then we would hope that $\hat{S}_x$ and $\hat{S}_y$ would not commute when they operate on a general normalized state

$|\psi\rangle$. Let's try it first in one order:

$$\begin{aligned}
\hat{S}_x\hat{S}_y\,|\psi\rangle &= \frac{\hbar^2}{4}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}-i\psi_2\\ i\psi_1\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}i\psi_1\\ -i\psi_2\end{bmatrix}\\
&= i\frac{\hbar^2}{4}\begin{bmatrix}\psi_1\\ -\psi_2\end{bmatrix}
\end{aligned}$$

Now let's try it in the other order:

$$\begin{aligned}
\hat{S}_y\hat{S}_x\,|\psi\rangle &= \frac{\hbar^2}{4}\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}\begin{bmatrix}\psi_2\\ \psi_1\end{bmatrix}\\
&= \frac{\hbar^2}{4}\begin{bmatrix}-i\psi_1\\ i\psi_2\end{bmatrix}\\
&= -i\frac{\hbar^2}{4}\begin{bmatrix}\psi_1\\ -\psi_2\end{bmatrix}
\end{aligned}$$

Clearly the two are not the same; one is the negative of the other. Therefore, $\hat{S}_x$ and $\hat{S}_y$ do *not* commute when operating on a general state ψ, as expected.

It is interesting to note the effect of $\hat{S}_z$ on this same general state:

$$\begin{aligned}
\hat{S}_z\,|\psi\rangle &= \frac{\hbar}{2}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}\\
&= \frac{\hbar}{2}\begin{bmatrix}\psi_1\\ -\psi_2\end{bmatrix}
\end{aligned}$$

Notice that except for the constant out front, the vector produced by $\hat{S}_z$ on this state is the same as the vector produced by $\hat{S}_x\hat{S}_y$ and $\hat{S}_y\hat{S}_x$. In fact, we can put the two together:

$$(\hat{S}_x\hat{S}_y\ -\hat{S}_y\hat{S}_x)\ |\psi\rangle = i\frac{\hbar^2}{2}\ |\psi\rangle$$

$$[\hat{S}_x\,,\,\hat{S}_y]\ |\psi\rangle\ =\ i\hbar\,\hat{S}_z\ |\psi\rangle$$

The term in brackets, $[\hat{S}_x, \hat{S}_y]$ is called the *commutator* of $\hat{S}_x$ and $\hat{S}_y$. It's defined by the term in parentheses above it: $(\hat{S}_x\hat{S}_y - \hat{S}_y\hat{S}_x)$. It works out for the commutators of all three spin angular momentum operators that:

$$[\hat{S}_x\,,\hat{S}_y]\ =\ i\hbar\,\hat{S}_z$$

$$\begin{aligned}[\hat{S}_y\,,\hat{S}_z] &= i\hbar\,\hat{S}_x \\ [\hat{S}_z\,,\hat{S}_x] &= i\hbar\,\hat{S}_y\end{aligned}$$

11.3 Quantifying Uncertainty

If a system is in an indeterminate state for a given observable, it means that we can't know exactly what we're going to measure if we do in fact make a measurement of that observable. We've seen that we can calculate the average of all the measurements we might make, suitably weighted by their probabilities; that's what is called the "expectation value" in quantum mechanics. And, we have seen how we can calculate the amplitude, and from that the probability, that we'll get any given possible measurement. With the spin-1/2 system we've been talking about, there are only two possible values that we might measure for the spin along any given axis: $+\hbar/2$ and $-\hbar/2$. As such, it's convenient just to list the probability of each. If you talk about other systems where there are a lot of possible measurements (including continuous systems such as the position of a particle), it becomes impractical to list the probabilities of each state. It would be nice to have some other way of quantifying our uncertainty.

11.3.1 Mean and Variance

Suppose you have a set of values a_j. By saying that this is a set, we mean that we have several values a_1, a_2, a_3, and so forth. The notation a_j, in this context, means that j can be replaced by any integer between 1 and the total number of values that you have in order to refer to that specific value. Suppose that we have N total values. The average of all of our values can be written as:

$$\langle a\rangle = \frac{1}{N}\sum_j a_j$$

The letter Σ is the capital Greek letter "sigma". This notation means that you sum together all of the values of a_j that you have. For instance, suppose you had just four values, a_1, a_2, a_3, and a_4, then:

$$\sum_j a_j = a_1 + a_2 + a_3 + a_4$$

Therefore, the mean (or average) value of a in this context is:

$$\langle a\rangle = \frac{1}{N}\sum_j a_j = \frac{1}{N}(a_1 + a_2 + a_3 + a_4)$$

To quantify the uncertainty on a set of values, we want to say something about how far, on average, a given value is from the mean of all the values. Thus, it's tempting to try to define the uncertainty as follows:

$$\frac{1}{N}\sum_j \left(a_j \;-\; \langle a\rangle\right)$$

Remember that addition is commutative. Realizing that the $\sum$ symbol just indicates a sum, i.e. a whole lot of addition, we can rewrite this as:

$$\frac{1}{N}\left(\sum_j a_j \;-\; \sum_j \langle a\rangle\right)$$

The second term in the subtraction is a sum over j of the average value. The average value doesn't depend on which a_j we're talking about; it's a constant, it's the same for all of them. Therefore, the sum of that number N times is just going to be equal to $N\,\langle a\rangle$. Making this substitution and distributing the $1/N$ into the parentheses:

$$\frac{1}{N}\sum_j a_j \;-\; \frac{1}{N}N\,\langle a\rangle$$

But we recognize the first term in this subtraction as just $\langle a\rangle$. So, the total result of this is zero. Clearly, this is not a good expression for the uncertainty in a. If you think about it, the average deviation of a_j from $\langle a\rangle$ *ought* to be zero. If $\langle a\rangle$ is the average value of a, then a_j should be below $\langle a\rangle$ about as often as it is above, so your sum will have a mix of positive and negative terms. The very definition of the average insures that this sum will be zero.

Instead, we shall define the *variance* as:

$$\Delta a^2 \;=\; \frac{1}{N}\sum_j (a_j \;-\langle a\rangle)^2$$

Here, we're using Δa to indicate the uncertainty in a. The variance is defined as the uncertainty squared.[1] The advantage of this expression is that because we're squaring the difference between each value a_j and the average value, we're always going to be summing together positive terms; there will be no negative terms to cancel out the positive terms. Therefore, this should be a reasonable estimate of how far, typically, the measurements a_j are from their average.

---

[1] If you know statistics, you may recognizing this as being very similar to how variance is defined there— only in statistics, we divide by $N-1$ rather than by N. The difference becomes unimportant as N gets large.

We can unpack this sum a bit, first by multiplying out the squared polynomial:

$$\Delta^2 \;=\; \frac{1}{N}\sum_j (a_j^2 \;-\; 2\langle a\rangle\, a_j \;+\; \langle a\rangle^2)$$

In order to clean this expression up, inside the parentheses both add and subtract $\langle a\rangle^2$:

$$\begin{aligned}
\Delta a^2 \;&=\; \frac{1}{N}\textstyle\sum_j (a_j^2 \;-\; 2\langle a\rangle\, a_j \;+\; 2\langle a\rangle^2 \;-\; \langle a\rangle^2) \\
&=\; \frac{1}{N}\textstyle\sum_j (a_j^2 \;-\; \langle a\rangle^2 \;+\; 2\langle a\rangle\,(\langle a\rangle - a_j)) \\
&=\; \frac{1}{N}\textstyle\sum_j a_j^2 \;-\; \frac{1}{N}\sum_j \langle a\rangle^2 \;+ \frac{1}{N}\,2\langle a\rangle \sum_j (\langle a\rangle - a_j)
\end{aligned}$$

Notice that the last term is going to be zero, as it includes the average difference between the mean and each observation. The second term is just going to be $\langle a\rangle^2$, because once again $\langle a\rangle$ is the same for all terms of the sum; the sum will yield $N\langle a\rangle^2$, canceling the N in the denominator. So, we have:

$$\Delta a^2 \;=\; \left\langle a^2\right\rangle \;-\; \langle a\rangle^2$$

11.3.2 Uncertainty in Quantum Mechanics

In order to bring this into quantum mechanics, we already know how to calculate the average $\langle a\rangle$, which we call the "expectation value". If the state of the system is $|\psi\rangle$ and the operator corresponding to the observable a is $\hat{A}$, then

$$\langle a\rangle \;=\; \left\langle \psi \middle| \hat{A} \middle| \psi \right\rangle$$

Similarly, now that we recognize that we can interpret $\hat{A}^2$ as just applying the operator $\hat{A}$ twice, we can calculate $\langle a^2\rangle$:

$$\left\langle a^2\right\rangle \;=\; \left\langle \psi \middle| \hat{A}^2 \middle| \psi \right\rangle$$

For example, let's consider the state $|\psi\rangle = |{+z}\rangle$ and the observable spin-z. We expect the uncertainty here to be zero, because we know exactly what we'll get if we measure spin-z. Let's see if it works out that way:

$$\begin{aligned}
\langle s_z \rangle &= \left\langle \psi \middle| \hat{S}_z \middle| \psi \right\rangle \\
&= \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar}{2}
\end{aligned}$$

As expected, the expectation value for spin-z is $+\hbar/2$. For the other part:

$$\begin{aligned}
\langle {s_z}^2 \rangle &= \left\langle +z \middle| \hat{S}_z \hat{S}_z \middle| -z \right\rangle \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar^2}{4}
\end{aligned}$$

If we take the difference $\langle {s_z}^2 \rangle - \langle s_z \rangle^2$, we get $\hbar^2/4 - \hbar^2/4 = 0$, as expected.

What if we want to know the uncertainty on S_x for this state?

$$\begin{aligned}
\langle s_x \rangle &= \left\langle +z \middle| \hat{S}_x \middle| +z \right\rangle \\
&= \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= 0
\end{aligned}$$

If the system is in the state $|+z\rangle$, we know that we have a 50% chance each for finding spin-x to be $+\hbar/2$ or $-\hbar/2$. Thus, it's no surprise that the average value of spin-x is zero, even though zero isn't a value we might measure. To figure out the variance:

$$\begin{aligned}
\langle {s_x}^2 \rangle &= \left\langle +z \left| \hat{S}_x \hat{S}_x \right| +z \right\rangle \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\hbar^2}{4} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{\hbar^2}{4}
\end{aligned}$$

Thus, in this case, the formal uncertainty Δs_x on the x-spin is $\hbar/2$.

11.4 The Heisenberg Uncertainty Principle

As previously stated, quantifying the uncertainty on a given observable for a given quantum state is more interesting when the observable we're talking about has a large number (or even a continuum) of different values it might take on. If you consider two different observables whose operators do not commute, then a system cannot be in a definite state for both of those observables at the same time. The Heisenberg Uncertainty Principle takes this observation, makes it stronger, and quantifies it.

Consider a quantum particle that can move along one direction. Its position is then x, and its momentum along that direction is p_x. The Heisenberg Uncertainty Principle states that:

$$\Delta x \, \Delta p_x \geq \frac{\hbar}{2}$$

We've gotten used to thinking of $\hbar$ as an angular momentum unit, because that's where it's shown up before. However, here, it's not really an angular momentum, though it still does of course have the same units (position times position over time). Instead, it represents the fundamental limit in quantum mechanics on how well you can know two different observables, position and momentum. If you know one of them *perfectly*, e.g. $\Delta x = 0$, then the uncertainty in the other one must be *infinite*. Although in more advanced quantum mechanics we use such states as they're a good approximation for a lot of things, they're not really physical. In reality, most quantum systems have a small amount of uncertainty in both position and momentum. That is, a particle doesn't have a definite position or a definite momentum, but the range of positions for which it has an appreciable amplitude is confined to a small space, and the range of momenta for which it has an appreciable amplitude is confined to a small range.

As a concrete example, let's consider an electron. For a non-relativistic electron, its momentum is just $p = mv$, where m is its mass and v is its speed. Therefore, $\Delta p = m\Delta v$, as the mass is well known and there is no uncertainty in it. What is a good uncertainty in speed to consider? For practical purposes, let's suppose that we're doing an experiment with an electron that requires it to be localized for 1 second. We don't want the uncertainty in the speed of the electron to cause our uncertainty in the position after one second to be greater than the uncertainty in the position was in the first place. So, we shall choose $\Delta v = \Delta x/t$, where we'll put in $t = 1$ second. If we then put this into the uncertainty principle

$$\begin{aligned}\Delta x\,\Delta p &\geq \frac{\hbar}{2}\\ \Delta x\, m_e\frac{\Delta x}{t} &\geq \frac{\hbar}{2}\\ \Delta x &\geq \sqrt{\frac{\hbar\, t}{2\, m_e}}\end{aligned}$$

If you put in the numbers, you find that the uncertainty on the position of this electron is 0.01 m, or one centimeter. For an electron, that's a lot! (One could argue about whether or not 1 second is a reasonable timescale. When we get to talking about atoms, we'll think more carefully about what a reasonable timescale is.)

Notice, however, that the uncertainty in the position goes down as the mass goes up. Imagine that you stood still your entire life. If you want to balance the uncertainty in your starting position with the uncertainty in your position resulting from the uncertainty in your velocity over your entire life, then you'd put in your age for t. Let us assume, optimistically, that you will live 100 years (3×10^9 seconds), and that your mass is 80 kg. If you put those numbers in to the equation above, you find out that the uncertainty on your position is 4×10^{-14} m. In other words, even though quantum uncertainty can be pretty important for an electron, on everyday scales for macroscopic objects the effect of quantum uncertainty is utterly negligible.

11.4.1 An Alternate Formulation

An alternate way to formulate Heisenberg's Uncertainty principle is:

$$\Delta E\,\Delta t \geq \frac{\hbar}{2}$$

The interpretation of this is a little less clear than in the case of position and momentum. Uncertainty in energy seems obvious enough; it's the square root of the variance of all the values of energy that might be measured for a particle in a given quantum state. But what is "uncertainty on time"? Rather than interpreting this as an uncertainty, we shall interpret it as a time interval. In a sense, that's the same thing; an interval of time is qualitatively similar to an uncertainty on what time it is.

What this means, then, is that the uncertainty in the energy of a quantum state is related to how long that state hangs around. If a system is in an energy eigenstate, then it has a definite energy and $\Delta E = 0$. Such a state *must* be stable then, for Δt has to be infinite. In other words, in the absence of any interactions, a particle in an energy eigenstate will stay, forever, in an energy eigenstate.

For small time intervals, however, there will be a finite uncertainty in the energy of a system. One thing that this means is that it becomes possible to violate the conservation of energy, so long as you do it so fast that nobody can catch you at it! Among other things, this leads to the possibility of quantum tunneling– that is, if a particle is up against a potential barrier it doesn't have enough energy to penetrate, there *is* some finite probability that the particle may be located *inside the barrier*. And, the particle may be able to cross the barrier, even though classically it could not.

Later, when we talk about atoms, states other than the ground state (i.e. lowest-energy state) of the atom aren't going to be perfectly stable. Over time, they will decay to the ground state, with a characteristic lifetime analogous to the half-life of a radioactive isotope. Although we will describe these excited states as being energy eigenstates, the fact that they decay tells us that they can't *exactly* be energy eigenstates. It also tells us that there must be some uncertainty as to the exact energy value associated with those states. There will be observational consequences of this, although in practice for real atoms these consequences are extremely difficult to observe.

11.4.2 Vacuum Energy

A second consequence of this formulation of the Heisenberg Uncertainty Principle is the possibility of vacuum energy. Consider a small region of space. Suppose that it's empty; that is, you've taken out everything you can take out of it, including atoms, light (photons), dark matter, and so forth. Make sure that there are no quantum systems anywhere with non-negligible probability for being found in this region of space. Over a finite time interval Δt, you can't be sure exactly how much energy there is in this region of space; your uncertainty in the amount of energy must be at least $\Delta E = \frac{\hbar}{2\Delta t}$. As a result, there *may* be energy in the vacuum.

What is the expectation value of this energy? You might predict that the expectation should be 0, even though the uncertainty has to be greater than zero. Figuring it out requires going into relativistic quantum mechanics, called quantum field theory. Unfortunately, even quantum field theory can't calculate that right, for naive estimates of what you'd get (the best we can really do) gives a value of the vacuum energy density that is so high that it would prevent galaxies from ever having formed in our Universe. The fact that you are reading this indicates that this estimate cannot be right. Indeed, quantum field theory estimates a value for the vacuum energy density that is 120 orders of magnitude too big! That's pretty far off. As such, we have to say that we don't completely understand the nature of vacuum energy.

What form would this vacuum energy take? We've already seen that in a finite time interval Δt, we can't say with certainty that the vacuum has zero energy. In quantum field theory, it becomes possible to create and destroy particles, as long as you obey all of the conservation laws. For example, two photons can interact and create an electron/positron pair, where a positron is the antimatter partner to an electron. If you don't have to worry about conserving energy, however, you can create a positron/electron pair out of *absolutely nothing*... as long as they re-annihilate back to absolutely nothing fast enough. For every fundamental particle that exists, this sort of thing is going on around us all the time.

What is the net energy density of the vacuum as a result of all of this? For a long time, many physicists assumed that a various terms would cancel out to zero; the naive calculations indicated something absurd, and the most natural result if those calculations are wrong is that things would cancel out. However, in the last ten years, observations of the expansion of the Universe have shown that the expansion is accelerating; indeed, these astronomical observations were the source of the 2011 Nobel Prize in Physics. We don't know what is causing this, and have given the name "dark energy" to whatever it is that is causing it. The simplest explanation for dark energy is that it is vacuum energy. Measurements from cosmology indicate a vacuum energy density corresponding to about 10^{-29} grams per cubic centimeter. That is, the energy density of vacuum energy is 29 orders of magnitude less than the mass-energy density of water. Obviously, we can ignore this in our every day life. However, if you look at the Universe as a whole, most of it is empty; our planet is a very special place that is, compared to most of the Universe, extremely dense with regular atoms. In the Universe as a whole, dark energy makes up three quarters of the energy density. Even though this density may be 120 orders of magnitude smaller than what naive estimates from our theory would suggest, it is coming to dominate the evolution of our Universe.

Chapter 12

Multiple Particle States

12.1 Indistinguishable Particles

Every electron is exactly the same as every other electron. Thus, all electrons are indistinguishable. This means that if you have a state with two electrons, you can swap the two electrons and it cannot change anything physically observable from that state.

To make this concrete, suppose the state $|\psi\rangle$ is a state with two electrons. Let's define $|\psi'\rangle$ as the state with those two electrons swapped. Then, the expectation value of any operator must be the same for these two different states:

$$\left\langle \psi \middle| \hat{\mathcal{O}} \middle| \psi \right\rangle = \left\langle \psi' \middle| \hat{\mathcal{O}} \middle| \psi' \right\rangle$$

Also, the *probability* for any measurement of any observable to be made must be the same for the two states. That is, if $\langle\phi|$ is an eigenstate of a given observable, then

$$|\langle \phi \,|\, \psi \rangle|^2 = |\langle \phi \,|\, \psi' \rangle|^2$$

If you think about it, however, this does *not* mean that the two states must be identical! However, they must be close enough such that *anything physically observable from the state* must be identical. Below, we will introduce the *exchange operator* as a way of quantifying the effect of identical particles on quantum states.

12.2 Notating Multiple Particle States

Before we go further, we need to refine our notation so that we can keep track of two different particles. We can construct a two-particle state by putting together two states for each individual particle with:

$$|\psi_1\rangle \otimes |\phi_2\rangle$$

The $\otimes$ operator indicates that we're putting these two states together to form a composite state. It's sometimes called a "direct product", but it's not really all that much like multiplication. Really, it just means that we're making some composed state that combines particle 1 in state $|\psi\rangle$ and particle 2 in state $|\phi\rangle$. The subscript indicates which particle we're talking about; the rest of the stuff inside the ket indicates the state of that particular particle.

For simplicity, we will often omit the $\otimes$ symbol in the "direct product", and just write the two states next to each other, e.g.

$$|\psi_1\rangle\,|\phi_2\rangle$$

Again, this does not mean that we're multiplying two ket vectors, which is something we can't do. Instead, it means that we're *composing* the states. If these were spin states, we would *not* represent this with two column vectors. Instead, we'd represent it with a *single* four-row column vector; the first two rows have the column vector representation of whatever state the first particle is in, and the second two rows have the column vector representation of whatever state the second particle is in.

If an operator operates on this state, it only affects the state for the particle it is an operator for. That is, if "spin-z for particle 2" is the observable we're talking about, then the operator $\hat{S}_{z2}$ only operates on (in this example) the state $|\phi_2\rangle$. Indeed, you can treat $|\phi_1\rangle$ as if it were a constant:

$$\hat{S}_{z2}\;|\psi_1\rangle\,|\phi_2\rangle \;=\; |\psi_1\rangle\;\hat{S}_{z2}\;|\phi_2\rangle$$

As an example, suppose that particle 1 is in the state $|{+z}\rangle$ and particle 2 is in the state $|{-z}\rangle$. If we apply the $\hat{S}_{z2}$ operator to this state, we get:

$$\begin{aligned}
\hat{S}_{z2}\;|{+z_1}\rangle\,|{-z_2}\rangle &= |{+z_1}\rangle\;\hat{S}_{z2}\;|{-z_2}\rangle \\
&= |{+z_1}\rangle\left(\frac{-\hbar}{2}\right)|{-z_2}\rangle \\
&= \left(-\frac{\hbar}{2}\right)|{+z_1}\rangle\,|{-z_2}\rangle
\end{aligned}$$

Here, we have taken advantage of the fact that $|{-z_2}\rangle$ is an eigenstate of $\hat{S}_{z2}$, and replaced the action of the operator with a simple multiplication by the eigenvalue.

There will be some operators (e.g. the forthcoming exchange operator) that don't operate on just one of the two particles, but on both at the same time.

Similarly, with inner products, bra versions of a state only “stick” to ket versions of a state on the straight side of the bra-ket notation if they are states for the same particle. Thus, suppose we had a composite state:

$$|\xi\rangle \;=\; |\psi_1\rangle\,|\phi_2\rangle$$

The corresponding bra vector is:

$$\langle\xi| \;=\; \langle\psi_1|\,\langle\phi_2|$$

Normalization of this state is then expressed as:

$$\begin{aligned}\langle\xi\,|\,\xi\rangle &= (\langle\psi_1|\,\langle\phi_2|)\;\;(|\psi_1\rangle\,|\phi_2\rangle)\\ &= \langle\psi_1\,|\,\psi_1\rangle\;\langle\phi_2\,|\,\phi_2\rangle\\ &= 1\end{aligned}$$

We’ve rearranged states here a bit. We moved the $|\psi_1\rangle$ from after the $\langle\phi_2|$ to before it. This should make you a little nervous; we’ve seen that with matrices and other things that aren’t simple numbers, multiplication is not necessarily commutative. However, again, in this case, when it comes to inner products, a state for a *different* particle can be treated as a constant with respect to inner products for the first particle. As such, it’s entirely legitimate to move $|\psi_1\rangle$ into, out of, and through inner products on particle 2 (at least in the case of the simple composed states we’re talking about here).

12.3 The Exchange Operator

The exchange operator, notated here by $\hat{P}_{12}$, just exchanges particle 1 for particle 2. In order to satisfy the conditions described in Section 12.1, a state composed of two *indistinguishable* particles (e.g. two electrons) must be an eigenstate of the exchange operator. Suppose that $|\xi\rangle$ is such a state. This means that

$$\hat{P}_{12}\;|\xi\rangle \;=\; c\;|\xi\rangle$$

where c is the eigenvalue. Suppose that we apply the exchange operator twice. What will happen? We should get back to the original state! We’ve just swapped the two particles back. Let’s apply this twice:

$$\begin{aligned}\hat{P}_{12}\,\hat{P}_{12}\;|\xi\rangle &= \hat{P}_{12}\;(c\;|\xi\rangle)\\ &= c\,\hat{P}_{12}\;|\xi\rangle\\ &= c^2\;|\xi\rangle\end{aligned}$$

If the result of applying this exchange operator twice must be the state we started with, then we must have $c^2 = 1$. This is regular old fashioned squaring, not taking the absolute square. That $c^2 = 1$ means that there are only two possibilities for the eigenvalue of the exchange operator working on a state of two indistinguishable particles: $c = 1$ or $c = -1$.

12.4 Fermions and Bosons

In quantum mechanics, there are two kinds of particles. **Fermions** are particles that are *antisymmetric* under the exchange operator; that is, if $|\xi\rangle$ is a two-particle state for two indistinguishable fermions, $\hat{P}_{12}\,|\xi\rangle = -\,|\xi\rangle$. **Bosons** are particles that are *symmetric* under the exchange operator; that is, if $|\xi\rangle$ is a two-particle state for two indistinguishable bosons, $\hat{P}_{12}\,|\xi\rangle = |\xi\rangle$. This is summarized below:

$$\hat{P}_{12}\;|\xi\rangle \;=\; \begin{cases} \;\;\;|\xi\rangle & \text{for a two-boson state} \\ \\ -\,|\xi\rangle & \text{for a two-fermion state} \end{cases}$$

Which particles are which? Particles that have half-integral spin— which includes the spin-1/2 electrons we've been talking about all this time— are fermions. Other fermions include protons, neutrons, quarks, and neutrinos. Particles with integral spin are bosons. Bosons include photons, pions, and the force carriers for the weak and strong nuclear forces.

How do you create a two-fermion state with a total z component of angular momentum equal to zero? The most obvious first thing to guess is just to assign each particle angular momentum in a different direction, so that they cancel:

$$|\xi\rangle \;=\; |{+z_1}\rangle\,|{-z_2}\rangle$$

However, this state doesn't work! Why not? Consider the operation of the exchange operator on it:

$$\hat{P}_{12}\;|{+z_1}\rangle\,|{-z_2}\rangle \;=\; |{+z_2}\rangle\,|{-z_1}\rangle$$

We started with particle one having positive z-spin and particle 2 having negative z-spin. After the exchange, it's the other way around. However, this isn't the same state, nor is it a constant times the original state. On other words, this state is not an eigenstate of the exchange operator. Therefore, it's not a valid quantum state if particle 1 and particle 2 are indistinguishable particles (e.g. if they're two electrons).

A valid two-fermion spin state with total angular momentum zero would be:

$$|\xi\rangle \;=\; \frac{1}{\sqrt{2}}\;|{+z_1}\rangle\,|{-z_2}\rangle \;-\; \frac{1}{\sqrt{2}}\;|{+z_2}\rangle\,|{-z_1}\rangle$$

To verify that this works, let's try the exchange operator on this state:

$$\begin{aligned}\hat{P}_{12}\,|\xi\rangle &= \hat{P}_{12}\left(\tfrac{1}{\sqrt{2}}\,|{+z_1}\rangle\,|{-z_2}\rangle \;-\; \tfrac{1}{\sqrt{2}}\,|{+z_2}\rangle\,|{-z_1}\rangle\right)\\ &= \tfrac{1}{\sqrt{2}}\,\hat{P}_{12}\,|{+z_1}\rangle\,|{-z_2}\rangle \;-\; \tfrac{2}{\sqrt{2}}\,\hat{P}_{12}\,|{+z_2}\rangle\,|{-z_1}\rangle\\ &= \tfrac{1}{\sqrt{2}}\,|{+z_2}\rangle\,|{-z_1}\rangle \;-\; \tfrac{1}{\sqrt{2}}\,|{+z_1}\rangle\,|{-z_2}\rangle\\ &= -\,|\xi\rangle\end{aligned}$$

Sure enough, this state is an eigenstate of the exchange operator. What's more, the eigenvalue is -1, which is required for fermions. (If you're wondering about why we mess about with all of the $1/\sqrt{2}$ coefficients, we do that so that $|\xi\rangle$ is properly normalized. You can verify that this is the case, and indeed doing so would be good practice in doing algebra with bra and ket vector representations of multiple particle states.)

12.5 The Pauli Exclusion Principle

The Pauli Exclusion Principle states that no two fermions may occupy the same quantum state. This principle is absolutely crucial to life as we know it; without it, we would not have the Periodic Table of chemistry, nor would we have a lot of the rest of the structure of matter. This doesn't mean, however, that only one electron in the Universe is allowed to have positive z spin! Obviously, we have many more than two electrons in the Universe. However, if you have a quantum state, such as an energy level in an atom, where you can put electrons, you can only put *two* electrons into that energy level. Why two, and not one? Because of electron spin; as long as the two electrons have opposite spin (or, more precisely, are in a combined spin state with spin angular momentum zero such that they are antisymmetric under exchange), then you can put two electrons into the same state. It *is* possible to have two electrons with the same spin, so long as something else is different about their quantum states. So, for example, you could have two electrons with the same spin if they were in different orbitals in an atom.

Why can't you put more than one fermion in the same state? Because it's impossible to construct an antisymmetric state vector two fermions in the same state. Suppose you have a state $|\psi\rangle$, and you want to put two fermions into it. We know that the state:

$$|\psi_1\rangle\,|\psi_2\rangle$$

won't work, because the exchange operator working on it just produces the same state back, not the negative of the same state:

$$\hat{P}_{12}\,|\psi_1\rangle\,|\psi_2\rangle \;=\; |\psi_2\rangle\,|\psi_1\rangle \;=\; |\psi_1\rangle\,|\psi_2\rangle$$

This is an eigenvalue of the exchange operator, which is good, but the eigenvalue is +1. This would work for bosons; indeed, because of this, you can put as many bosons as you want all into the same state. However, for fermions, the eigenvalue of the exchange operator working on the two-particle state needs to be −1. If we try to construct an antisymmetric wave vector with both of these electrons in the same state:

$$\frac{1}{\sqrt{2}}\left|\psi_1\right\rangle\left|\psi_2\right\rangle \;-\; \frac{1}{\sqrt{2}}\left|\psi_2\right\rangle\left|\psi_1\right\rangle$$

we just end up with 0, which isn't a state at all. Thus, if you have two indistinguishable fermions, there *must* be something different about their states; you can't put more than one fermion into a single quantum state.

12.6 Entangled Particles

When two particles' quantum state is a combined quantum state, we say that those two particles are *entangled*. Most of the time we encounter such states, we don't worry about it too much. The two electrons in the ground state of Helium have entangled states, because they are indistinguishable particles. You can't talk about the state of one electron without talking about the state of another.

Entangled quantum states become more interesting when you separate the two particles. Suppose that there is some sort of reaction that produces two electrons that have a total spin angular momentum of zero. We've seen before that the state of these two electrons is then:

$$\frac{1}{\sqrt{2}}\left|+z_1\right\rangle\left|-z_2\right\rangle \;-\; \frac{1}{\sqrt{2}}\left|+z_2\right\rangle\left|-z_1\right\rangle$$

Although the total z angular momentum of this combined state is 0, a definite value, the angular momentum of an individual electron is *not* in a definite state. Now suppose that you separate these two electrons; it may be that the reaction that produces them sends them shooting off in two directions, which for discussion purposes we shall call "left" and "right".

Now let's suppose that somebody far off to the left detects the left electron and measures its z-spin. This measurement will collapse the wave function of the left electron, putting it into a state of definite z spin. However, because it's a combined state for the two electrons, you can't collapse the wave function of just one of them; you have to collapse the entire state all at once. Therefore, if somebody measures the z spin of the left electron, the wave function of the right electron *also* collapses at that moment, even if nobody has made a measurement on it. If the left observer measures that the left electron is spin up, then anybody off to the right *will* observe that the right electron is spin down; the right electron is no longer in an indefinite state, even though nothing was done to it.

This behavior of entangled particles is what Einstein referred to as “spooky action at a distance”. (citation needed.) Not only was he disturbed by the stochastic nature ☑ of quantum mechanics, he was also bothered by what *seemed* to be communication faster than the speed of light. Does some sort of signal traverse from one electron to the other electron in order to communicate the fact that their mutual wave function has collapsed? Together with two other physicists, Podolsky and Rosen, Einstein argued that this behavior indicated that quantum theory had to be incomplete. In 1935, they published a paper describing what is now known as the “EPR Paradox” (Einstein et al., 1935). If quantum mechanics is indeed incomplete, then there would need to be some sort of “local hidden variable” that tells a particle which way its wave function *should* collapse when that particle is measured. This variable is “hidden” because it is not accounted for in quantum mechanics. In the early 1960’s, physicist John Bell proposed experiments that would test the EPR paradox by being able to tell the difference between the standard predictions of quantum mechanics and the predictions of a theory that had some sort of local hidden variables (citation needed). ☑ Experiments performed since then have shown that in fact standard quantum mechanics does predict the correct results, and that therefore there are no local hidden variables. The fact is that, somehow, the wave function of an electron can collapse when another electron is measured— and that other electron may, at least in principle even if this is not realizable in practice, be light-years away. This raises philosophical issues associated with the interpretation of quantum mechanics, but also indicates that quantum mechanics remains a very robust theory.

Chapter 13

The Schrödinger Equation

13.1 Where we are so far

We have focused primarily on electron spin so far because it's a simple quantum system (there are only two basis states!), and yet it still shows much of the peculiar nature of reality on the quantum level. In particular, we've seen the following things in the theory of quantum mechanics:

- A "system" (e.g. the angular momentum vector of an electron) may be an an *indefinite* state, also sometimes called a "mixture" of states, where an observable doesn't have a set value. Rather, the state of the system is such that if the observable were measured, there is a probability of different values being observed. The mathematical theory represents this by allowing states to be sums of coefficients times orthogonal basis states. For example, with angular momentum of a spin-1/2 particle such as an electron, the basis states are $|{+z}\rangle$ and $|{-z}\rangle$.
- Observables may take on *quantized* values. For example, every time you measure the z component of angular momentum of an electron, you get either $+\hbar/2$ or $-\hbar/2$. This is in sharp contrast to what you'd see in classical physics.
- What propagates in quantum mechanics is *amplitudes.* For example, if an electron is in state $|\psi\rangle$, the amplitude to measure it to have z angular momentum $+\hbar/2$ is $\langle +z \,|\, \psi\rangle$. The *probability*, which is what we can really find in experiments, is the absolute square of the amplitude; in this example, that would be $|\langle +z \,|\, \psi\rangle|^2$.
- Different observables may be *orthogonal* (the second use of this term). If they are, then a system can not be in a definite state for those two observables at

the same time. The projections of angular momentum along different axes are orthogonal; position and momentum along the same direction are orthogonal.

- Observables in quantum mechanics are paired with *operators*. A quantum mechanical operator operates on a quantum state (represented by a ket vector), and the result of that operation is another (non-normalized) quantum states (i.e. another ket vector). For example, if we call the z component of angular momentum spin-z or just s_z, the operator that goes with it is $\hat{S}_z$, the spin-z operator. Operators are quite abstract, and form a mathematical part of the theory that is useful, but is difficult to interpret and associate directly with something that you could observe.

- A state that is a definite state for a given observable is an *eigenstate* of that operator. (We would also say that the ket vector that represents that state is an *eigenvector* of the operator; if we're representing operators as matrices, then the column vector that represents the state is an *eigenvector* of the operator.) An operator working on one of its eigenstate returns a constant times the same state. That constant is called the *eigenvalue* associated with the eigenstate. If this operator corresponds to an observable, that eigenvalue must be a real number, and corresponds to the physical measurement you'd make of that observable. For example:

  $$\hat{S}_z \left|+z\right\rangle \;=\; \frac{\hbar}{2}\left|+z\right\rangle .$$

  This equation is the *eigenvalue equation*, in this case specifically for the z-spin operator and the $|+z\rangle$ state. The state $|+z\rangle$ is a state of definite z-spin, so it is an eigenstate of the z-spin operator $\hat{S}_z$. The eigenvalue equation for this state and this operator includes the constant $\hbar/2$, which is the actual value of the z component of spin angular momentum that an electron in state $|+z\rangle$ has.

- You can find the *expectation value* for a system in state $|\psi\rangle$ for a given observable by sandwiching the observable's operator between $\langle\psi|$ and $|\psi\rangle$. For example, the expectation value for z-spin for a given electron is $\left\langle \psi \,\middle|\, \hat{S}_z \,\middle|\, \psi \right\rangle$. The expectation value is the *average* value you'd get if you measured the observable for that state. That is, if you took a large number of systems in that state and measured the observable for all of those systems, you'd get different results, with probabilities for each result predicted by the mathematics of quantum mechanics. The average of all those results would be the expectation values.

Although the eigenvalue equation is fairly abstract, it's a very important part of the mathematical theory of quantum mechanics. The only direct connection it has to what we might observe in the lab is that it extracts (in the form of the eigenvalue) the measured quantity for the observable that you'd get for a given eigenstate (i.e.

definite state) of that observable's operator. However, the operator itself doesn't represent any particular *physical* operation you might perform in the lab.

From a broader point of view, the eigenvalue equation is the equation you can use to *figure out* what states are possible definite states for a given operator, and therefore what values you might measure for the observable associated with that operator.

13.2 Stating the Equation

As was briefly mentioned in Section 8.5, the energy operator is an operator so important to quantum mechanics that it gets its own proper name, the Hamiltonian, and the eigenvalue equation for it also gets its own name, the Schrödinger equation. It is this equation that allows us to figure out the energy states of a system, and it could be argued that energy states are the most important states in quantum mechanics. It is energy levels in atoms that provides all of the structure that gives us the Periodic Table of the Elements, and it is transitions between those energy levels that we observe in a number of both terrestrial and astrophysical contexts.

At the most base level, the Schrödinger equation is just the energy operator eigenvalue equation:

$$\hat{H}\,|\psi\rangle \;=\; E\;|\psi\rangle$$

$\hat{H}$ is the Hamiltonian; it's the operator that corresponds to energy as an observable. Solutions $|\psi\rangle$ to this equation are the eigenstates of energy. The value E that goes with a given solution $|\psi\rangle$ is the energy associated with that state. Technically, this equation is called the *Time-Independent Schrödinger Equation.* (There is also a full Schrödinger equation that describes how quantum states evolve in time.)

A full investigation of the Hamiltonian requires differential calculus, so we won't fully present it here. However, you can break the Hamiltonian into two parts. In doing so, we're going to go to a *wave function* representation of the state vector $|\psi\rangle$. Whereas we have used column vectors to represent spin states, it is more traditional (and more useful) to represent energy states as functions of position $\psi(x, y, z)$. As with a regular function, $\psi(x, y, z)$ is just something into which you can plug a position (i.e. values of x, y, and z) and get a number— although in this case that number can be a complex number. Dividing the Hamiltonian into two parts and writing the state as a wave function yields this form of the Schrödinger equation:

$$\hat{K}\psi(x,y,z) \;+\; V(x,y,z)\psi(x,y,z) \;=\; E\,\psi(x,y,z)$$

The Hamiltonian here has been divided into the kinetic energy operator $\hat{K}$, and the potential energy $V(x, y, z)$. Note that the action of the potential energy operator is just multiplying the wave function by the potential energy! The aforementioned differential calculus is buried inside $\hat{K}$. In fact, there isn't a *single* Schrödinger equation. Rather, there's a different one for each form of the potential $V(x, y, z)$. This also means that the solutions $\psi(x, y, z)$ will be different for each potential.

Although the full time-independent Schrödinger equation is in fact a function in the full 3-D space that we live in, for much of what we do below we will simplify it and consider only one dimensional systems. This makes dealing with it conceptually simple, but does not obscure any of the essential physical results. Such systems can in fact be realistic. For example, if you consider a mass moving on a spring attached to a wall, that is essentially a one-dimensional system, as the mass moves only forwards and backwards along the direction the spring is oriented.

13.3 Free Particles & the de Broglie Wavelength

A particle is called a "free particle" if its potential is constant. That is, there are no potential energy wells or barriers anywhere. It's simplest to choose that constant potential energy to be zero, as that reduces the Schrödinger equation to:

$$\hat{K}\psi(x) \ = \ E\,\psi(x)$$

(in the one-dimensional case). Solutions to this equation are called "plane-wave" solutions. They are states with definite momentum $p = E^2/2m$ (which is exactly what you'd expect if you compare momentum and kinetic energy in classical physics). However, their position is completely undetermined; there is equal probability for *any* x, which is what you'd expect for a state of definite momentum given the Heisenberg Uncertainty Principle. The functional form of $\psi(x)$ is just a standard wave:

$$\psi(x) \ = \ A\,\cos(2\pi\,x/\lambda) \ \pm \ iA\,\sin(2\pi\,x/\lambda)$$

where A is a constant (a complex number) that normalizes the wave. The ± depends on whether the wave is moving to the right (i.e. momentum is in the $+x$ direction) or to the left (i.e. momentum is in the $-x$ direction). The normalization condition will only put a constraint on the absolute square of A, meaning that there will be many complex numbers that satisfy it. As such, there isn't one single solution to this equation; however, all of the solutions do give the same predictions for *measurable* things such as the probability of finding the electron at a given spot. The value λ that shows up in these equations is the *wavelength*; that is it's the range of x over which it takes the sine or the cosine to go through one complete cycle. Note that although $\psi(x)$ varies with space, the probability of finding x at any given position, $|\langle x\,|\,\psi\rangle|^2$, does not! See Section 13.7 for more details about this.

In this case, the energy levels are *not* quantized. E can be anything in the equation above. A different energy E does correspond to a different wavelength in the plane wave represented by ψ. In these solutions, the energy E of the particle is related to the wavelength λ of the wave function by:

$$E \;=\; \frac{h^2}{2\,m\,\lambda^2}$$

It's more traditional to express this wavelength, called the *de Broglie wavelength*, in terms of the momentum of the particle:

$$\lambda \;=\; \frac{h}{p}$$

You can get this equation directly from the previous equation by using the relationship $E = p^2/2m$, that results from the combination of kinetic energy $E = \frac{1}{2}mv^2$ and momentum $p = mv$. The constant h here is a version of Plank's Constant, related to $\hbar$ by $h \;=\; 2\pi\,\hbar$.

For example, what is the de Broglie wavelength of an electron moving at $1{\times}10^6$ m/s (a "typical" speed for an atomic electron)? We would plug the right numbers into this equation:

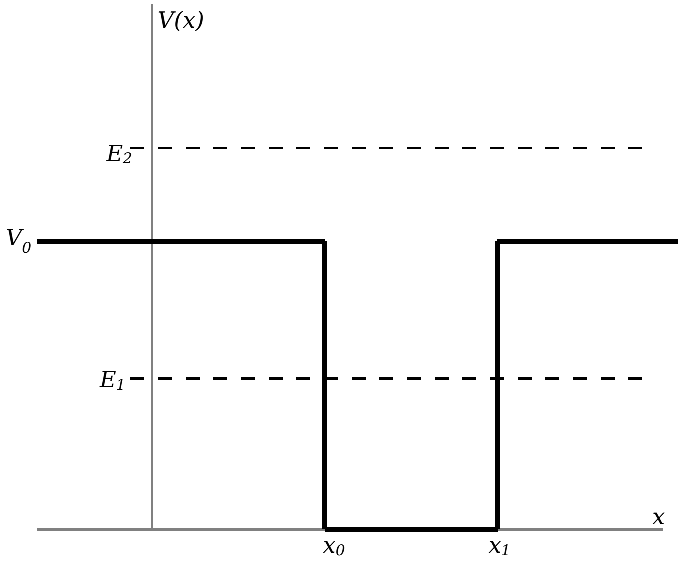
$$\begin{aligned}\lambda \;&=\; \frac{h}{p} \;=\; \frac{h}{mv} \\ &=\; \frac{6.626\times10^{-34}\,\mathrm{kg\,m^2\,s^{-1}}}{(9.109\times10^{-31}\,\mathrm{kg})(1\times10^{6}\,\mathrm{m\,s^{-1}})} \\ &=\; 7\times10^{-10}\,\mathrm{m} \;=\; 0.7\,\mathrm{nm}\end{aligned}$$

For comparison, this is about 1/1000 the wavelength of visible red light.

Many of the physical effects peculiar to quantum mechanics show up as wave interference between different components of a wave function $\psi(x)$. All waves, including those that derive from classical physics (such as waves on a string, sound waves, or electromagnetic (i.e. light) waves), show interference effects. The fact that the wave function, this abstract mathematical object which is used to figure out things about the state of a particle, also shows interference effects is what we mean when we say that sometimes particles behave like waves. In general, the longer the wavelength of a wave (i.e. the larger λ is), the easier it is to see interference effects. The de Broglie wavelength indicates that wavelength is inversely proportional to momentum. For a non-relativistic particle (which is implied here, as the Schrödinger equation assumes non-relativistic particles), $p = mv$. Thus, for particles moving at a given velocity, the larger m is, the smaller λ is. This is why it is so difficult to observe quantum interference effects for larger objects; the effective wavelength, and thus the typical separations that you'd need to see those effects, becomes tiny.

13.4 Quantized Energy Levels in Bound Systems

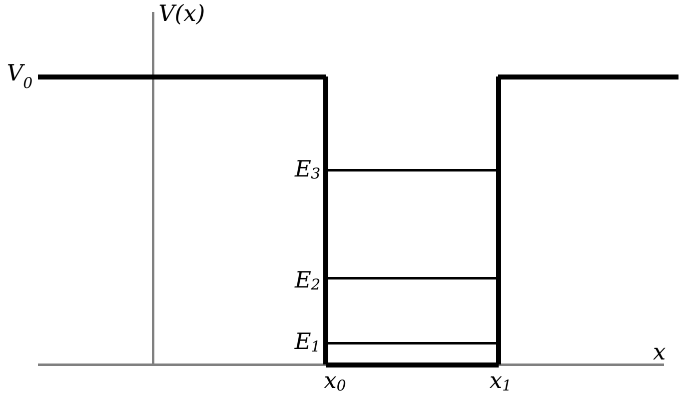
Whereas a free particle has a continuum of energy states available to it, if the particle is bound in a potential its available energy states are quantized. For a particle to be bound in a potential, the potential must limit a region of space that the particle could access classically. That is, if there is an energy well of some depth, the particle's energy over the bottom of the well must be *less* than the depth of the well:



A finite square potential well, with potential energy 0 chosen to be the bottom of the well. (Sometimes, one chooses the top of the well to be potential energy 0.) If a particle's total energy is less than the depth of the well V_0, as is the case for energy E_1 in the figure, that particle is bound in the well. If the particle's total energy is greater than the depth of the well, as is the case for energy E_2 in the figure, that particle is free, although its wave function is still influenced by the presence of the well.

In this example, a particle would be classically limited to the range of position $x_0 < x < x_1$ (see Section 2.2.2). Quantum mechanically, the particle is *most likely* to be found in that range, but it turns out there is a finite non-zero probability that the particle is found outside the classically allowed volume!

In this square well, there are a finite number of energy states available to a particle. Classically, a bound particle could have *any* energy $0 \leq E < V_0$. However, quantum mechanically, the particle *must* be in one of specific separated states. This is analogous to what we've seen with angular momentum of a spin-1/2 particle, where the projection of angular momentum along one axis must take on one of two specific separated values. For the square well, the allowed energy levels qualitatively looks something like:



In this example, there are only three allowed energy states, and the energies of those states are E_1, E_2, and E_3. Notice that the lowest allowed energy level is *not* 0! That is, you can't have a particle that's got no kinetic energy. This fits with the Heisenberg Uncertainty Principle. By binding the particle into the potential well, you've set constraints on the particle's position: it's most likely to be between x_0 and x_1. The particle isn't equally probable to be anywhere, so the uncertainty on the particle's position Δx is finite (and indeed will be something close to $x_1 - x_0$). As such, there must be a corresponding uncertainty in the particle's momentum Δp, so it's impossible for the particle to be in a zero-momentum state.

13.5 The Simple Harmonic Oscillator

One important potential energy function is the Simple Harmonic Oscillator, or SHO. This is the potential energy of a spring (so long as you don't stretch of squish the spring too much). It also turns out to be a decent approximation, at least for lower energy levels, for a number of quantum systems. One such system is the vibrational energy states of a Hydrogen molecule H_2. The form of this potential, in one dimension, is:

$$V(x) \;=\; \frac{1}{2} m\,\omega^2\,x^2$$

Here, m is the mass of the particle moving in the potential. ω is the "natural frequency of oscillation" for the potential; for a *classical* spring, it would correspond to $2\pi/T$, where T is the period of oscillations. (Of course, for a classical spring, the system could also have any energy!)

The solution to the one dimensional Schrödinger equation for this potential gives the following energies for the energy eigenstates:

$$E_n \;=\; (n+\frac{1}{2})\,\hbar\,\omega$$

where n is an integer $0, 1, 2, \ldots$. As written, this potential is an infinitely high potential ($V(x)$ just keeps going up as x gets farther and farther from 0.) As such, there are an infinite number of allowed energy levels. Of course, as an approximation to a real physical system, usually the approximation will get worse and worse as x gets farther and farther from 0, which means that the solutions less and less of a good approximation to the real energy system for higher and higher energy levels.

13.6 The Hydrogen Atom

The potential that an electron in a Hydrogen atom experiences results from the electrostatic interaction between the electron (which is negatively charged) and the proton (which is positively charged). (A Hydrogen nucleus is composed of a single proton.) It is traditional (and convenient) to choose the zero level of the potential to be when the electron is extremely far away from the proton. This means that the potential energy gets more and more negative as the electron gets closer and closer to the proton. The form of this potential is:

$$V(r) \;=\; -\frac{1}{4\pi\epsilon_0}\,\frac{e^2}{r}$$

In this equation, ϵ_0 is a fundamental constant related to the effective strength of the electromagnetic force; it's value is $8.854\times10^{-12}\,\mathrm{C}^2\,\mathrm{m}^{-1}\,\mathrm{J}$. The letter e indicates the elementary charge; it is the charge on the proton, and the absolute value of the charge on the electron. It's value is 1.602×10^{-19} C, where C is "Coulombs", the SI unit of charge. Finally, r indicates the distance between the electron and the proton. If we consider the proton to be at the origin, then $r = \sqrt{x^2+y^2+z^2}$, with (x, y, z) indicating the position of the electron. The Schrödinger equation for an electron in a Hydrogen atom is then:

$$\hat{K}\psi(\vec{r}) \;-\; \frac{1}{4\pi\epsilon_0}\,\frac{e^2}{r}\,\psi(\vec{r}) \;=\; E\,\psi(\vec{r})$$

Here, we've written $\psi(\vec{r})$ as a shorthand for $\psi(x, y, z)$; the vector form of r, $\vec{r}$, indicates that the wave function depends on the electron's displacement from the origin. In fact, it's more common and more practical to express the wave function in terms of spherical coordinates $(r\,\theta, \phi)$, where r is the distance from the origin, θ is the angle off of the z axis, and ϕ is the angle off of the x axis in a projection on to the x-y

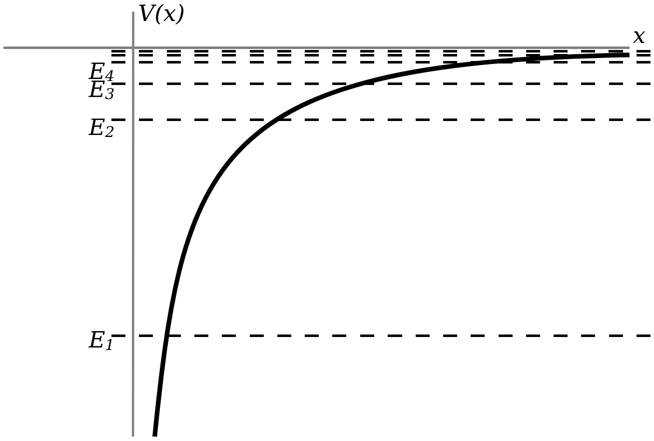
plane. (Even though the *potential* is spherically symmetric, in that it only depends on the distance from the origin, it turns out that the wave function solution is *not* always spherically symmetric. The reason for this is angular momentum, and that will be discussed in the next section.)

As with the other potentials discussed above, the allowed energy levels are quantized. In the Hydrogen atom, that energy level is specified by a *principle quantum number n*. The energy levels in a Hydrogen atom are:

$$E_n = \frac{-13.6\,\mathrm{eV}}{n^2}$$

The ground state of Hydrogen has an energy of −13.6 eV. That is, if you want to tear the electron off of the Hydrogen atom (a process known as *ionization*), you need to somehow provide at least 13.6 eV of energy in order to give the electron enough energy to make it out to extremely large distances away from the proton.

The figure below slows the Hydrogen atom potential and the first few energy levels. Higher and higher energy levels are less tightly bound (it takes less energy to ionize the atom, freeing the electron). They also get closer and closer together.



In fact, there are *three* quantum numbers associated with the solution to the Hydrogen atom. In addition to the principle quantum number n, there is also the total *orbital* angular momentum quantum number l, and the orbital z-angular momentum quantum number m. You could notate the energy eigenstates in a Hydrogen atom, corresponding to a state that the electron could actually be in, by $|n, l, m\rangle$. The ground state for a Hydrogen atom, in this notation, would be $|1, 0, 0\rangle$. For actual electrons, there's a fourth quantum number you have to specify: s, the z-spin of the electron. For every state n, l, m in a Hydrogen atom, there are in fact *two* different electron states, one where the electron has z-spin $+\hbar/2$, one where the electron has z-spin $-\hbar/2$. We'll call this quantum number s, and it will have either the value $+1/2$ or $-1/2$.

In Hydrogen, to very good approximation the energy of an eigenstate $|n, l, m\rangle$ only depends on the principal quantum number n, regardless of l and m. That is, the eigenvalue associated with states $|2, 1, 0\rangle$ and $|2, 0, 0\rangle$ are exactly the same (and are equal to $E = (-13.6\mathrm{eV})/n^2 = -3.4\,\mathrm{eV}$). In atoms with more than one electron, things get more complicated, and the energy of a given state may depend on the other quantum numbers.

13.7 Interpretation of the Wave Function $\psi(x)$

In general, it is best to view $\psi(x)$ the same way that you view $|\psi\rangle$. It's an abstract mathematical object that represents the state of the system. Quantum mechanics is then a theory, a mathematical model of reality that includes rules for manipulating $\psi(x)$ (or other representations of $|\psi\rangle$) in order to make predictions about the results of experiments, such as probabilities for observing particles in certain states, or expectation values for certain values.

It turns out that there is one particularly simple rule that can be applied to $\psi(x)$ in order to learn something about the state of the system. if $\psi(x)$ is a properly normalized single-particle wave function, then the construction $\psi^*(x)\psi(x)dx$ is the probability of finding that particle between position x and position $x + dx$, where dx is a small range of x. (By "small", we mean small enough that $\psi(x)$ does not appreciably change over the range.) As an example, consider the free particle wave function:

$$\psi(x) \;=\; A\; [\cos(2\pi\, x/\lambda) \;\pm\; i \sin(2\pi\, x/\lambda)]$$

If we want to find the probability for finding a particle at a given position, we multiply this function by its complex conjugate:

$$\begin{aligned}
\psi^*(x)\psi(x)dx \;&=\; A^*A \;\; \left[\cos\left(\tfrac{2\pi\, x}{\lambda}\right) \;\mp\; i \sin\left(\tfrac{2\pi\, x}{\lambda}\right)\right) \;\; \left(\cos\left(\tfrac{2\pi\, x}{\lambda}\right) \;\pm\; i \sin\left(\tfrac{2\pi\, x}{\lambda}\right)\right] \, dx \\
&=\; A^*A \;\; \left[\cos^2\left(\tfrac{2\pi\, x}{\lambda}\right) \;+\; \sin^2\left(\tfrac{2\pi\, x}{\lambda}\right)\right. \\
&\qquad\qquad \left.\pm\, i \cos\left(\tfrac{2\pi\, x}{\lambda}\right) \sin\left(\tfrac{2\pi\, x}{\lambda}\right) \;\mp\; i \cos\left(\tfrac{2\pi\, x}{\lambda}\right) \sin\left(\tfrac{2\pi\, x}{\lambda}\right)\right] \, dx \\
&=\; A^*A \;\; \left[\cos^2\left(\tfrac{2\pi\, x}{\lambda}\right) \;+\; \sin^2\left(\tfrac{2\pi\, x}{\lambda}\right)\right] \, dx
\end{aligned}$$

To simplify this further, we can use the trigonometric identity $\sin^2\phi + \cos^2\phi = 1$ (this applies for all ϕ). Thus, we are left with:

$$\psi^*(x)\psi(x)dx \;=\; |A|^2 \; dx$$

That is, the probability of finding a free particle at *any* x within a given range dx is always the same. This corresponds to an infinite uncertainty in position x, which is what we need given that this state has a definite momentum $p = h/\lambda$.

The construction $\psi^*(x)\psi(x)dx$ works for *any* one-dimensional wave function for calculating the probability of finding the particle at a given position. Using the three dimensional version of this construction on the solutions to the Hydrogen atom is what gives us the "electron cloud" diagrams you may have seen for electron orbitals. More about that in the next chapter.

Chapter 14

Atomic Orbitals

In this chapter, we will be talking primarily about the Hydrogen atom. However, a lot of what we are saying will be an approximation to other atoms. In the Hydrogen atom, there are just two particles interacting, an electron and a proton. The proton is more than a thousand times more massive than the electron. That means that as a result of their interaction, the electron moves far more than the proton does. As such, we can view it as an electron moving about in the potential of the proton. Except for the fact that it provides this potential, to first order we can ignore the proton, and just consider the electron as the quantum mechanical particle.

For other atoms, there are additional electrons. The energy levels that we have calculated for Hydrogen can very easily be adapted to take into account a nucleus with more protons— all you have to do is multiply all of the energy levels by Z^2, the square of the number of protons in the nucleus (which is also the positive charge of the nucleus). We can then *approximate* other atoms by putting electrons into all of these energy levels. Because electrons are fermions, we can only put two into any given orbital. (Two because there are two possible spin states for an electron.) However, this implicitly assumes that the electrons are interacting only with the nucleus, and not with each other. That approximation will allow us to get a lot of insight into the structure of (for example) the Periodic Table of the elements, but is too much of an approximation to be able to figure out precise energy levels.

14.1 The Schrödinger Equation

The one-dimensional Coulomb potential was presented in section 13.6. Of course, the real Hydrogen atom is three-dimensional. The Schrödinger equation that results is almost identical:

$$\hat{K}\,\psi(\vec{r}) \;+\; V(r)\,\psi(\vec{r}) \;=\; E\,\psi(\vec{r})$$

First, remember that when we say $\psi(\vec{r})$, that's a shorthand for $\psi(x, y, z)$. On the left, we've broken the Hamiltonian $\hat{H}$ into the kinetic energy part ($\hat{K}$) and the potential. Notice that in here, we don't have $V(x, y, z)$, but only $V(r)$. When we say r without the little vector arrow, we mean the distance from the origin, that is, $r = \sqrt{x^2 + y^2 + z^2}$. The potential here is *spherically symmetric*. Because it depends only on the origin, if you rotated the whole system through any angle, the potential would be no different.

If we substitute in the correct expression for the potential $V(r)$, this equation becomes:

$$\hat{K}\,\psi(\vec{r}) \;-\; \frac{1}{4\pi\,\epsilon_0}\frac{e^2}{r}\,\psi(\vec{r}) \;=\; E\,\psi(\vec{r})$$

As noted in the previous chapter, the potential is always negative. As r gets very large (i.e. the electron is very far away from the proton), the potential approaches zero. Again, this is just a convention; we could add any constant we wanted to the potential without changing the physics of what's happening. We've chosen this because it's convenient not to have to worry about the nuclei of atoms that are far away. The solution to this equation will be individual functions $\psi(\vec{r})$, each corresponding to a different allowed state, each with a corresponding energy level. The solutions that represent an atom— where the electron is bound to the proton— have $E < 0$. Classically, if a particle moves in this potential, that will set a maximum distance away from the origin that the particle could reach.

14.2 The Orbitals

The specific functions $\psi(\vec{r})$ (or $\psi(x, y, z)$, or $\psi(r, \theta, \phi)$ that work with the Schrödinger equation are the eigenfunctions of the Hamiltonian operator. Because we don't know necessarily when we start what those functions are, even though we've now specified the operator, we have to go through a process in order to figure out what the possible functions $\psi(\vec{r})$ are.[1] Therefore, we call those functions the "solutions" of the Schrödinger equation. The exact functional form of these solutions aren't particularly illuminating. However, they do have some general features, the existence of which underlie all of atomic structure. In the previous chapter, we talked about some solutions that could be described by energy levels. For instance, the solutions to the one-dimensional simple harmonic oscillator have evenly spaced energy levels indexed by an integer n. In three dimensions, it will take three different "quantum numbers" to index the solutions. The numbers we choose to represent the states are the ones that have a most direct physical interpretation.

---

[1] That process is much more advanced than what could be covered here and involves differential equations.

We call the solutions "orbitals" rather than just energy levels. In Hydrogen, there are usually multiple orbitals that share the same energy. In other atoms, fewer orbitals will share the same energy. These orbitals represent states available to the electron. Each state has an energy, a total angular momentum, and a z component of orbital angular momentum associated with it.[2] Those three values are the observables for which the orbitals are eigenstates. It's not surprising that the orbitals are energy eigenstates, because we produced them by finding states represented by wave functions that solve the Schrödinger equation, which *is* the energy eigenfunction equation. It is less obvious why they would be angular momentum eigenstates. It turns out that it's possible to break the kinetic energy operator into two parts, a "radial" part and a part that *is* the orbital angular momentum operator. That means that the Schrödinger equation includes the orbital angular momentum eigenvalue equation inside it.

These orbitals are *not*, however, eigenstates for position or for momentum. In particular, not being position eigenstates, electrons in atomic orbitals do not have definite position. Rather, there is a probability density for them to be at different positions, just as an electron in the $|{+z}\rangle$ spin state is not in a definite state of x spin and has probabilities to be found with positive and negative x spin. While the name "orbital" suggests that the electron is circling the nucleus in a manner analogous to how the Earth circles the Sun, this is not what is happening at all. The electron doesn't follow any particular path through space around the nucleus. Instead, the orbital is a probability cloud representing the effective amplitude for the electron to be found in any one tiny region of space around the nucleus. One consequence of this is that the negative charge associated with the electron is spread throughout this cloud. Whereas the Earth's mass is always at the position the Earth is in its orbit at any given moment, there is no single position for an electron in an atom, so there's no single place where the electron's charge is. The charge of the electron is more diffuse. Additionally, the center of the cloud is right at the nucleus. For the ground state, where the probability distribution is spherically symmetric, to something "far" from the atom (i.e. far enough away that the probability for the electron being found that far away or farther is negligible) the atom acts as if it were entirely neutral, with the effective charge of the electron being at exactly the same place as the effective charge of the proton.

Of the three quantum numbers that represent the electron orbitals, the quantum number n, the principle quantum number, is sometimes called the shell number. The

---

[2] Remember that if a system is in a definite state for the z component of angular momentum, it can not have a definite value for x and y angular momentum. This is true for orbital angular momentum as well as for the intrinsic spin of particles. Whereas in a classical system, x, y, and z components of angular momentum would give you three "degrees of freedom", three things that could be varied, the orthogonality of those observables in quantum mechanics means that you only have two: total angular momentum, and one component.

average distance of the electron from the nucleus is determined *primarily* by n. In a single-electron atom, to first order the energy of the atom is determined entirely by n. (There are second order effects, such as the magnetic interaction between the spin and orbit of the electron, that are beyond the scope of this class.) The ground state has $n = 1$, and higher shells have larger values of n.

The second quantum number, l, indexes the total orbital angular momentum of an electron in that state. It too must be a non-negative integer; that is, it can be 0 or a positive integer. The angular momentum represented by l does *not* include the spin angular momentum of the electron. Electrons remain electrons, and as such their total spin angular momentum is $\frac{\sqrt{3}}{2}\hbar$, and the possible projection along any axis is quantized to $+\hbar/2$ and $-\hbar/2$. The *total* angular momentum in the electron cloud of an atom depends on the orbital and spin angular momenta of all the electrons. Combining those angular momenta is fairly complicated, and involves taking into account the fact that electrons are indistinguishable particles. It turns out that for a given state, the angular momentum quantum number l must be less than the principle quantum number n. Thus, for the ground state, $n = 1$, we know that $l = 0$. The next shell out, $n = 2$, there are two possible values of l: $l = 0$ and $l = 1$. The orbital angular momentum associated with a given value of l is $\hbar\sqrt{l(l+1)}$. Notice that this means that the ground state orbital has zero orbital angular momentum! This highlights the degree to which these orbitals are not analogous to planets circling stars, for a planet circling a star assuredly does have orbital angular momentum.

The third quantum number, m, indexes the z projection of the angular momentum. Just as electron spin angular momentum is quantized, so is electron orbit angular momentum. However, there's a difference. Instead of having half-integral values, orbital angular momentum has integral values. As a spin-1/2 particle, we could say that the quantum number for the total spin of every electron is $s = 1/2$. The z projection, which we've called s_z, is then either $+\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$. We could say that there are quantum numbers, perhaps m_s, associated with electron spin that give the electron two possible spin states, one with $m_s = 1/2$, the other with $m_s = -1/2$.

In electron orbits, the total angular momentum l is an integer. m can be either positive or negative (representing angular momentum that's in the positive or negative z directions). Unlike electron spin, m can also be zero. However, m cannot get any larger than l. Thus, for a given value of l, there are $2l + 1$ possible projections: $m = -1$, $m = -1 + 1$, …, $m = 0$, …, $m = l - 1$, $m = l$.[3] The z component of orbital angular momentum associated with an orbital with quantum number m is just $m\hbar$. Notice that, just as with electron spin, it's impossible to have a z component of angular momentum that is equal to the total angular momentum of the state. Just like electron spin, the x, y, and z components of angular momentum are all represented

---

[3] So for $l = 1$, there are three possible values of m: -1, 0, and 1. For $l = 2$, there are five possible values of m: -2, -1, 0, 1, and 2. And so forth.

by non-commuting operators. That is, an orbital can only be in a definite state for one of them. Thus, while a classical particle would have multiple different states (effectively) for a given l and m (as there would be different x and y projections of angular momentum available), a quantum particle's orbit is *completely specified* by just the total angular momentum quantum number l, and the z projection quantum number m.

The solution to the Schrödinger equation provides a dazzling wealth of orbitals available to the electron in the Hydrogen atom. Strictly speaking, we've solved for the orbitals as energy eigenstates. This means that if an electron is in one of those states, in the absence of observations or interactions it will stay in that state. (If it's in an indeterminate energy state, if you somehow manage to measure the energy of the atom you will collapse the atom's wave vector and it will drop into an eigenstate.) However, observationally, atoms that are in excited states do not stay there forever. After a while, they will spontaneously decay, with the electron dropping down to a lower state, and eventually with the electron reaching the ground state (which *is* stable. This would imply that the atom must somehow be interacting with something, if it is able to change from one eigenstate to another. Indeed, it does; it is interacting with the electromagnetic field. Even if there isn't any light (i.e. any "excitations of the electromagnetic field") around us, the field is always there. What's more, there are always virtual photons, as a result of the energy/time version of Heisenberg's Uncertainty Principle (Section 11.4.1). The interaction of the atom with the electromagnetic field yields a probability in any given time interval that the atom may emit a photon and drop to a lower energy state. The energy of that photon corresponds exactly to the difference in energy between the upper and lower states of the transition.

14.3 Visualizing Orbitals

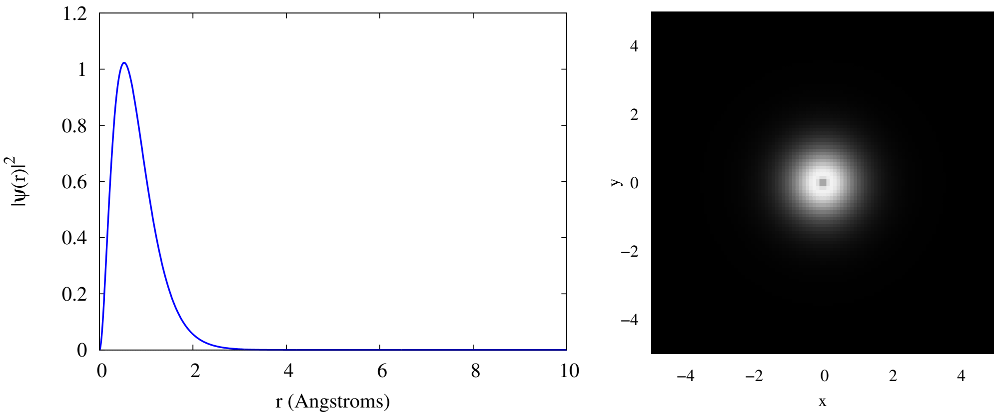
Historically, the energy available orbitals in an electron were identified by observing the wavelengths of light that would be emitted as the electron changed energy levels. Until we understood that light was quantized in photons, it was difficult to explain this as different energy levels. However, it was recognized that there were specific spectroscopic "lines" (i.e. wavelengths associated with a given transition) for each atom. The nomenclature that grew up around these lines is what gives us the nomenclature that we use today to name the various orbitals. Alas, the reasons for the letters that we use have little to do with the physics of the atom and everything to do with how the spectroscopic lines looked to the scientists who identified them. Each orbital is indicated by a form nx, where n is the principle quantum number and x is a letter corresponding to the angular momentum quantum number l. That letter is s for orbitals with $l = 0$; it is p for orbitals with $l = 1$; it is d for orbitals with $l = 2$;

and it is f for orbitals with $l = 3$.

14.3.1 s Orbitals

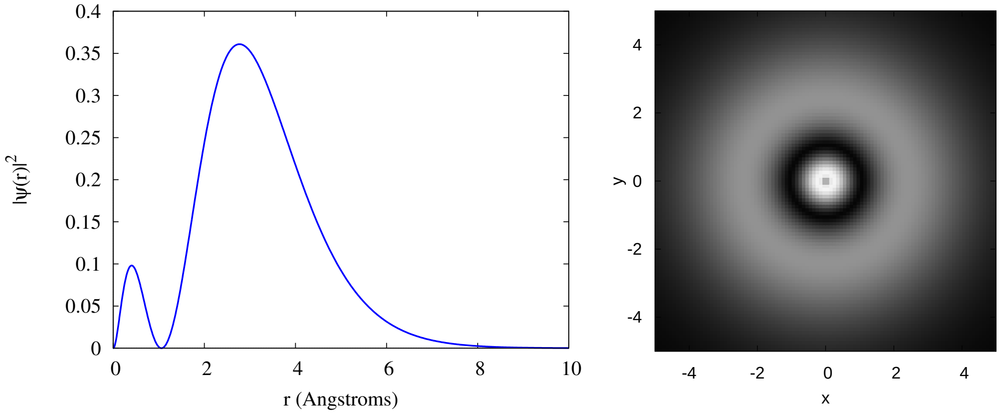
Orbits with $l = 0$ are called s orbitals. Although this is not where the letter comes from, it's useful to think of these as "spherical" orbitals, because they are spherically symmetric. However, they aren't just spheres! Again, remember that the probability cloud for the electron is a fuzzy ball around the nucleus, representing where the electron is likely to be found.

The plot below shows two visualizations of the 1s orbital. On the left is a plot of $\Psi^*(r)\Psi(r)$. This gives the *probability density* for the electron to be found at radius r. That is, you must pick a small range dr around the r you're interested in, and multiply this probability density by that dr. You then get the probability for finding the electron with that dr of your chosen r. On the right is a cut through the $x - z$ plane showing the probability density as a function of position. Lighter colors mean more probability of finding the electron at that position. Notice that there is a darker spot at the center. This corresponds to the probability dropping to zero at $r = 0$, as seen in the left plot. In both cases, distances are plotted in terms of Angstroms; one Angstrom is 10^{-10} m which, as you can see from the plot, is about the size of an atom.

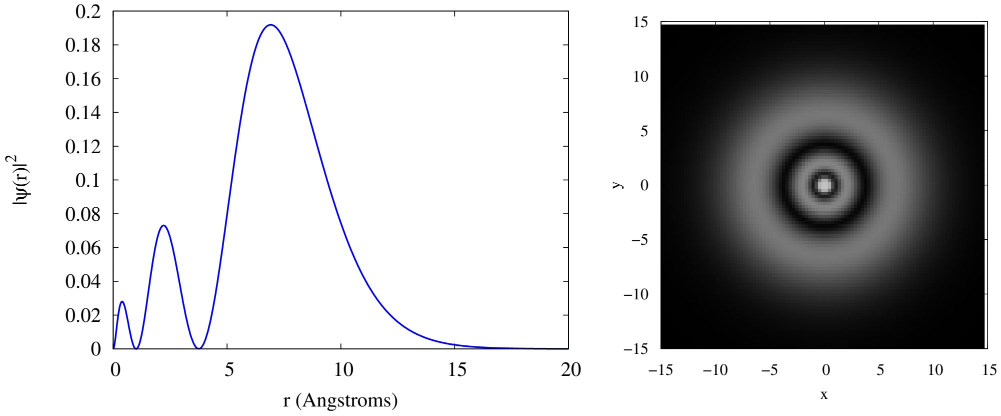


If you go to the 2s orbitals, an additional bump is added to the radial wave function. Also, the average distance the electron is from the center of the atom gets larger. While the probability clouds for a 1s and 2s orbital overlap, most of the probability for a 2s electron is outside most of the probability for a 1s electron. This means that to some extent, when working out the properties of an atom with two electrons in the 1s shell and one 2s electron (that would be Lithium), we can treat the nucleus plus the 1s shell as a single spherical ball of net charge +1. While this isn't perfect, this does lend some support to the approximation we'll make for multi-electron atoms that each electron is moving in a nuclear potential and not interfering too much with other electrons.

Below are the same two plots for the 2s orbital. The scale of the axes is the same as the scale used previously in the 1s orbital, so that you may compare the plots directly.

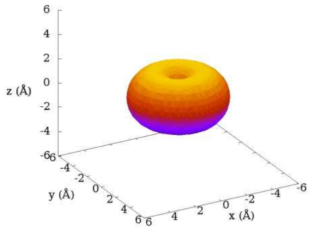


As we move to the 3s orbital, we have to expand the limits of our plots, as the electron is starting to have more and more probability to be at greater radius. In the plots below, you can see that the electron cloud still has reasonable probability density at a radius of 15 Angstroms. You can also see that the 3s orbital is three concentric fuzzy spherical shells; equivalently, the radial function has three bumps. Again, sizes are in Angstroms.

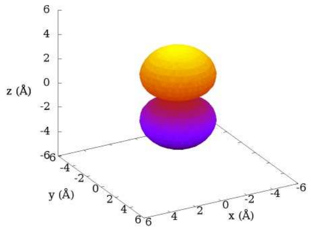


14.3.2 p Orbitals

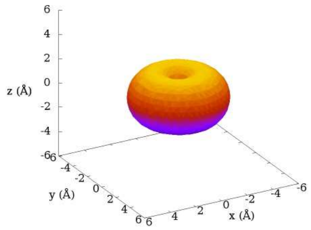
Orbitals with total angular momentum quantum number $l = 1$ are called p orbitals. Remember that l must always be less than n. As such, the first shell has no p orbitals; it only has s orbitals. This means that you can only put two electrons (with opposite spin) in the first shell. In the second shell, you can put eight total electrons. You can put two electrons in the 2s orbital, and six in the 2p orbital. Why six? For $l = 1$, there are three possible values for m, the quantum number that indexes the z component of angular momentum: $m = 1$, $m = 0$, and $m = -1$. Below is three plots showing what the $2p$ orbitals look like.



$m = +1$

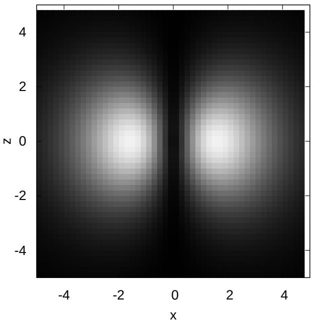


$m = 0$

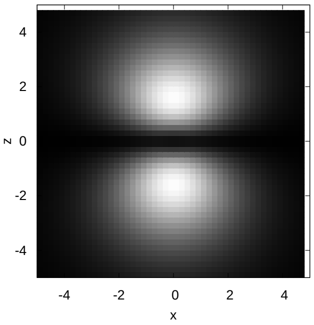


$m = -1$

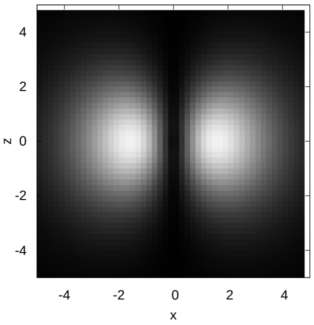
What we've drawn here is a surface of constant probability. In reality, the p-orbitals aren't hard shells, as this picture would seem to indicate. Rather, just as with the s-orbitals, they're fuzzy, with higher probability towards the "center" of the distribution (which may not be at the origin!) and less probability away from it. As a way of visualizing this, the plots below show a cut in the $x - z$ plane of the three orbitals depicted above:



$m = +1$



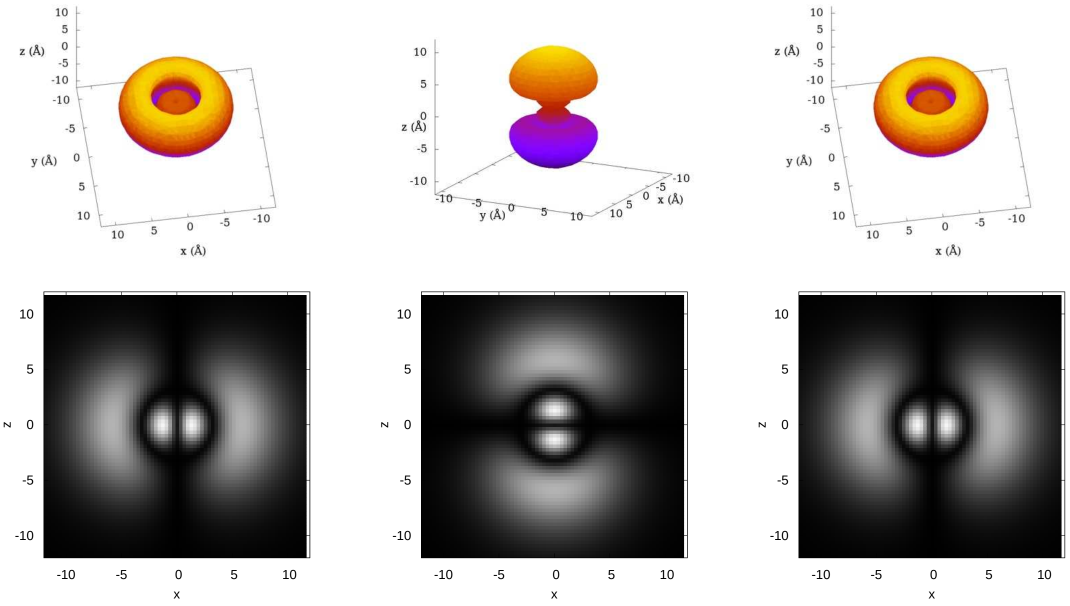
$m = 0$



$m = -1$

Notice that the $m = +1$ and $m = -1$ $2p$ orbitals look identical. If you imagine rotating the plot around a vertical axis through the center of the plot, both of those orbitals look like two lobes, one over the other. The $m = 0$ orbital looks different, however. If you imagine rotating it around a vertical axis, you get an orbital that looks like a thick donut.

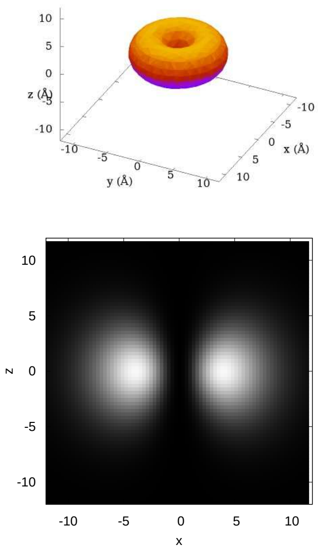
The p orbitals for higher values of n get more interesting. Just as the s orbitals become versions of themselves nested inside each other, the same thing happens with the p orbitals. Below are the three $3p$ orbitals:



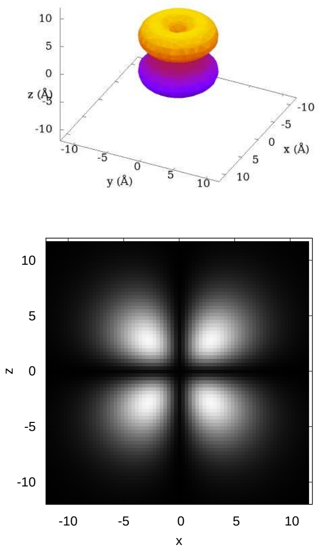
$m = +1$ $m = 0$ $m = -1$

14.3.3 *d* Orbitals

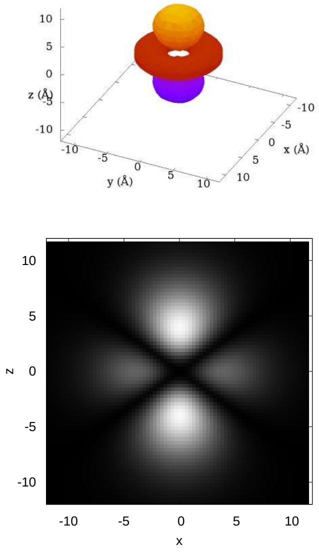
When $l = 2$, the orbitals that are the solutions of the Hydrogen atom Schrödinger equation are called d orbitals. These orbitals only exist for shells with $n = 3$ and greater, again because l must be less than n. As we saw with the p orbitals, the probability density for the electron in space is the same for $+m$ and $-m$. As such, we'll only plot the positive-m versions of the orbitals. As before, in addition to a 3d plot showing "shells" at a constant probability level, there is a 2d plot showing a cut in the x-z plane.



$m = 2$

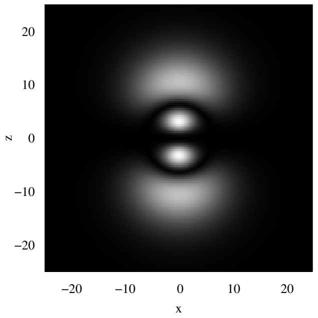


$m = 1$

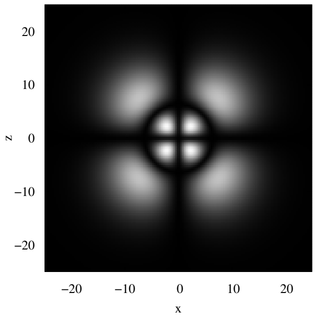


$m = 0$

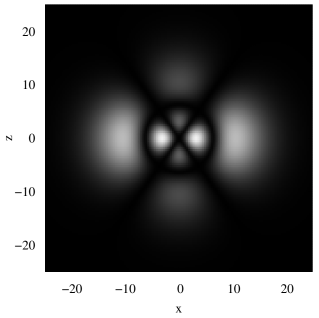
As with the p orbitals, as we go to the d orbitals in higher shells they get more interesting. Plotted below are the cuts through the x-z plane for the $4d$ orbitals:



$m = 2$



$m = 1$



$m = 0$

Chapter 15

The Periodic Table of the Elements

In Section 13.4, we saw the example of a square potential well. In that example, there were three bound states. This is the solution for a single electron moving in that potential, so each of these levels represents a possible energy state available to that electron. There are in fact six total states available, because there are two spin states available to an electron as well. Under the approximation that electrons do not interact with each other, we could take exactly these energy levels and put up to six electrons in the potential well. Because electrons are fermions, no two electrons can be in the same state. (If we were putting bosons into the potential well, there would be no limit, as you can put multiple bosons in one state.)

The process of constructing the periodic table of the elements is similar to the process of filling up this square well with electrons. In the previous chapter, we saw that the states available to an electron are indexed by three quantum numbers: n, the principle quantum number, l, the total orbital angular momentum quantum number, and m, the quantum number indexing the z projection of orbital angular momentum. In addition, there is electron spin, allowing two electrons to go into each $|n, l, m\rangle$ state.

On the periodic table, the "atomic number", usually indicated as the largest number in a display and often represented with the letter Z, is the total number of protons in the nucleus of the atom. The charge on the proton is exactly opposite the charge on the electron; whereas electrons have a charge of -1.602×10^{-19} C, protons have a charge of $+1.602 \times 10^{-19}$ C. Thus, for a neutral atom, the number of electrons is equal to the number of protons. Chemistry is all about the dynamics of electrons as atoms interact with each other, form bonds, trade electrons, and so forth. Therefore, from a chemical point of view, it might be more useful to think of Z as the number of electrons in a neutron atom of an element. (What if there is an additional electron added, making the atom negative, or if there is an electron removed, making the atom positive? In that case, we call it an ion, but we still name the ion based on the number of protons. A Chlorine atom with an extra electron would be called a negative Chlorine ion.)

Elements are constructed by filling in electron states until the number of electrons matches the number of protons in a nucleus. The number of states available at each shell is dictated by how angular momentum functions under quantum mechanics: as we saw in the previous chapter, l must be less than n, and m varies from $-l$ through l. If you put these two things together, it would be fair to say that angular momentum at the quantum level is responsible for the structure of the periodic table of the elements, the chemical properties of the different elements, and thus for chemistry and life as we know it. The number of states available is influenced, for instance, by the fact that angular momentum can only have definite states for projection along one axis at a time. The structure of the periodic table would be very different if x, y, and z angular momentum operators all commuted.

15.1 Interacting Electrons, Energy Levels, & Filled Shells

In fact, electrons *do* interact with each other. In the previous chapter, we made arguments that these interactions should be smaller than the interaction with the nucleus. Because electron probability clouds are spread out, and outer shell clouds only have relatively small overlap with inner shell clouds, often, especially when viewing inner shells, you can approximate them as just lowering the net effective charge of the proton. That is, if you look at a Sodium atom, it has 11 electrons. The first 10 electrons will fill up the $1s$, $2s$, and $2p$ states. That leaves the outermost electron in the $3s$ state. Because there isn't a whole lot of probability for that $3s$ electron to be found where the inner electrons are usually found, you could approximate the situation for that outer electron that it's orbiting a ball of charge with a net charge of $+1$ (in atomic units), neglecting the fact that that charge is made up if $+11$ in the tiny nucleus and -10 in the outer electron cloud. However, even though interactions between electrons are secondary to the interaction between each electron and the nucleus, they are there, and they do ultimately have a lot of influence as to how elements at different places on the periodic table behave.

One of the primary effects of electron interactions is that the s, p, and d orbitals for a given value of n are *not* at exactly the same energy. In a Hydrogen atom— or any ion that only has one electron— they are, to a fairly good approximation. If there is more than one electron, however, the electron-electron interactions modify the energies of these states. In general, levels with higher l will be higher energy states than levels with lower l but the same n. In the absence of something external (such as a magnetic field), levels of different m but at the same n and l will still have approximately the same energy. Sometimes, you will find levels with a higher n but a lower l to be at a lower energy level than levels with a lower n and higher l. For instance, the $4p$ states tend to be filled before the $3d$ states. This isn't always a hard and fast rule; sometimes you will see the states filled out of the "standard" order. The interactions between electrons make the entire system a many-body system, and many-body systems are often notoriously difficult to solve in Physics.

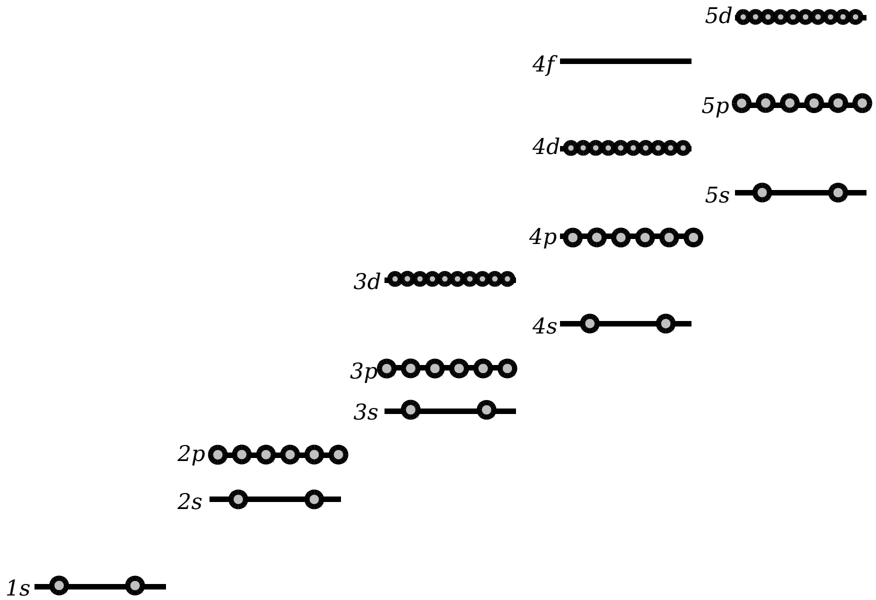
For the most part, atoms are "happiest" (if you will allow for some anthropomorphization for purposes of discussion) if the number of electrons equals the number of protons. If there is one too many electrons, the ion will generally be happy to give away one of its negative electrons to the first positive charge that goes along. Likewise, if there is one too few electrons, the ion has an extra positive charge, and will tend to snap up any spare electrons in its vicinity.

However, this is not the only consideration for atom happiness. Atoms also like to have a filled shell. That is, Helium is more chemically stable than Hydrogen, because whereas Hydrogen only has one of two possible electrons in the $1s$ state, Helium has entirely filled the $n = 1$ shell by placing two electrons in the $1s$ state. Likewise, Neon, with 10 electrons, has filled up both $1s$ states, both $2s$ states, and all six $2p$ states, making it a very chemically stable element. The elements down the right column of the Periodic Table are called "noble gasses". They are so called because they are chemically stable, and don't tend to interact with other atoms or form molecules. (They're noble, and thus above it all, or some such. Doubtless sociologists of science love to tear apart this nomenclature to display cultural bias in scientists.) The reason they are so stable is that each one of these noble gasses is an element that has just completely filled a set of p orbitals. (The one exception is Helium. It has completely filled the $n = 1$ shell, where there are no p orbitals.) Ne has completely filled its set of $2p$ orbitals. Ar has completely filled its set of $3p$ orbitals. Kr has completely filled its set of $4p$ orbitals. And so forth.

You can get a first guess at the chemical properties of an element by comparing how close it is to a noble gas. If an element has just one or two electrons more than a noble gas, the easiest way for it to be more like a noble gas would be for it to lose an extra electron. Elements like these are more apt to form positive ions than negative ions. An example is Sodium. Sodium has atomic number 11. The first 10 electrons fill up the $1s$, $2s$, and $3p$ orbitals; that is, they're like a Neon inner core. Then, just outside that, is a single $3s$ electron. If Sodium loses that electron, then it is electrically positive, but now it has a happy noble-gas-like electron configuration. In contrast, Chlorine has 2 electrons in the $3s$ shell and 5 electrons in the $3p$ shell. All it needs is one more electron to have a full $3p$ shell, giving it the electronic configuration of Krypton. If you put these two elements together, each Cl atom will tend to take away an electron from each Na atom, leaving the Cl a negative ion and the Na a positive ion. Those two ions then will have an electrostatic attraction towards each other as a result of their opposite charges. The result is a crystal, Sodium Chloride, more commonly known as salt. In this case, the bonds holding the crystal together are "ionic bonds". In most molecular bonds, an electron is shared between elements. In this case, however, the Sodium is so eager to get rid of an electron and the Chlorine is so greedy for another one that effectively the electron transfers all the way across from the Na to the Cl.

15.2 Filling Up Orbitals

You can use the diagram below to figure out where an element will fall on the periodic table. Start with the lowest energy states, and fill in available states with electrons. Keep filling them in until you have as many electrons as you need. If you have just a single spare electron in an s shell, then that is an element like Hydrogen, Lithium, or Sodium; it goes on the left column of the periodic table. As you fill in more and more levels, you move farther and farther to the right on the table. If you completely fill a p orbital, then you're on the very right side of the table, and the element you've constructed is a noble gas.



States available to electrons in atoms. States higher in the diagram are (usually!) states at higher energy, although the exact spacing on the diagram should not be interpreted as meaningful. Although not drawn, there are 14 states available in the $4f$ orbital. Additionally, to fill out the current modern periodic table, we'd need to extend this to include the $5f$, $6s$, $6p$, $7s$, and $7p$ orbitals.

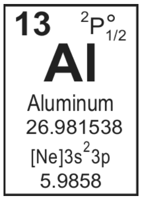
Because the chemical properties of an element are approximately determined by how far away it is from a noble gas, constructing the periodic table in this manner means that every column should have somewhat similar chemical properties. Thus, each column on the periodic table is called a "Group". Each row on the periodic table is called a "Period". The name "Periodic Table" is not arbitrary. When something is periodic, it means that it has regular cycles. The period of the Earth's orbit around the Sun is about 365 days; after that much time, the Earth is back where it started. On the periodic table, it's element numbers that are periodic. After you add just enough electrons to fill up a p orbital, the next electron you add is as if you were starting over on a new shell.

However, there is a difference with the Periodic Table. In a simple orbit such as the Earth going about the Sun, each period is exactly the same length. However, on the Periodic Table of the Elements, periods get longer and longer. As you go to higher and higher values of n, there are more and more states available. At a given n, you can have values of l between 0 and $n-1$. Thus, for $n=1$, there are only two states available, and thus the first period has only two elements: H and He. At $n=2$, there are now eight states available (two s states and six p states), so there are eight elements in the period: Li, Be, B, C, N, O, F, and Ne. You might then expect there to be 18 elements in the third period, as for $n=3$ there are 18 states: two s states, six p states, and ten d states. However, it turns out that the $4s$ states are at a lower energy level than the $3d$ states. Thus, the third period only fills up the $3s$ and $3d$ states, and has eight elements just like the second period. In the fourth period, starting with Potassium, we fill in the $4d$, $3d$, and $4p$ states (*approximately* in that order), and now have eighteen elements. The same thing happens with f orbitals; it isn't until the sixth period, after the $6s$ states are filled, that the $4f$ orbitals start to get filled.

The fact that as you go to higher and higher periods, there are more and more states available before you can completely fill a p orbital, is what gives the periodic table its iconic "stepped" structure. The first step comes after Hydrogen and Helium. The first period only has the $1s$ orbital available, and only has two elements in it. The second and third periods each have 8 elements, filling up the $2s$, $2p$, $3s$, and $3p$ orbitals. The fourth period now has 18 elements in it, because in addition to the $4s$ and $4p$ orbitals, it also has to fill up the $3d$ orbitals.

15.3 Reading a Periodic Table

If you look at a periodic table, there is a variety of information you may find on it. Every periodic table includes the symbol of the element (one or two big bold letters at the center of the element's box), and the atomic number of the element (the number of protons in that element, usually shown as a number in the upper left). For example, below is the entry for Aluminum from the NIST Periodic Table (Dragoset et al., 2003).



13 ${}^{2}P^{\circ}_{1/2}$

Al

Aluminum

26.981538

$[Ne]3s^{2}3p$

5.9858

Here, you can see the symbol for the element is Al, and the atomic number is 13.

Usually, below the element symbol, you will find the atomic weight of the element. This is in units of “atomic mass units” or amu; one amu is equal to 1.66×10^{-27} kg. In this example, the atomic weight is given to eight significant figures, and is 26.981538 amu. The atomic weight in amu is *approximately* the number of protons and neutrons— which, together are just called “nucleons”— in the element, but there are a number of complications. First is the complication that for some elements, there are multiple *isotopes*. Different isotopes have the same number of protons but different numbers of neutrons. For example, the atomic weight of Chlorine is 35.45 amu. This is largely because in nature, we find Chlorine in two isotopes: Cl-35 and Cl-37, with 35 and 37 total nucleons respectively. There is a second complication, however. The mass of a nucleus is *not* exactly equal to the sum of the masses of the protons and neutrons that compose it! Each nucleus has what’s called a *binding energy*. This binding energy is equivalent to the 13.6eV of energy that holds an electron on to a Hydrogen atom. It is the total energy for all of the nucleons in their bound states, and is negative for a stable nucleus. This binding energy is taken away from the effective mass of the nucleus, using the conversion $E = mc^2$. In fact, exactly the same thing is true for atoms! However, the binding energy compared to the mass of atoms is something like one part in a billion, so as such when dealing with chemical reactions and other electronic transitions, we can approximate mass as being conserved. Nuclear binding energy can get up to a few percent of the total mass. (It is this difference that makes nuclear power so much more efficient, in terms of energy produced per mass of fuel used, than chemical power.)

15.3.1 Electronic Configuration

Often, but not always, the periodic table will include one or two sets of symbols intended to convey information about the ground state of the atom’s electron cloud. The one seen more rarely includes information about the spin, orbital, and total angular momentum of the atom. In the NIST entry for Aluminum above, you can see this information in the upper right as ${}^{2}\mathrm{P}^{\mathrm{O}}_{1/2}$. The letter in the middle represents the total orbital angular momentum of all of the electrons the atom put together; S means $l = 0$, P means $l = 1$, D means $l = 2$, and F means $l = 3$. For Hydrogen, the single electron is in the $1s$ orbital, which has no orbital angular momentum, so that letter is S. For Aluminum, the letter is P. This results from the single electron in the $3p$ state. There are two electrons in each of the $1s$, $2s$, and $3s$ states, and none of them have any orbital angular momentum. However, for Aluminum, there are also six electrons in the $2p$ state. But, because that shell is filled, there will be as many electrons with z-angular momentum of $+\hbar$ as there are with $-\hbar$ (two each, in this case), so all of their orbital angular momentum cancels out.

The superscripted number before the letter tells you about the electron spin state. It is equal to $2s + 1$, where s is the *net* electron spin divided by $\hbar$. For Hydrogen, this is ${}^2S_{1/2}$. In Hydrogen, the net electron spin is $1/2$, because there is just one electron that has spin $+\hbar/2$, so $2s = 1$. In Helium, that number is 1, because the two electrons have spins in the opposite directions, so $s = 0$. For aluminum, the net electron spin is the result of a single electron in the $2p$ state (as all of the filled states will have as many spin up as spin down electrons, thereby cancelling out each other's angular momentum). Thus, the net electron spin is $1/2$ (as always in units of $\hbar$), so the number we see in the example above is $(2)(1/2) + 1 = 2$. Finally, the subscripted number after the letter is J, the quantum number associated with the *total* electronic angular momentum of the atom. J represents a combination of orbital and spin angular momentum for the electrons. For Hydrogen, $J = 1/2$, because the angular momentum is entirely in the spin of the electron; for Aluminum, it's also $J = 1/2$, but the reason is more complicated. Both the spin and orbit of the $3p$ electron contribute, but it would have been possible for them to combine yielding either $J = 1/2$ or $J = 3/2$. For Helium $J = 0$ because there is no net angular momentum: there's no orbital angular momentum for two electrons in the $1s$ state, and the two spins cancel each other out. (As a caution, adding angular momenta in quantum mechanics can become complicated for cases with higher numbers than these examples.[1]) Similar to what we see with orbital angular momentum, the physical amount of angular momentum for an atom with total orbital quantum number J is $\hbar\sqrt{J(J+1)}$.

(The superscripted O— it's a capital O, not a zero— on the notation you see for Aluminum indicates that Aluminum has "odd parity". Parity is another quantum

---

[1] For example, with Aluminum, it turns out that for the total spin+orbit angular momentum of the $3p$ electron to be in a definite state $J = 1/2$, neither the orbital z component nor the spin z component may individually be in definite states. If we write the state of the $3p$ electron as $|m, s_z\rangle$, for an Aluminum atom to have $J = 1/2$ and $J_z = 1/2$ (i.e. we've specified the orientation as well as the total angular momentum of our Aluminum atom), then the angular momentum state of the outermost electron would be

$$\sqrt{\frac{2}{3}}\,|{+1}, -1/2\rangle \;-\; \sqrt{\frac{1}{3}}\,|0, +1/2\rangle$$

.

property that you don't need to worry about here.)

You may also see a series of letters and numbers that tell you how many electrons there are in each orbital. For Hydrogen, this is $1s^1$. That is, there is but a single electron in the $1s$ orbital. For Helium, it's $1s^2$: there are two electrons in the $1s$ orbital. For Nitrogen, it's $1s^2 2s^2 2p^3$. The $1s$ and $1s$ orbitals are filled, and the $2p$ orbital is half-filled. For periods after the second, it's often conventional not to list the full state, but to list the noble gas that has the same configuration as the inner core of electrons, and then just the states of the electrons outside of that. In our example of Aluminum above, the configuration is $[\mathrm{Ne}]3s^2 3p$. It's got all the electrons that Neon does— $1s^2 2s^2 2p^6$— plus an additional two electrons in the $3s$ state and one in the 3p state.

The number at the very bottom of the box for Aluminum, 5.9858, is the *ionization potential* in eV for Aluminum. Many periodic tables will not include this number. This is the amount of energy it takes to remove one electron from the atom, forming a positive ion.

Chapter 16

Matter

We've concentrated primarily on electrons throughout this course. Indeed, in our everyday life, it is the interactions of electrons that, together with photons (light), drive most of what we do. In this final chapter, we'll peer down inside the atom to see what the most fundamental particles are, and then extend our view out to bulk states comprised of large numbers of electrons.

16.1 The Standard Model of Particle Physics

Our best current understanding of Physics at the most basic level is that it is composed of a number of fundamental particles. These particles are, as best we can tell, points, much like electrons (which are in fact one of the fundamental particles). They have various properties associated with them, including mass, spin (angular momentum), electric charge, and others. It is from these fundamental particles that all of the matter we interact with is built. (However, matter built from these particles only makes up 5% of the total density of the Universe! See Section 16.7.)

Broadly, we can divide the particles into two categories: fermions and bosons. Matter is built from fermions, and the interactions between matter— the four forces— are carried by bosons.

The fermions are divided into three generations; it is only the first generation that we ever interact with on a daily basis. Each generation includes two *quarks* and two *leptons*. In the first generation, the quarks are called the "up" and "down" quarks. The two leptons are the familiar electron, and the electron neutrino. Each of the other two generations has an additional two quarks, heavier than the quarks in the first generation; and an additional two leptons, including a heavier version of the electron, and a corresponding neutrino. Additionally, for each fermion, there is a corresponding antiparticle. For example, the anti-electron is called the "positron". Antimatter is rare. In the very early Universe, we believe that matter and antimatter were present in equal quantities, but for some reason that is not understood matter had a *slight* advantage. A particle of antimatter, when it meets its corresponding matter particle, will mutually annihilate both, releasing their mass as energy according to the conversion $E = mc^2$. That slight advantage that matter had in the very early Universe is what's left over to build up stars, galaxies, and us today.

Generation	**Fermion**	**Symbol**	**Charge**	**Spin**	mc^2
First	Down Quark	d	-1/3	1/2	5.05 MeV
	Up Quark	u	+2/3	1/2	2.49 MeV
	Electron	e	-1	1/2	0.511 MeV
	Electron Neutrino	ν_e	0	1/2	*
Second	Strange Quark	s	-1/3	1/2	101 MeV
	Charm Quark	c	+2/3	1/2	1.27 GeV
	Muon	μ	-1	1/2	106 MeV
	Muon Neutrino	ν_μ	0	1/2	*
Third	Bottom Quark	b	-1/3	1/2	4 GeV
	Top Quark	t	+2/3	1/2	172 GeV
	Tauon	τ	-1	1/2	1.78 GeV
	Tau Neutrino	ν_τ	0	1/2	*

Table of fermions from the standard model of particle physics. All data is from the PDG (Nakamura and Particle Data Group, 2010). Charges are in units of e, the elementary charge. *: There are three neutrinos. Although masses are not known, we have limits that the heaviest one is less than 18.2 MeV, the middle one is less than 0.19 MeV, and the lightest one is less than 1.3 eV. However, the mass eigenstates and the flavor (i.e. "type of neutrino") eigenstates of the neutrinos are not the same, so it's impossible to identify a given mass with a given type of neutrino.

Quarks are never observed in isolation. For the most part, they are observed in bound states called *hadrons*. The proton and neutron are two three-quark particles. There is a huge zoo of additional hadronic particles, including baryons (made of three quarks) and mesons (made up of a quark and an antiquark). The proton and the neutron are the only stable hadrons. Indeed, even the neutron is not stable unless it's bound into an atom; a free neutron will decay to a proton, an electron, and an antineutrino in about 15 minutes. "Virtual" mesons are found inside the nucleus. Transient mesons and other sorts of baryons are made in particle accelerators, and also when cosmic rays hit the Earth's atmosphere.

In addition to the fermions that make up matter, there are four forces through which matter interacts. The most familiar of these forces are gravity and the electromagnetic force. In our current theories of physics, gravity is described by General Relativity, and is *not* included in the Standard Model of Particle Physics. We believe that we will one day be able to produce a working theory of quantum gravity, but we have yet to successfully do that. We expect this theory to include the graviton as a massless spin-2 boson.

The electromagnetic force is the best understood of the four forces. It unifies the electrostatic force and the magnetic force. The photon is the particle that carries the electromagnetic force; we see it as light. Radio waves, infrared radiation, ultraviolet radiation, x-rays, and gamma rays are all forms of light at wavelengths different from those our eye can detect. All of these are made up of photons, the quanta of the electromagnetic field. It is the electromagnetic force that forms the potential in which electrons move in atoms, and it is that interaction that governs the interactions between atoms.

Because the photon and the (presumed) graviton are massless, both gravity and electromagnetism are long-range forces. In contrast, the other two forces are short ranged. The strong nuclear force is the force that binds quarks together into protons and neutrons, and that ultimately binds protons and neutrons together into nuclei. The bosons that carry the strong nuclear force are called gluons. They may be massless, although a moderate mass isn't ruled out. However, other properties of the strong force limit it to a short-range force. The weak nuclear force is, as its name suggests, much weaker than the strong nuclear force, and has only a secondary effect in nuclei. The charge carriers of the weak force are indeed massive, limiting it to a short range force. The weak force is the *only* force other than gravity that interacts with *all* of the particles in the standard model. Neutrinos, in particular, only interact via the weak force, making them extremely hard to detect. The weak force is responsible for much radioactive decay; it is a result of the weak force, for instance, that a free neutron will decay into a proton, an electron, and an antineutrino. The bosons that carry the weak force are called "intermediate vector bosons." Their name is the same as their symbol. There are three, the W^+, W^-, and Z^0 bosons.

Force	**Boson**	**Symbol**	**Charge**	**Spin**	**Mass**
Electromagnetic	Photon	γ	0	1	0
Strong	Gluon	g	0	1	0?
Weak	W^+ Boson	W^+	$+1$	1	80 GeV
	W^- Boson	W^-	-1	1	80 GeV
	Z Boson	Z^0	0	1	91 GeV

Table of bosons in the standard model of particle physics. All charges are in units of e, the elementary charge. Not included is the hypothesized graviton, which would be a massless, chargeless, spin-2 boson.

16.2 Nuclei and Atoms

Quarks bind together to make protons and neutrons. A proton is composed of two up quarks and a down quark, and a neutron is composed of two down quarks and an up quark. Together, protons and neutrons are called nucleons. Nucleons can themselves bind together to make nuclei. These nuclei are always positively charged, with the total charge depending on the total number of protons. Nuclei are so called because they sit at the nucleus of atoms; an atom is a nucleus that has gathered negative electrons into the various electron states (i.e. orbitals) allowed by the electric potential created by its positive charge. The smallest nucleus is the Hydrogen nucleus, composed of but a single proton; it is about 10^{-15} m across. The largest stable nuclei have a more than 200 nucleons in them. Nuclei with more nucleons than that tend to be unstable, and spontaneously fission into smaller nuclei.

The *binding energy*— that is, the potential energy that results from combining nucleons together to make nuclei— is an appreciable fraction (1% or so) of the mass energy of the nucleons. The nucleus with the greatest binding energy per nucleon is Iron-56, which makes it (in a sense) the most stable nucleus. You can get energy out by fusing lighter elements together until you get to Iron-56; after that, it costs energy to build up heavier elements.

Of course, in nature, most materials are mostly electrically neutral, at least on Earth. It turns out that most of the baryonic material in space is in the form of plasma (mostly ionized Hydrogen, i.e. free protons and free electrons) filling the void between galaxies inside galaxy clusters. On Earth, though, for the most part if there's a free electron, it will be captured by the first nucleus that comes by with an extra positive charge. Hence, in our everyday experience, all things are composed of atoms. We organize our understanding of the various different types of atoms via the Periodic Table, as was discussed at length in Chapter 15.

16.3 Molecules

Atoms can bond together. Sometimes, if one atom is able to completely steal an electron from another atom (as is the case with Chlorine and Sodium atoms, where a Sodium atom will donate an electron to a Chlorine atom), the resulting ions will then stick together as a result of the electrostatic attraction between their opposite net charge.

More common, however, are molecules made from what is called covalent bonds. The electrons in the outer (unfilled) shell of an atom are known as "valence" electrons. Depending on the electronic configuration of an atom, it will have one or more effective valence electrons. In a molecule, the valence electrons are no longer associated with a single atom, but instead are shared between the electrons. In terms of the quantum mechanics involved, you wouldn't find a solution to Schrödinger's Equation for just the potential of one atom. Rather, you create a joint potential for the two atoms (including the effects of inner-shell electrons), and determine a solution for the system as a whole. The result is an electron wave function that indicates the electron probability cloud is shared between two or (for more complicated molecules) more of the atoms that composes the molecule. Just as nuclei have a binding energy, molecules have a binding energy, meaning that it is a lower energy state for these atoms to bind together and share an electron than it is for them to stay separate. Although this binding energy is typically a billionth of the mass energy of atoms, it is enough to create the vast majority of energy producing processes (e.g. burning gas to power a car) that we are familiar with in our everyday lives.

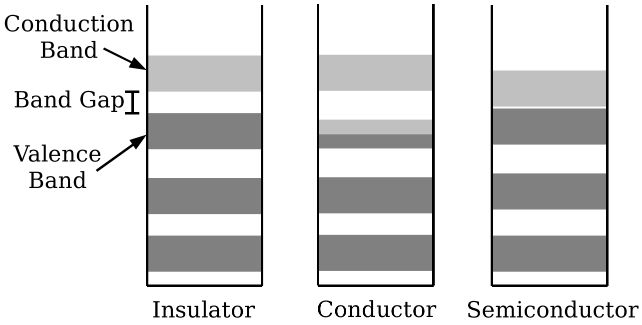
Finding these solutions to multi-atom potentials is an extremely difficult problem, and cannot be solved analytically (as the Hydrogen atom may be). Describing the quantum mechanical state and electron orbitals of any molecule more complicated than something like H_2 generally involves both heavy-duty numerical calculations on computers and heavy-duty quantum chemists.

16.4 Solids

Roughly speaking, a solid is when a large collection of molecules are held together and fixed in place. They aren't completely still, unless a solid is at absolute zero temperature. (And that's not possible, as a result of Heisenberg's Uncertainty Principle.) Most solids are at a higher temperature; the solids you interact with every day are approximately at room temperature, about 20°C or 290 K. At such a temperature, the molecules are vibrating about, each with about 1/40 eV of kinetic energy in that vibration. Solids are held together different ways. Some solids form crystals, where every atom is bonded one way or another to neighboring atoms. Other solids do not strictly have covalent or ionic bonds holding everything together, but are held together by intermolecular forces resulting from the attraction of one electron cloud to a neighboring nucleus and the like.

In some particularly interesting solids, valence electrons don't end up just being shared between neighboring atoms in molecules, but are rather associated with the solid as a whole. These solids *may* be conductors, if there are quantum states available for electrons to move about. When the valence electrons are associated with the solid as a whole, you can approximate the potential in which they move as as three-dimensional square well the size of the solid. Of course, this isn't exactly true, as there will be localized potential wells where each atom in the solid is. However, it is reasonable to approximate the solid as a fixed lattice, with valence electrons potentially free to move through it.

If you associate the valence electrons with the solid as a whole, you then need to think about the states available to those electrons. Since we are talking about a chunk of material that may have a huge number of atoms (e.g. about 10^{22} atoms if we're talking a mere 1g sample of Copper), we're also talking about a huge number of valence electrons, and a huge number of available states. Depending on how you model it, you can think of the states as resulting from the lattice, or as resulting from the effective square well potential in which the electrons move. The nature of the underlying lattice *does* matter. Typically, the states available to the valence electrons come in bands of energy, with gaps between the bands; this is one way in which a simple square well does not reflect the nature of the potential (where the distribution of states would be continuous). Three examples of solids with bands and band gaps are shown below. Each diagram is an energy level diagram. In a shaded band, there are *many* electron states stacked one on top of the other. Because electrons are fermions, however, each individual state may only have two electrons (two, not one, because of electron spin).



Electron energy states in the band gap model of solids. Increasing energy is upwards on all three diagrams. Dark gray bands indicate bands of many energy states that are filled. Light Gray bands indicate bands of many empty energy states.

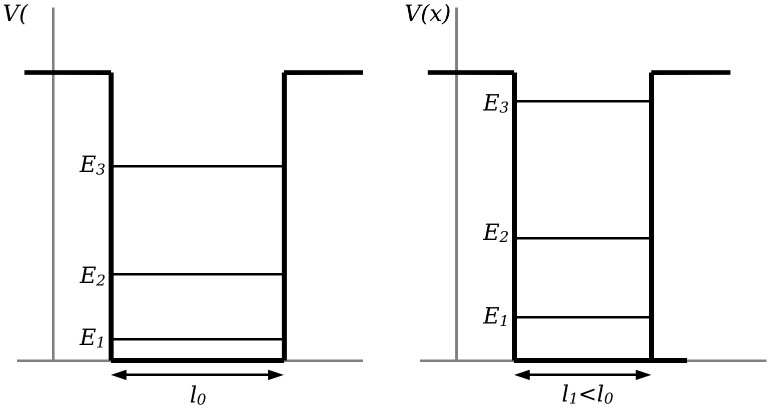
In the figure on the left, the top filled band is called the *valence band*. Again, this band, and the band below it, represent a huge number of closely spaced energy states for valence electrons. The energy of the top filled state (assuming the solid as a whole is in its lowest overall energy state) is called the *Fermi energy*. These states are *not* places where the electron can be, in the sense of places in space. Electrons in higher states have more energy than electrons in lower states, so they are in fact moving around. However, they're not moving around in a way that would allow electrons to flow in a coordinated fashion from one side of the solid to another. Rather, they're moving around in the same way that an electron in a higher orbital of an atom that has some kinetic energy associated with it. The electron is still tied to the solid, and stuck in the energy state it's in. But, that state is not localized to one nucleus; it's associated with the insulator as a whole. That there are no empty states nearby is why nothing can change that would allow for electric conduction.

The first empty band above the filled band is the *conduction band*. If you want to conduct electricity— that is, allow charge to move through the solid— you need to be able to get electrons out of the states where they're fixed, and into states that have empty states nearby. If there is a band gap between the valence band and the conduction band, this is not very easy; it takes a lot of energy to get an electron out of the state where it normally is and into a state where there are many neighboring states, allowing it to move in a coordinated way through the solid. The conduction band is the band where there are lots of empty states that the electron can make transitions between (or occupy in a superposition of states) so as to allow it to move through the solid and thereby carry electric current. (Electric current ultimately is just the transfer of net electron charge from one side of the conductor to the other.) *Every* material will ultimately carry electron current, if you apply enough potential to it. However, the wider the band gap, the more energy it takes to liberate an electron from the valence band and into the conduction band, thereby allowing it to move freely. If there is a wide gap, we would consider the material an insulator.

In the middle figure above, the material is a conductor. Here, the Fermi energy is in the *middle* of a band of available energy states. As such, electrons in the top filled states have many nearby states available to them. They may freely change states, and thus are able to go into states that can correspond to electrons transferring charge through the material.

The rightmost figure above is a semiconductor. Here, there is a band gap between the valence and conduction bands, but that band gap is quite small. Because of thermal excitations— electrons bumped about by the vibrating atoms resulting from the non-zero temperature— a tiny fraction of the electrons will in fact be up in the conduction band. However, the conductivity of this material remains small, because not very many electrons are up there. (The conductivity will go up with temperature, however, as more electrons statistically get bumped up into the conduction band.) The typical band gap in a semiconductor is around 1–2 eV, which is a lot larger than the 1/40 eV that is the average energy available to one particle at room temperature. By *doping* a semiconductor— that is, adding impurities— you can either add additional valence electrons inside the band gap at an energy just below the conduction band, or additional valence states at an energy inside the band gap just above the valence band. By putting such doped semiconductors together, you can create devices with interesting electrical properties, such as diodes and transistors.

One thing that is interesting about solids is what happens when you try to compress the solid. In a square well, if you decrease the width of the well the energy of all the states increases:



If you think about squeezing a solid, what this means is that you're pushing the states occupied by the valence electrons to higher energy levels, and therefore you must be putting energy into the solid. The need to exert energy on a solid when squeezing will manifest as a pressure (force per area) that resists the force trying to squeeze the solid. While you might think that it is the atoms and molecules themselves resisting being pushed closer together that makes a solid resist being compressed, this is not the whole case. In fact, this additional energy put into the valence electrons— which, remember, are not associated with individual electrons, but with the solid as a whole— contributes significantly to the restorative pressure of a compressed solid. This pressure is called *Fermi degeneracy pressure*. In this circumstance, "degeneracy" is a technical term referring to electrons all being packed into states as tightly as they can. The Fermi degeneracy pressure that resists the compression of a solid is a direct result of those states rising to higher energies as a result of the compression.

16.5 Liquids, Gasses, and Plasmas

If you put enough energy into a solid, eventually it will melt. At this point, the atoms and molecules in the solid are no longer held together in the crystal, lattice, or other structure. Instead, they have enough energy to break whatever bonds (covalent, ionic, or otherwise) holding them together, and now they can flow past each other. The atoms and molecules are still largely packed together as closely as they can go, and there still are bonds of a sort holding the broadly together, but no longer are they fixed in place. Such a state would be called a liquid.

If you raise the temperature of a liquid enough (to the "boiling point"), and continue to add energy, you can break the residual forces holding the liquid together, and give each molecule enough energy that it starts to bounce about freely. Add enough energy, and the liquid becomes a gas. At this point, each molecule or atom of the gas moves around freely. Molecules do very regularly collide with each other, but they're not in constant interaction any more as is the case with a liquid.

Indeed, if you investigate the momentum and kinetic energy states available to molecules in a gas, the states are mostly empty. Unlike the valence electrons in a solid, which mostly fill up the lowest states available to them, there are many, many empty lower energy states for every gas molecule. As such, it's easy to change the energy of a gas molecule by a very small fraction of its current energy, as there are so many empty states about. At this point, you can approximate the available energy states as a continuum, and the gas molecules behave like classical particles. A traditional gas is described by the *Maxwell Boltzmann* distribution, which specifies the fraction of gas molecules that will be moving at any velocity given the temperature of the gas and the mass of each molecule. This description applies to the air around us, to the gas between the stars, and to the gas in atmospheres of stars.

Solid, liquid, and gas are the standard "three states of matter." If you ionize a gas— that is, if you tear an electron off of a substantial fraction of the gas atoms— it becomes a plasma, the fourth state of matter. Because the individual particles in a plasma are electrically charged (either positive ions or negative electrons), electric and magnetic fields can greatly influence the behavior of a plasma. There are a few ways to create a plasma. One is just to raise the temperature of the gas high enough so that the average kinetic energy of any particle is high enough that collisions will tend to ionize gas molecules. Another is to shine ionizing radiation— generally ultraviolet or x-ray radiation— on the gas. Interstellar gas around young massive stars is typically mostly ionized as a result of the radiation from those stars, even though the temperature of the gas itself isn't high enough to maintain that ionization. A final way to ionize gas is to run high energy particles through it. For instance, if you can shoot an electron beam through dilute gas, it will tend to ionize the gas it passes through. This is how plasma discharge tubes are created.

16.5.1 Quantum Gasses

It *is* possible to create a fundamentally quantum gas, however. If you can lower the temperature of a gas enough while allowing it to stay as a gas e.g. by keeping it at a low enough density that it does not condense into a liquid or solidify), you can get to the point that a substantial fraction of the gas is occupying the lowest states available to it. At this point, the gas is no longer adequately described by classical physics. If the gas is composed of fermions (i.e. each molecule has net half-integral spin), you will have what's called a "Fermi gas", that is analogous in many ways to the valence electrons in a solid. If, on the other hand, the gas is composed of bosons, and you can lower its temperature enough, it's possible to create a "Bos-Einstein condensate", where a substantial fraction of the gas molecules all drop into the *same* state (something that would be impossible for atoms). At this point, you can see coherent quantum phenomena for the whole gas, such as interference, because of all the molecules or atoms that are in the same state. A Bos-Einstein condensate was first created in 1995 (Anderson et al., 1995); this work received the Nobel Prize in physics in 2009.

16.6 Planets, Stars, Galaxies, and Clusters

Once you get past the sizes of everyday solids, liquids and gasses, you enter the realm of astronomical objects. In our Solar System, such objects range from lowly asteroids, through dwarf planets such as Pluto or Ceres, through rocky planets such as the Earth or Mars, on up through the gas giants such as Saturn or Jupiter. However, the vast majority of the mass of our Solar System is in the Sun, the star about which everything else orbits. The Sun is a ball of gas, 300,000 times the mass of the Earth. It is composed of about 74% Hydrogen, about 24% Helium, and 2% everything else. This is a very different composition from the Earth, which is composed mostly of heavier elements. However, the composition of the Sun is more typical of the composition of the Universe as a whole— indeed, the Sun has a greater than average fraction of heavy elements!

Stars collect together into galaxies, gravitationally bound systems of millions, billions, or even trillions of stars. Galaxies themselves collect together into groups and clusters, which may themselves have thousands of galaxies. The groups and clusters we can identify are part of a filamentary structure that fills the Universe. Most galaxy groups and clusters are found in this filamentary web, with vast voids between them. On the largest scales, the universe is homogeneous. That is, if you look at one spot in the universe about a billion or so light-years across, it looks pretty much the same as another, with galaxy clusters on filaments surrounding voids.

16.7 Dark Matter and Dark Energy

All of the matter discussed so far in this chapter only makes up 5% of the energy density of the Universe. (We can talk about mass density and energy density interchangeably, as mass is a form of energy. The amount of energy E in mass m can be found through the famous conversion $E = mc^2$.) Several different lines of evidence have shown us that most of the mass in galaxy clusters is not the luminous mass we can observe. Indeed, this mass can't be baryonic at all. Dark Matter is the name given to this mass; it's not dark the way dust is, which absorbs light is. It neither absorbs nor emits light; it only interacts with light gravitationally (and has been observed partly through the gravitational lensing effect). Dark Matter interacts only via gravity and, perhaps, the Weak Force. In this way, it is similar to neutrinos. We have not identified a particle that can make up Dark Matter. We're very sure that it's there, and we're very sure that it makes up most of galaxies and galaxy clusters, but we don't know exactly what it is.

Dark Matter, however, itself only makes up about 20-25% of the energy density of the Universe. In the late 1990's, astronomers discovered that the expansion of the Universe is *accelerating*; this discovery was awared the Nobel Prize in Physics in 2011. This is not something you would expect from regular matter or from Dark Matter. With matter (including Dark Matter), as well as normal forms of energy such as radiation (light), the gravitational effect is attractive. The result would be to tend to slow down the expansion of the Universe. For the Universe to be accelerating, there must be something else in it. Dark Energy is the name given to this unknown substance that evidently makes up about 75% of the energy density of the Universe and that has a negative gravitational effect. Dark Energy is even more unknown than Dark Matter, and indeed some believe that it doesn't really exist as a substance, but is a pointer to our theories of gravity breaking down. Many believe that the most likely candidate for Dark Energy is vacuum energy (see Section 11.4.2), but for now nature of Dark Energy remains one of the primary outstanding unanswered questions in both astronomy and fundamental physics.

Bibliography

Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E., and Cornell, E. A. (1995). Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor. *Science*, 269:198–201.

Arndt, M., Nairz, O., Vos-Andreae, J., Keller, C., van der Zouw, G., and Zeilinger, A. (1999). Wave-particle duality of C_{60} molecules. *Nature*, 401:680–682.

Bernstein, J. (2010). The Stern Gerlach Experiment. *ArXiv e-prints*.

Dragoset, R. A., Musgrove, A., Clark, C. W., and Martin, W. C. (2003). Periodic table: Atomic properties of the elements. Technical Report NIST SP 966, National Institute of Standards and Technology, Washington, DC.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47(10):777–780.

Nakamura, K. and Particle Data Group (2010). Review of Particle Physics. *Journal of Physics G Nuclear Physics*, 37(7):075021–+.

Schlosshauer, M. (2004). Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 76:1267–1305.